Quantitative Methods

Quantitative Methods (1)

The Time Value of Money

1. The Time Value of Money and Interest Rates

The **time value of money** (TVM) refers to the fact that $1 today is worth more than $1 in the future. This is because the $1 today can be invested to earn interest immediately. The TVM reflects the relationship between present value, future value, time, and interest rate. The time value of money underlies rates of return, interest rates, required rates of return, discount rates, opportunity costs, inflation, and risk. It reflects the relationship between time, cash flow, and interest rate.

There are three ways to interpret interest rates:

- **Required rate of return** is the return required by investors or lenders to postpone their current consumption.
- **Discount rate** is the rate used to discount future cash flows to allow for the time value of money (that is, to bring a future value equivalent to present value).
- **Opportunity cost** is the most valuable alternative investors give up when they choose what to do with money.

In a *certain world*, the interest rate is called the risk-free rate. For investors preferring current to future consumption, the risk-free interest rate is the rate of compensation required to postpone current consumption. For example, the interest rate paid by T-bills is a risk-free rate of interest.

In an *uncertain world*, there are two factors that complicate interest rates:

- **Inflation**: When prices are expected to increase, lenders charge not only an opportunity cost for postponing consumption but also an inflation premium that takes into account the expected increase in prices. The **nominal cost of money** consists of the **real rate** (a pure rate of interest) and an **inflation premium**.

- **Risk**: Companies exhibit varying degrees of uncertainty concerning their ability to repay lenders. Lenders therefore charge interest rates that incorporate **default risk**. The return that borrowers pay thus comprises the nominal risk-free rate (real rate + an inflation premium) and a default risk premium.

Compounding is the process of accumulating interest over a period of time. A compounding period is the number of times per year that interest is paid. **Continuous compounding** occurs when the number of compounding periods becomes infinite; interest is added continuously.

Discounting is the calculation of the present value of some known future value. Discount rate is the rate used to calculate the present value of some future cash flow. Discounted cash flow is the present value of some future cash flow.

2. Calculate the Effective Annual Rate

```
  There are three ways to quote interest rates for investments paying interest more than
```

- **Periodic interest rate** is the rate of interest earned over a single compounding period. For example, a bank may state that a particular CD pays a periodic quarterly interest rate of 3% that compounds 4 times a year.

- **Stated annual interest rate**, also called **quoted interest rate**, is the annual rate of interest that does not account for compounding within the year. It is the annual interest rate quoted by financial institutions and equal to the periodic interest rate multiplied by the number of compounding periods per year. For example, the stated annual interest rate of the above CD is 3% x 4 = 12%. It is strictly a quoting convention, and it does not give a future value directly.

- **Effective annual rate (EAR)** is the annual rate of interest that takes full account of compounding within the year. The periodic interest rate is the stated annual interest rate divided by m, where m is the number of compounding periods in one year: EAR = (1 + periodic interest rate)^m - 1. Note that the higher the compounding frequency, the higher the EAC.

For example, a $1 investment earning 8% compounded semi-annually actually earns 8.16%: $(1 + 0.08/2)^2 - 1 = 8.16$. The annual interest rate is 8%, and the effective annual rate is 8.16%.

Example

If the nominal interest rate is 8%, find the effective annual rate with quarterly compounding.

Method 1: By Formula

m = 4, EAR = $(1 + 0.08/4)^4 - 1 = 0.0824$

The effective interest rate with quarterly compounding is 8.24%.

Method 2: Texas Instruments

You will use the Interest Conversion (ICONV) worksheet

1. Press 2nd ICONV to select the worksheet
2. NOM will be displayed with the previous value
3. Press 2nd [CLR WORK] to clear the worksheet

Proceed as shown below:

Keystrokes: Display

2nd ICONV: NOM = previous value

2nd CLRWORK: NOM = 0.00

8 ENTER: NOM = 8.00

DownArrow: EFF = 0.00

DownArrow: C/Y = previous value

4 ENTER: C/Y = 4.00

DownArrow: EFF = 0.00

CPT: EFF = 8.24

Method 3: HP 12C

After you have set the calculator to END of period and cleared the financial registers, key in the nominal interest rate as a percentage.

Proceed as shown below:

Keystrokes: Display

g END: Previous value

f CLEAR FIN: 0.00000000

f CLEAR REG: 0.00000000

8 ENTER: 8.00000000

4n i 0.66666667

100 CHS PV: -100.00000000

FV: 108.2432160

100-: 8.24321600

We can also calculate the periodic interest rate given the effective annual interest rate.

3. The Future Value and Present Value of a Single Cash Flow

```
When you make a single investment today, its future value, received N years from now,
```

- FV = future value at time n
- PV = present value
- r = interest rate per period
- N = number of years

A key assumption of the future value formula is that interim interest earned is reinvested at the given interest rate (r). This is known as **compounding**.

In order to receive a single future cash flow N years from now, you must make an investment today in the following amount:

Notice that the future cash flow is discounted back to the present. Therefore, the interest rate is called the **discount rate**.

You should be able to calculate PVs and FVs using your calculator.

- N = number of periods
- I/Year = yield in market place or the required rate of return
- PV = present value
- PMT = payment amount per period
- FV = the future value of the investment

One can solve for any of the above variables. Just input the other variables and solve for the unknown. Using the calculator on the test will prove to be a very time-efficient manner of calculating present values and future values.

Example 1

An analyst invests $5 million in a 5-year certificate of deposit (CD) at a local financial institution. The CD promises to pay 7% per year compounded annually. The institution also allows him to reinvest the interest at the same CD rate for the duration of the CD. How much will the analyst have at the end of five years if his money remains invested at 7% for five years with no withdrawals of interest?

Before using the Texas Instruments BAII PLUS and HP 12C calculator, it is essential to ensure that your settings are correct. The default settings on the calculator are not necessarily the settings you need when making the calculations. Follow these steps to ensure that your calculator is correctly set.

Texas Instruments BAII PLUS Settings

- Press 2nd QUIT 2nd [CLR TVM] to clear the worksheet.
- Press 2nd [P/Y] to enter payments per year and/or compounding periods per year.
- The P/Y label and current value are displayed. The default value is 12. You must now key in 1 and then ENTER since you want 1 payment per year.
- If the question says there are 12 payments per year, you would change this to 12.

HP 12C Settings

- Turn the calculator on by pressing the ON key.
- Clear the memory and set decimals to 2 places by pressing the following keys:
- CLEAR REG f 2 - 0.00 will display.
- CLEAR FIN - this clears all the data in the financial mode.

The calculator keys to press are:

If you are given the FV and need to solve for PV, the calculator keys to press are:

When compounding periods are not annual

Some investments pay interest more than once a year. When you calculate these amounts, make sure that periodic interest rates correspond to the number of compounding periods in the year. For example, if time periods are quoted in quarters, quarterly interest rates should be used.

When compounding periods are other than annual

- r_s = the quoted annual interest rate
- m = the number of compounding periods per year
- N = the number of years.

Example 2

An analyst invests $5 million in a 5-year certificate of deposit (CD) at a local financial institution. The CD promises to pay 7% per year, compounded semi-annually. The institution also allows him to reinvest the interest at the same CD rate for the duration of the CD. How much will the analyst have at the end of five years if his money remains invested at 7% for five years with no withdrawals of interest?

The calculator keys to press are:

Note that the answer is greater than when the compounding was annual. This is because interest is earned twice a year instead of only once.

If the number of compounding periods becomes infinite, interest is compounding continuously. Accordingly, the future value N years from now is computed as follows:

4. The Future Value and Present Value of a Series of Equal Cash Flows (Ordinary Annuities, Annuity Dues, and Perpetuities)

```
<b>Annuity</b> is a finite set of sequential cash flows, all with the same value.<p> <
```

Ordinary annuity has a first cash flow that occurs one period from now (indexed at t = 1). In other words, the payments occur at the end of each period.

- **Future value of a regular annuity**

  where

  - A = annuity amount
  - N = number of regular annuity payments
  - r = interest rate per period

- **Present value of a regular annuity**

Annuity due has a first cash flow that is paid immediately (indexed at t = 0). In other words, the payments occur at the beginning of each period.

- **Future value of an annuity due**

  This consists of two parts: the future value of one annuity payment now, and the future value of a regular annuity of (N -1) period. Calculate the two parts and add them together. Alternatively, you can use this formula:

  Note that, all other factors being equal, the future value of an annuity due is equal to the future value of an ordinary annuity multiplied by (1 + r).

- **Present value of an annuity due**

  This consists of two parts: an annuity payment now and the present value of a regular annuity of (N - 1) period. Use the above formula to calculate the second part and add the two parts together. This process can also be simplified to a formula:

  Note that, all other factors being equal, the present value of an annuity due is equal to the present value of an ordinary annuity multiplied by (1 + r).

Hint: Remember these formulas - you can use them to solve annuity-related questions directly, or to double-check the answers given by your calculator.

A **perpetuity** is a perpetual annuity: an ordinary annuity that extends indefinitely. In other words, it is an infinite set of sequential cash flows that have the same value, with the first cash flow occurring one period from now.

This equation is valid for a perpetuity with level payments, positive interest rate r. The first payment occurs one period from now (like a regular annuity). An example of a perpetuity is a stock paying constant dividends.

Example: Future value of a regular annuity

An analyst decides to set aside $10,000 per year in a conservative portfolio projected to earn 8% per annum. If the first payment he makes is one year from now, calculate the accumulated amount at the end of 10 years.

Method 1: Using a formula

- Identify the given variables: A = 10,000, r = 0.08, N = 10

- Identify the appropriate formula: $FV = A \times \{[(1 + r)^N - 1] / r\}$
- Solve for the unknown: $FV = 10{,}000\ \{[(1 + 0.08)^{10} - 1] / 0.08\} = \$144{,}865$

Method 2: Using a calculator

Texas Instruments settings:

- 2nd P/Y = 1 and key in 1 ENTER.
- SET END since this is a regular annuity. You do this by pressing 2nd BGN 2nd SET until you see END displaying. Press 2nd SET twice if necessary. After setting to END, you must always press 2nd QUIT and then continue.

Exploration: Change the problem to an annuity due (i.e., SET BGN) and compare the amounts. (The answer is $156,454.87 - a difference of $11,589.25)

5. The Future Value and Present Value of a Series of Uneven Cash Flows

```
   A series of uneven cash flows means that the cash flow stream is uneven over many time
```

Present Value

When we have unequal cash flows, we must first find the present value of each individual cash flow and then the sum of the respective present values. (This is usually accomplished with the help of a spreadsheet.)

Future Value

Once we know the present value of the cash flows, we can easily apply time-value equivalence by using the formula to calculate the future value of a single sum of money (LOS a).

Example

John wants to pay off his student loan in three annual installments: $2,000, $4,000 and $6,000, respectively, in the next three years. How much should John deposit into his bank account today if he wants to use the account balance to pay off the loan? Assume that the bank pays 8% interest, compounded annually.

6. The Cash Flow Additivity Principle

```
   The <b>additivity principle</b>: Dollar amounts indexed at the same point in time are
```

Suppose we are considering two series of cash flows (A and B). The annual interest rate is 5%. We want to know the future value of combined cash flows at $t = 3$.

- We can calculate the future value of each series and add them up. The future value of series A is $100 \times 1.05^2 + 100 \times 1.05 + 100 = 315.25$ and the future value of series B is $150 \times 1.05^2 + 150 \times 1.05 + 150 = 472.875$. The future value of A + B is 788.125.

- Alternatively, we can add the cash flows of each series first and then find the future value of the combined cash flows: $250 \times 1.05^2 + 250 \times 1.05 + 250 = 315.25 = 788.125$.

We can use this principle to solve many uneven cash flow problems if we add dollars indexed at the same point in time. Consider a cash flow series, A, with $100 indexed at t = 1, 2, 3 and 5, and $0 at t = 4. This series is an almost-even cash flow, flawed only by the missing $100 at t = 4. How do we find the present value of this series?

- We can create an annuity B with $100 indexed at t = 1, 2, 3, 4, 5. It's easy to find the present value of this series.
- Then we isolate an easily evaluated cash flow B - A; it has a single cash flow of $100 at t = 4. It's also easy to find the present value of this single cash flow.
- We then subtract the present value of B - A from the present value of B.

Organizing, Visualizing, and Describing Data

1. Data Types

```
<p> </p><b>Data</b> can be defined as a collection of numbers, characters, words, text
```

Numerical versus Categorical Data

From a statistical perspective, data can be classified as numerical data and categorical data.

Numerical data are values to represent measured or counted quantities as a number. They can be further split into two types:

- **Continuous data** can be measured and take on any numerical value in a specified range of values. There are normally lots of decimal places involved and (theoretically, at least) there are no gaps between permissible values (i.e., all values can be included in the data set).

  Examples would include the height of a person and the time to complete an assignment. These values can be measured using sufficiently accurate tools to numerous decimal places.
- **Discrete data** result from a counting process and therefore are limited to a finite number of values. That is, the values in the data set can be counted. There are distinct spaces between the values, such as the number of children in a family or the number of shares comprising an index.

Categorical data are values that describe a quality or characteristic of a group of observations and usually take only a limited number of values that are mutually exclusive. They can be further classified into nominal data and ordinal data.

Nominal data are not amenable to being organized in a logical order.

- Nominal measurement represents the weakest level of measurement.
- It consists of assigning items to groups or categories.
- No quantitative information is conveyed and no ordering (ranking) of the items is implied.
- Nominal scales are qualitative rather than quantitative.

Religious preference, race, and sex are all examples of nominal scales. Another example is portfolio managers categorized as value or growth style will have a scale of 1 for value and 2 for growth. Frequency distributions are usually used to analyze data measured on a nominal scale. The main statistic computed is the mode. Variables measured on a nominal scale are often referred to as categorical or qualitative variables.

Ordinal data are categorical values that can be logically ordered and ranked.

- Measurements on an ordinal scale are categorized.
- The various measurements are then ranked in their categories.
- Measurements with ordinal scales are ordered with higher numbers representing higher values. The intervals between the numbers are not necessarily equal.

Example 1

On a 5-point rating scale measuring attitudes toward gun control, the difference between a rating of 2 and a rating of 3 may not represent the same difference as that between a rating of 4 and a rating of 5.

Example 2

Two categories might be value and growth. Within each category, the portfolio managers measured will be weighted according to performance on a scale from 1 to 10, with 1 being the best- and 10 the worst-performing manager.

There is no "true" zero point for ordinal scales, since the zero point is chosen arbitrarily. The lowest point on the rating scale in the example was arbitrarily chosen to be 1. It could just as well have been 0 or -5.

Cross-Sectional versus Time-Series versus Panel Data

Based on how they are collected, data can be categorized into three types.

Time-series data is a set of observations for a single observational unit collected at usually discrete and equally spaced time intervals. Examples: the daily closing price of a certain stock recorded over the last six weeks, weekly sales figures of ice cream sold during a holiday period at a seaside resort.

Cross-sectional data are observations that come from different observational units at a single point in time. The underlying population should consist of members with similar characteristics. For example, suppose you are interested in how much companies spend on research and development expenses. Firms in some industries, such as retail, spend little on research and development (R&D), while firms in industries such as technology spend heavily on R&D. Therefore, it's inappropriate to summarize R&D data across all companies. Rather, analysts should summarize R&D data by industry and then analyze the data in each industry group.

Panel data is a mix of time-series and cross-sectional data that consists of observations through time on one or more variables for multiple observation units.

Structured versus Unstructured Data

Based on whether or not data are in a highly organized form, they can be classified into structured and unstructured types.

Structured data are highly organized in a pre-defined manner, usually with repeating patterns. Market data, fundamental data and analytical data are typical examples.

Unstructured data do not follow any conventionally organized forms. Common types are text, audio, video. They are typically alternative data as they are usually collected from unconventional sources such as individuals, business processes and sensors.

Typically, unstructured data must first be transformed into structured data that financial models can process.

2. Organizing Data for Quantitative Analysis

```
    Raw data are typically organized into either a one-dimensional array, which is suitabl
```

3. Summarizing Data Using Frequency Distributions

```
    <p> </p>Very often, the data available is vast, leading to a situation where dealing w
```

An **interval**, also called a **class**, is a set of values within which an observation falls.

- Each interval has a lower limit and an upper limit.
- Intervals must be all-inclusive and non-overlapping.

A **frequency distribution** is a tabular display of data categorized into a small number of non-overlapping intervals. Note that:

- Each observation can only lie in one interval.
- The total number of intervals will incorporate the whole population.
- The range for an interval is unique. This means a value (observation) can only fall into one interval.

It is important to consider the number of intervals to be used. If too few intervals are used, too much data may be summarized and we may lose important characteristics; if too many intervals are used, we may not summarize enough.

A frequency distribution is constructed by dividing the scores into intervals and counting the number of scores in each interval. The actual number of scores and the percentage of scores in each interval are displayed. This helps in the analysis of large amount of statistical data, and works with all types of measurement scales.

- **Absolute frequency** is the actual number of observations in a given interval.

- **Relative frequency** is the result of dividing the absolute frequency of each return interval by the total number of observations.

- **Cumulative absolute frequency** and **cumulative relative frequency** are the results from cumulating the absolute and relative frequencies as we move from the first to the last interval.

The following steps are required when organizing data into a frequency distribution together with suggestions on constructing the frequency distribution.

- Identify the highest and lowest values of the observations.

- Setup classes (groups into which data is divided). The classes must be mutually exclusive and of equal size.

- Add up the number of observations and assign each observation to its class.

- Count the number of observations in each class. This is called the class frequency.

The **relative frequency** for a class is calculated by dividing the number of observations in a class by the total number of observations and converting this figure to a percentage (multiplying the fraction by 100). Simply, relative frequency is the percentage of total observations falling within each interval. It is another way of analyzing data; it tells us, for each class, what proportion (or percentage) of data falls in that class.

Let's look at an example.

The following table shows the holding period returns of a portfolio of 40 stocks.

The highest HPR is 32% and the lowest one is -27%. Let's use 6 non-overlapping intervals, each with a width of 10%. The first interval starts at -27% and the last one ends at 33%. Therefore, the entire range of the HPRs is covered.

Hint: If, in an examination, your relative frequency column does not sum to 1 (or 100%), you know that you have made a mistake.

4. Summarizing Data Using a Contingency Table

```
    <p> </p>To find patterns between variables we can use a <b>contingency table</b>, whic
```

The table below shows the favorite leisure activities for 50 adults - 20 men and 30 women.

Entries in the "Total" row and "Total" column are called **marginal frequencies**. They represent the frequency distribution for each variable. Entries in the body of the table are called **joint frequencies**.

One benefit of having data presented in a contingency table is that it allows one to more easily perform basic probability calculations.

There's a 16/50 (32%) probability that the person sampled likes TV as his/her favorite leisure activity, while the probability that a random participant is female is 30/50 (60%). What's more, computing conditional probabilities is made easier using contingency tables, e.g., the probability that a person's favorite leisure activity is to dance given that the person is male is 2/20=10%, while the conditional probability that a person is male given that sports are preferred is 10/16 (62.5%).

One application is for evaluating the performance of a classification model (using a **confusion matrix**). It is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

Another is to investigate a potential association between two categorical variables by performing a chi-square test of independence. For example, we can test whether the two variables being examined - in this case, gender and favorite way to eat ice cream - are actually independent as they've been assumed throughout. This is done by computing, for each cell, the expected frequency E, comparing it to the observed frequency O, and then performing a chi-squared test.

Note the los says "interpret a contingency table.": you probably don't need to know how to perform a chi-square test of independence for the exam.

5. Data Visualization

```
    <b>Data visualization</b> is the graphical representation of information and data. By
```

Histogram and Frequency Polygon

A **histogram** is a bar chart that displays a frequency distribution. It is constructed as follows:

- The class frequencies are shown on the vertical (y) axis (by the heights of bars drawn next to each other).
- The classes (intervals) are shown on the horizontal (x) axis.
- There is no space between the bars.

From a histogram, we can see quickly where most of the observations lie. The shapes of histograms will vary, depending on the choice of the size of the intervals.

The **frequency polygon** is another means of graphically displaying data. It is similar to a histogram but the bars are replaced by a line joined together. It is constructed in the following manner:

- Absolute frequency for each interval is plotted on the vertical (y) axis.
- The midpoint of each class (interval) is shown on the horizontal (x) axis.
- Neighboring points are connected with a straight line.

Unlike a histogram, a frequency polygon adds a degree of *continuity* to the presentation of the distribution.

It is helpful, when drawing a frequency polygon, first to draw a histogram in pencil, then to plot the points and join the lines, and finally to rub out the histogram. In this way, the histogram can be used as an initial guide to drawing the polygon.

A **cumulative frequency distribution chart**is the sum of the class and all classes below it in a frequency distribution.

Bar Chart

A **bar chart** is a way of summarizing a set of categorical data. The height of each bar is proportional to a specific aggregation (for example the sum of the values in the category it represents). The categories could be something like an age group or a geographical location. A bar chart usually compares different categories. It is useful for looking at a set of data and making comparisons.

Although they look the same, bar charts and histograms have one important difference: they plot different types of data. Plot discrete data on a bar chart, and plot continuous data on a histogram.

A bar chart is used for when you have categories of data: Types of movies, music genres, or dog breeds. Itâ€™s also a good choice when you want to compare things between different groups. You could use a bar graph if you want to track change over time as long as the changes are significant (for example, decades or centuries). If you have continuous data, like peopleâ€™s weights or IQ scores, a histogram is best.

Grouped bar charts or **stacked bar charts** can present the frequency distribution of multiple categorical variable simultaneously.

Tree-Map

Tree-maps are an alternative way of visualizing the hierarchical structure of a tree diagram while also displaying quantities for each category via area size. Each category is assigned a rectangle area with their subcategory rectangles nested inside of it.

The main advantages:

- identify the relationship between two elements in a hierarchical data structure;
- accurately display multiple elements together;
- show ratios of each part to the whole;
- visualize attributes by size and color coding.

The downside to a tree-map is that it doesn't show the hierarchal levels as clearly as other charts.

Word Cloud

A **word cloud** is a novelty visual representation of text data, typically used to depict keyword metadata on websites, or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color.

Line Chart

A **line chart** is generally used to show trend of a measure (or a variable) over time. Using a line chart, one can see the pattern of any dependent variable over time like share price, EPS of a company, weather recordings (like temperature, precipitation or humidity), etc.

Why do you need a Line Chart?

1. To see changes of a dependent variable over time.

1. To identify trends and spot spikes and dips

1. To compare patterns of multiple sections

Scatter Plot

A **scatter plot** (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.

Relationships between variables can be described in many ways: positive or negative, strong or weak, linear or nonlinear.

Scatter plots are a powerful tool for finding patterns between two variables, for assessing data range, and for spotting extreme values.

Through the use of a â€œLine of Best Fitâ€ or a trend line, scatter plots to help identify trends.

Two common issues have been identified with the use of scatter plots â€" over-plotting and the interpretation of causation as correlation.

Over-plotting occurs when there are too many data points to plot, which results in the overlapping of different data points. It can make relationship identification between variables challenging.

Concerning correlation, it is important to remember that correlation does not mean that the changes observed in one variable are responsible for the changes observed in another variable. Correlation should not be interpreted as causation. Causation implies that an event occurring will have an impact on an outcome.

Heat Map

A **heat map** is data analysis tool that uses color the way a bar graph uses height and width. It organizes and summarizes data in a tabular format and represents it using a color spectrum. It is often used in displaying frequency distributions or visualizing the degree of correlation among different variables.

Guide to Selecting among Visualization Types

The key consideration when selecting among chart types is the intended purpose of visualizing data. That is, whether it is for exploring/presenting distributions or relationships or for making comparisons. Exhibit 34 in the textbook is a great summary for the purpose.

6. Measures of Central Tendency

```
   <p> </p>A <b>population</b> consists of an entire set of objects, observations, or scc
```

A **parameter** is a numerical quantity measuring some aspect of a population of scores. The mean, for example, is a measure of central tendency. Parameters are rarely known and are usually estimated by statistics computed in samples.

Estimates of these parameters taken from a sample are called **statistics**.

Hint: One way to easily remember these terms is to recall that "population" and "parameter" both start with a "p," and "sample" and "statistic" both start with a "s."

Measures of central tendency specify where data are centered.

The **population mean** is the average for a finite population. It is unique; a given population has only one mean.

where:

- N = the number of observations in the entire population
- X_i = the ith observation
- Î£X_i = add up X_i, where i is from 0 to N

The **sample mean** is the average for a sample. It is a statistic and is used to estimate the population mean.

where n = the number of observations in the sample

Arithmetic Mean

The arithmetic mean is what is commonly called the average. It is the most widely used measure of central tendency. When the word "mean" is used without a modifier, it can be assumed to refer to the arithmetic mean. The mean is the sum of all scores divided by the number of scores.

- All interval and ratio data sets (e.g., incomes, ages, rates of return) have an arithmetic mean.
- All data values are considered and included in the arithmetic mean computation.
- A data set has only one arithmetic mean. This indicates that the mean is unique.
- The arithmetic mean is the only measure of central tendency where the sum of the deviations of each value from the mean is always zero. **Deviation** from the arithmetic mean is the distance between the mean and an observation in the data set.

The arithmetic mean has the following disadvantages:

- The mean can be affected by extremes, that is, unusually large or small values (outliners). In such cases we can either do nothing, or delete all the outliners (trimmed mean), or replace the outliners with another value (winsorized mean).
- The mean cannot be determined for an open-ended data set (i.e., n is unknown).

Median

In English, the word "mediate" means to go between or to stand in the middle of two groups, in order to act as a referee, so to speak. The median does the same thing; it is the value that stands in the middle of the data set, and divides it into two equal halves, with an equal number of data values in each half.

To determine the median, arrange the data from highest to lowest (or lowest to highest) and find the middle observation. If there are an odd number of observations in the data set, the median is the middle observation (n + 1)/2 of the data set. If the number of observations is even, there is no single middle observation (there are two, actually). To find the median, take the arithmetic mean of the two middle observations.

The median is less sensitive to extreme scores than the mean. This makes it a better measure than the mean for highly skewed distributions. Looking at median income is usually more informative than looking at mean income, for example. The sum of the absolute deviations of each number from the median is lower than the sum of absolute deviations from any other number.

Note that whenever you calculate a median, it is imperative that you place the data in order first. It does not matter whether you order the data from smallest to largest or from largest to smallest, but it does matter that you order the data.

Mode

Mode means fashion. The mode is the "most fashionable" number in a data set; it is the most frequently occurring score in a distribution and is used as a measure of central tendency. A set of data can have more than one mode, or even no mode. When all values are different, the data set has no mode. When a distribution has one value that appears most frequently, it is said to be **unimodal**. A data set that has two modes is said to be **bimodal**.

The advantage of the mode as a measure of central tendency is that its meaning is obvious. Like the median, the mode is not affected by extreme values. Further, it is the only measure of central tendency that can be used with nominal data. The mode is greatly subject to sample fluctuations and, therefore, is not recommended for use as the only measure of central tendency. A further disadvantage of the mode is that many distributions have more than one mode. These distributions are called "multimodal."

The Weighted Mean

The weighted mean is computed by weighting each observed value according to its importance. In contrast, the arithmetic mean assigns equal weight to each value. Notice that the return of a portfolio is the weighted mean of the returns of individual assets in the portfolio. The assets are weighted on their market values relative to the market value of the portfolio. When we take a weighted average of forward-looking data, the weighted mean is called **expected value**.

Example

A year ago, a certain share had a price of $6. Six months ago, the same share had a price of $6.20. The share is now trading at $7.50. Because the most recent price is the most reliable, we decide to attach more relevance to this value. So, suppose we decide to "weight" the prices in the ratio 1:2:4, so that the current share price is twice as important as the price from six months ago, which in turn is twice as important as the price from last year.

The weighted mean would then be: (1 x 6 + 2 x 6.2 + 4 x 7.5) / (1 + 2 + 4) = $6.91. If we calculated the mean without weights, we'd get: (6 + 6.2 + 7.5) / 3 = $6.57. The fact that we've given more importance to the most recent (higher) share price inflates the weighted mean relative to the un-weighted mean.

The Geometric Mean

The geometric mean has three important properties:

- It exists only if all the observations are greater than or equal to zero. In other words, it cannot be determined if any value of the data set is zero or negative.
- If values in the data set are all equal, both the arithmetic and geometric means will be equal to that value.
- It is always less than the arithmetic mean if values in the data set are not equal.

It is typically used when calculating returns over multiple periods. It is a better measure of the compound growth rate of an investment. When returns are variable by period, the geometric mean will always be less than the arithmetic mean. The more dispersed the rates of returns, the greater the difference between the two. This measurement is not as highly influenced by extreme values as the arithmetic mean.

The Harmonic Mean

The harmonic mean of n numbers x_i (where i = 1, 2, ..., n) is:

The special cases of n = 2 and n = 3 are given by:

and so on.

For n = 2, the harmonic mean is related to arithmetic mean A and geometric mean G by:

The harmonic mean may be viewed as a special type of weighted mean in which an observation's weight is inversely proportional to its magnitude. It is used most often when the data consists of rates and ratios, such as P/Es.

Which Mean to Use

It depends on lots of factors: outliners, symmetric distribution? compounding?

The mean, median, and mode are equal in symmetric distributions. The mean is higher than the median in positively skewed distributions and lower than the median in negatively skewed distributions. Extreme values affect the value of the mean, while the median is less affected by outliers. Mode helps to identify shape and skewness of distribution.

Harmonic Mean <= Geometric Mean <= Arithmetic Mean

7. Quantiles

```
  <p> </p>The median is the value that divides a distribution in half. Quantities such a
```

Quartile

There are 3 quartiles in a data set. Between them, they divide the data into 4 equal parts or quarters. The first quartile is called the lower quartile and is often denoted as Q_1. The second quartile is obviously just the median, as it is the middle value of the data set. The third quartile is called the upper quartile and is often denoted as Q_3.

You should note that Q_1 effectively splits the data set into the lower 25% of values and the upper 75% of values whereas Q_3 splits the data into the lower 75% of values and the upper 25% of values.

The distance between Q_1 and Q_3, namely $Q_3 - Q_1$, is called the **inter-quartile range**; it gives an indication of the spread of the middle 50% of the data set.

Quintile

There are 4 quintiles in a data set. Between them, they divide the data into 5 equal parts or fifths. Quintiles are not very commonly used.

You should note that the first quintile effectively splits the data set into the lower 20% of values and the upper 80% of values; the second quintile splits the data set into the lower 40% of values and the upper 60% of values, and so on.

Decile

There are 9 deciles in the data set. Between them, they divide the data into 10 equal parts, or tenths.

Obviously, the fifth decile is the median, as it is the middle value in the data set.

Percentile

There are 99 percentiles in the data set. Between them, they divide the data into 100 equal parts, or hundredths.

The fiftieth percentile is the median, as it is the middle value in the data set. The sixty-third percentile divides the data set into the lower 63% of values and the upper 37% of values, and so on.

Note that the 75th percentile is also the same value as Q_3, for example.

These types of values are used to rank investment performance, such as the performance of mutual funds.

In calculating these values, it is important to first order the data set, as we did with the median. Once this is done, it is necessary to find the position of the value that you are calculating, and then the value itself (the procedure is exactly the same as that for calculating the median).

A **box and whiskers plot** is a convenient way of visually displaying the data distribution through their quartiles. It shows a *five number summary* in a chart.

The main part of the chart (the â€œboxâ€) shows where the middle portion of the data is: the interquartile range. At the ends of the box, you find the first quartile (the 25% mark) and the third quartile (the 75% mark). The bottom of the chart is the minimum (the smallest number in the set) and the top is the maximum (the largest number in the set). Finally, the median is represented by a horizontal bar in the center of the box.

A boxplot tells us:

- Are there any outliers, and what are their values?
- Is the data symmetrical?
- How tightly is the data grouped.
- Is the data skewed and if so, in what direction?

8. Measures of Dispersion

```
    <p> </p><b>Dispersion</b> is defined as "variability around the central tendency."  Ir
```

There are two types of dispersions:

- **Absolute dispersion** is the amount of variability without comparison to any benchmark. Measures of absolute dispersion include range, mean absolute deviation, variance, and standard deviation.
- **Relative dispersion** is the amount of variability in comparison to a benchmark. Measures of relative dispersion include the coefficient of variance.

The Range

The **range** is the simplest measure of spread or dispersion. It is equal to the difference between the largest and the smallest values. The range can be a useful measure of spread because it is so easily understood.

The range is very sensitive to extreme scores because it is based on only two values. It also cannot reveal the shape of the distribution. The range should almost never be used as the only measure of spread, but it can be informative if used as a supplement to other measures of spread, such as the standard deviation or semi-interquartile range.

Example: The range of the numbers 1, 2, 4, 6,12,15,19, 26 = 26 - 1 = 25

The Mean Absolute Deviation

The deviation from the arithmetic mean is the distance between the mean and an observation in the data set. The **mean absolute deviation** (MAD) is the arithmetic average of the absolute deviations around the mean.

In calculating the MAD, we ignore the signs of deviations around the mean. Remember that the sum of all the deviations from the mean is equal to zero. To get around this zeroing-out problem, the mean deviation uses the absolute values of each deviation.

MAD is superior to the range as a measure of dispersion because it uses all the observations in the sample. However, the absolute value is difficult to work with mathematically.

Sample Variance and Sample Standard Deviation

The **variance** is a measure of how spread out a distribution is. It is computed as the average squared deviation of each number from its mean. The formula for the variance in a *population* is:

where:

- Î¼ = the mean
- N = the number of scores

To compute variance in a sample:

This gives an unbiased estimate of Ïƒ^2. Since samples are usually used to estimate parameters, s^2 is the most commonly used measure of variance.

Note for the sample variance, we divide by the sample size minus 1, or N - 1. In the math of statistics, using only N in the denominator when using a sample to represent its population will result in underestimating the population variance, especially for small sample sizes. This systematic understatement causes the sample variance to be a biased estimator of the population variance. By using (N - 1) instead of N in the denominator, we compensate for this underestimation. Thus, using N - 1, the sample variance (s^2) will be an unbiased estimator of the population variance (Ïƒ^2).

The major problem with using the variance is the difficulty interpreting it. Why? The variance, unlike the mean, is in terms of units squared. How does one interpret squared percentages or squared dollars?

The solution to this problem is to use the standard deviation. The formula for the **standard deviation** is very simple: it is the square root of the variance. This is the most commonly used measure of spread.

The variance and the standard deviation are measures of the average deviation from the mean.

An important attribute of the standard deviation as a measure of spread is that if the mean and standard deviation of a normal distribution are known, it is possible to compute the percentile rank associated with any given score. In a normal distribution, about 68% of the scores are within one standard deviation of the mean and about 95% of the scores are within two standards deviations of the mean.

Dispersion and Means

Geometric Mean â‰ˆ Arithmetic Mean - Standard Deviation/2

The more disperse the returns, the larger the gap between the two means.

9. Downside Deviation and Coefficient of Variation

```
<p> </p>Downside risk measures include target deviation, short-fall probability (cover
```

Target Semi-Deviation

Downside risk assumes security distributions are non-normal and non-symmetrical. This is in contrast to what the capital asset pricing model (CAPM) assumes: that security distributions are symmetrical, and thus that downside and upside betas for an asset are the same.

Downside deviation is a modification of the standard deviation such that only variation below a minimum acceptable return is considered. It is a method of measuring the below-mean fluctuations in the returns on investment.

The minimum acceptable return can be chosen to match specific investment objectives.

Semi-deviation will reveal the worst-case performance to be expected from a risky investment.

The semivariance is not used in bond portfolio management extensively because of "ambiguity, poor statistical understanding, and difficulty of forecasting".

Coefficient of Variation

A direct comparison of two or more measures of dispersion may be difficult. For example, the difference between the dispersion for monthly returns on T-bills and the dispersion for a portfolio of small stocks is not meaningful because the means of the distributions are far apart. In order to make a meaningful comparison, we need a relative measure, to standardize the measures of absolute dispersion.

It is often useful to compare the relative variation in data sets that have different means and standard deviations, or that are measured in different units. **Relative dispersion** is the amount of variability present in comparison to a reference point or benchmark. The **coefficient of variation** (CV) is used to standardize the measure of absolute dispersion. It is defined as:

It gives a measure of risk per unit of return, and an idea of the magnitude of variation in percentage terms. It allows us direct comparison of dispersion across data sets. The lower the CV, the better; investments with low CV numbers offer less risk per unit of return. This measurement is also called **relative standard deviation** (RSD).

Note that because s and X-bar have the same units associated with them, the units effectively cancel each other out, leaving a unitless measure which allows for direct comparison of dispersions, regardless of the means of the data sets.

The CV is not an ideal measure of dispersion. What if the expected return is zero!? Generally, the standard deviation is the measure of choice for overall risk (and beta for individual assets).

Example

The mean monthly return on T-bills is 0.25% with a standard deviation of 0.36%. For the S&P 500, the mean is 1.09% with a standard deviation of 7.30%. Calculate the coefficient of variation for T-bills and the S&P 500 and interpret your results.

T-bills: CV = 0.36/0.25 = 1.44

S&P 500: CV = 7.30/1.09 = 6.70

Interpretation: There is less dispersion relative to the mean in the distribution of monthly T-bill returns when compared to the distribution of monthly returns for the S&P 500 (1.44 < 6.70).

10. Interpret Skewness

```
    <p> </p>If a distribution is symmetrical, each side of the distribution is a mirror im
```

A distribution is skewed if one of its tails is longer than the other (that is, it is not symmetrical). A symmetrical distribution has no skewness, (the skewness is zero). Skewness refers to the degree of *asymmetry* of a distribution. It occurs due to the existence of extremely large or small values in the data set. It allows us to see if large positive or negative deviations dominate.

A **positively skewed distribution** means that it has a long tail in the positive direction (a long right tail). It is sometimes called "skewed to the right." This type of distribution is characterized by many small losses and a few extreme gains.

For a positively skewed distribution, the mode is less than the median, which is less than the mean.

Recall that the mean is affected by outliers. In a positively skewed distribution, there are large positive outliers which will tend to "pull" the mean upward. An example of a positively skewed distribution is that of housing prices. Suppose that you live in a neighborhood with 100 homes. Ninety-nine of those homes sell for $600,000 and there is one house that sells for $3,000,000. The median and the mode will be $600,000, but the mean will be $1,300,000. The mean has been "pulled" upward by the existence of one distinctive home in the neighborhood.

A **negatively skewed distribution** has a long tail in the negative direction (a long left tail). It is sometimes called "skewed to the left." It is characterized by many small gains and a few extreme losses.

For the negatively skewed distribution, the mean is less than the median, which is less than the mode. In this case, there are large negative outliers which tend to "pull" the mean downward.

Distributions with positive skew are more common than distributions with negative skew. One example is the distribution of income. Most people make under $60,000 a year, but some make quite a bit more, with a small number making many millions of dollars per year. The positive tail therefore extends out quite a long way, whereas the negative tail stops at zero.

In a more psychological example, a distribution with a positive skew typically results if the time it takes to make a response is measured. The longest response times are usually much longer than typical response times, whereas the shortest response times are seldom much less than typical response times.

Negatively skewed distributions do occur, however.

Tips on how to remember these relative locations:

- The mean is always in the direction of the skew. For example, a positively (negatively) skewed distribution skews to the right (left), so its mean is on the right (left). This is because the mean is unduly influenced by extreme values.
- The median is always in the middle.

a typical exam question: There is a certain probability distribution with the characteristics described below:

- Mean = 100
- Highest possible value = 200
- Lowest possible value = 20

What type of distribution is this?

When a distribution is normal, the dispersion to the left of the mean is the same as the dispersion to the right of the mean. The highest number above (200) is 100 units larger than the mean, whereas the lowest number (20) is 80 units below the mean. Thus, the distribution is not symmetrical. It is skewed to the right.

11. The Shape of the Distributions: Kurtosis

```
<p> </p>We've just discussed skewness, which refers to the deviation from a normal dis
```

Kurtosis is based on the size of a distribution's tails. It is the statistical measure that tells us when a distribution is more or less peaked than a normal distribution.

- Distributions with small tails (that is, less peaked than normal) are called "**platykurtic**." If a return distribution has more returns with large deviations from the mean, it is platykurtic.

- Distributions with relatively large tails (that is, more peaked than normal) are called "**leptokurtic**." If a return distribution has more returns clustered closely around the mean, it is leptokurtic.
- A distribution with the same kurtosis as the normal distribution is called "**mesokurtic**."

Kurtosis is critical in a risk management setting. Most research on the distribution of securities returns has shown that returns are not normal. Actual securities returns tend to exhibit both skewness and kurtosis (sounds like fungus!). Skewness and kurtosis are critical for risk management because if securities returns are modelled after a normal distribution, predictions from those models will take into consideration the potential for extremely large negative outcomes. In fact, most risk managers put very little emphasis on the mean and standard deviations of a distribution and focus more on the distribution of returns in the tails of the distribution, since this is where the risk is.

To calculate skewness (when N is large):

where Î¼ is the mean and Ïƒ is the standard deviation.

The normal distribution has a skew of 0, since it is a symmetric distribution.

If skewness is positive, the average magnitude of positive deviations is larger than the average magnitude of negative deviations.

To calculate **excess kurtosis** (when N is large):

The kurtosis equals excess kurtosis + 3.

For all normal distributions, kurtosis is equal to 3; excess kurtosis is equal to 0.

A leptokurtic distribution has an excess kurtosis greater than 0, and a platykurtic distribution has an excess kurtosis less than 0.

12. Correlation Between Two Variables

```
<p> </p>Variance and standard deviation measure the dispersion of a single random vari
```

Given two random variables, R_i and R_j, the **covariance** between the two variables is:

Facts about covariance:

- Covariance of returns is negative if, when the return on one asset is above its expected value, the return on the other asset is below its expected value (an average inverse relationship between returns).
- Covariance of returns is 0 if returns on the assets are unrelated.
- Covariance of returns is positive if, when the return on one asset is above its expected value, the return on the other asset is above its expected value (an average positive relationship between returns).
- The covariance of a random variable with itself (own covariance) is its own variance.

Example

Suppose that the future short-term outlook for the economy is favorable with a probability 0.6 and unfavorable with a probability of 0.4. For two stocks, F and G, returns are 0.25 and 0.2, respectively, in favorable conditions, and 0.01 and 0.02, in unfavorable conditions. Calculate cov (R_f, R_g).

We must firstly calculate the expected value of the return on each stock:

$E[R_f]$ = 0.6 x 0.25 + 0.4 x 0.01 = 0.154

E[R_g] = 0.6 x 0.2 + 0.4 x 0.02 = 0.128

cov (R_f, R_g) = E[{R_f - E(R_f)} x {R_g - E(R_g)}] = 0.6 x [{0.25 - 0.154}x {0.2 - 0.128}] + 0.4 x [{0.01 - 0.154}x {0.02-0.128}] = 0.010368

The fact that the answer is positive indicates that the return on both stocks is above (or below) the expected value at the same time. We know that this is the case because both returns are higher in favorable conditions and lower in unfavorable conditions. Had we obtained a negative answer, logic would have told us that we had made an error somewhere.

The **correlation** between two random variables, R_i and R_j, is defined as:

Alternative notations are corr(R_i, R_j) and Ï$_{ij}$.

Properties of correlation:

- Correlation is a number between -1 and +1.
- A correlation of 0 indicates an absence of any linear (straight-line) relationship between the variables.
- Increasingly positive correlation indicates an increasingly strong positive linear relationship (up to 1, which indicates a perfect linear relationship).
- Increasingly negative correlation indicates an increasingly strong negative linear relationship (down to -1, which indicates a perfect inverse linear relationship).

The correlation between two variables represents the degree to which these variables are related. It is important to keep in mind that correlation does not necessarily mean causation. For example, there is a high positive relationship between the number of fire-fighters sent to a fire and the amount of damage done. Does this mean that the fire fighters cause the damage? Or is it more likely that the bigger the fire, the more fire fighters are sent and the more damage is done? In this example, the variable "size of the fire" is the causal variable, correlating with both the number of fire-fighters sent and the amount of damage done.

The relationship among covariance, standard deviation, and correlation:

Using the figures from the previous example, we first need to calculate the two standard deviation terms:

Var(R_f) =[{R_f - E(R_f)}^{2}] x P(R_f) = 0.6 x [{0.25-0.154}^{2}] + 0.4 x [{0.01-0.154}^{2}] = 0.013824. Hence, Ïƒ(R_f) = 0.117576

Var(R_g) = [{R_g - E(R_g)}^{2}] x P(R_g) = 0.6 x [{0.2-0.128}^{2}] + 0.4 x [{0.02-0.128}^{2}] = 0.007776. Hence, Ïƒ(R_g) = 0.088182

Also, we know that cov(R_f,R_g) = 0.010368.

So, correlation = Ï(R_f, R_g)= cov(R_f, R_g) / Ïƒ(R_f) x Ïƒ(R_g) = 0.010368 / (0.117576 x 0.088182) = 0.99999.

This indicates an almost perfect positive linear relationship between R_f and R_g.

Calculate covariance given a joint probability function:

We can calculate covariance using the joint probability function of the random variables if that can be estimated. The joint probability function of two random variables, X and Y, denoted P(X, Y), gives the probability of joint occurrences of values X and Y. For example, P(3, 2) is the probability that X equals 3 and Y equals 2.

Probability Concepts

1. Introduction

```
   An <b>experiment</b> is the act of making an observation or taking a measurement. For
```

A particular result of an experiment is called an **outcome**. It is the value assigned to a random variable. For example, there are six possible outcomes to the experiment of tossing dice.

A **random variable** is a quantity whose future outcomes are uncertain. For example, when you toss dice, the number on top is a random variable; you are unsure which number will come up.

Any outcome or specified set of outcomes of a random variable is called an **event**. For example, one event in the dice-tossing experiment is observing an odd number (three possible outcomes: 1, 3, 5).

Mutually exclusive events are events that only occurs one at a time. In the above example, event A (observe an odd number) and event B (observe an even number) are mutually exclusive because they cannot occur simultaneously; a number can be either odd or even, but cannot be both.

Exhaustive events are events that cover all the distinct possible outcomes. In the above example, event A and event B are exhaustive because they cover all six possible outcomes of the dice-tossing experiment.

As a general rule (and it is one you can use with confidence), if probabilities of two or more events add up to 1 and the events share no common outcomes at all, those events will always be mutually exclusive and exhaustive. However, it is important to remember that events can be mutually exclusive and exhaustive (as you've just seen), or can be one but not the other, or can be neither mutually exclusive nor exhaustive.

The two defining properties of probability

In general, probability is the likelihood that an event will happen. P(E) stands for "the probability of event E." The two defining properties of probability are:

- 0 â‰¤ P(E) â‰¤1: the probability of any event E is a number between 0 and 1. The probability of 0 means that the event can never happen and the probability of 1 means that the event is certain to happen.
- The sum of the probabilities of any list of mutually exclusive and exhaustive events equals 1. For example, the sum of the possibilities of two events (observing an odd number event and observing an even number event in the dice-tossing example) is 1, as these events are mutually exclusive and exhaustive.

Empirical probability

Empirical probability is a probability based on relative frequency of occurrence. It is estimated on the basis of historical data. For example, based on historical data over a 10-year period, the probability of default for real estate mortgage loans is 7%. We cannot estimate the empirical probability for an event without historical data. For empirical probabilities to be accurate, relationships must be stable over time.

Priori probability

Priori probability is a probability based on logical analysis rather than observation or personal judgment. For example, when you toss dice fairly, the probability of rolling an even number is 50%.

Empirical and priori probabilities generally do not vary from person to person, and they are often grouped as objective probabilities.

Subjective probability

Subjective probability is a probability based on personal or subjective judgment. For example, based on his own judgment, Bill believes that the probability that IBM's revenue will increase in 2005 is 60%.

According to the **Dutch Book Theorem**, one of the most important probability results theories for investments, inconsistent probabilities create profit opportunities. Investors should eliminate the profit opportunity and inconsistency through buy and sell decisions exploiting inconsistent probabilities.

Suppose that:

- If event E occurs, the values of two assets, A and B, will both rise.
- The price of asset A reflects a higher probability of event E than the price of asset B and thus inconsistent probabilities exist.

If all other factors are equal, asset A is overvalued compared with asset B.

- If event E does occur, the price of asset A will not rise as much as the price of asset B. This is because the occurrence of event E is mostly incorporated into the price of asset A.
- If event E does not occur, the prices of both assets will fall, but the price of asset A will decline more than the price of asset B. This is because, compared with asset B, the price of asset A understates the probability that event E may not occur.

Therefore, investors can profit by buying an undervalued asset (i.e., B) and selling an overvalued asset (i.e., A). Conservative investors will buy asset B and reduce or fully liquidate their position in asset A. Aggressive investors will buy asset B and short asset A. This strategy is known as the **pairs arbitrage trade**, which involves using the proceeds from the short sale of one stock to purchase another.

Note that the above discussion is based on the assumption that the occurrence of event E will increase the values of the two assets (A and B). If the occurrence of event E will reduce the value of assets A and B, asset B is overvalued if compared with asset A. To profit from inconsistent probabilities, investors should buy asset A and sell asset B.

Example

Suppose that if a hike in oil price occurs, the stock prices of American Airlines (AA) and British Airways (BA) will decline. The stock price of AA reflects a 0.7 probability of a hike in oil price, whereas the stock price of BA reflects only a 0.4 probability. In this situation, the stock of AA is undervalued when compared with the stock of BA. A conservative investor can profit by buying the stock of AA and reducing or eliminating his holdings in the stock of BA. An aggressive investor can profit by buying the stock of AA and shorting the stock of BA.

2. Unconditional, Conditional, and Joint Probabilities

```
   The <b>complement of event A</b> is the event that A does <i>not</i> occur. It is expr
```

Probabilities are either unconditional or conditional.

Unconditional probability, also called **marginal probability**, is simply the probability of an event occurring. It refers to the probability of an event that is not conditioned on the occurrence of another event. For example, what is the probability that a stock earns a return above the risk-free rate? An unconditional probability can be considered as a stand-alone probability. It is expressed as P(A).

A **conditional probability** is the probability of an event given that another event has occurred. It is denoted as P(A | B) ("the probability of A given B").

For example, what is the probability that the total of two dice will be greater than 8 if the first dice is a 6? This can be computed by considering only outcomes which could occur if the first dice is a 6 and determining the proportion of these outcomes that total more than 8. There are six outcomes for which the first dice is a 6, and of these, there are four that total more than 8 (6,3; 6,4; 6,5; 6,6). The probability of a total greater than 8, given that the first dice is 6, is therefore 4/6 = 2/3. More formally, this probability can be written as: P(total>8 | Dice 1 = 6) = 2/3. In this equation, the expression to the left of the vertical bar represents the event and the expression to the right of the vertical bar represents the condition. Thus, it would be read as "The probability that the total is greater than 8, given that Dice 1 is 6, is 2/3." In more abstract form, P(A|B) is the probability of event A given that event B occurred.

A **joint probability** is the probability of both events A and B *happening together*. It is denoted as P(AB) ("the probability of A and B"). For example, Kevin is assessing the probability that both airfare and oil prices increase. Such a probability is a joint probability.

If two events are mutually exclusive, then they cannot occur together, so the joint probability of two mutually exclusive events is 0.

A **joint probability function** of two random variables, X and Y, gives the probability of joint occurrences of the values of X and Y.

If we know the conditional probability P(A|B) and we want to know the joint probability P(AB), we can use the following **multiplication rule for probabilities**:

$$P(AB) = P(A|B) \times P(B)$$

Example 1

If someone draws a card at random from a deck and then, without replacing the first card, draws a second card, what is the probability that both cards will be aces? Event A is that the first card is an ace. Since 4 of the 52 cards are aces, p(A) = 4/52 = 1/13. Given that the first card is an ace, what is the probability that the second card will be an ace as well? Of the 51 remaining cards, 3 are aces. Therefore, p(B|A) = 3/51 = 1/17 and the probability of A and B is: 1/13 x 1/17 = 1/221.

Example 2

The probability of an increase in oil price, P(B), is 0.4. The probability of an increase in airfare given an increase in oil price, P(A|B), is 0.3. The joint probability of an increase in both oil price and airfare, P(AB), is 0.3 x 0.4 = 0.12.

Hint:

- Look out for the words "given that" or "you are told that," which will help you know that the probability is conditional. In the absence of such information, the probability will be unconditional.
- The letter after the | is the event that we know has definitely occurred, whereas the letter before the | is the event whose probability we are trying to calculate.

3. Addition Rule for Probabilities: the Probability that at Least One of Two Events Will Occur

```
   If we have two events, A and B, that we are interested in, we often want to know the p
```

Such probabilities are calculated using the **addition rule for probabilities**.

$$P(A \text{ or } B) = P(A) + P(B) - P(AB)$$

The logic behind this formula is that when P(A) and P(B) are added, the occasions on which A and B both occur are counted twice. To adjust for this, P(AB) is subtracted.

If events A and B are mutually exclusive, the joint probability of A and B is 0. Consequently, the probability that either A or B occurs is simply the sum of the unconditional probabilities of A and B: P (A or B) = P(A) + P(B).

What is the probability that a card selected from a deck will be either an ace or a spade? The relevant probabilities are:

P(ace) = 4/52; P(spade) = 13/52

The only way in which an ace and a spade can both be drawn is to draw the ace of spades. There is only one ace of spades, so:

P(ace and spade) = 1/52.

The probability of an ace or a spade can be computed as:

P(ace) + P(spade) - P(ace and spade) = 4/52 + 13/52 - 1/52 = 16/52 = 4/13.

Consider the probability of rolling dice twice and getting a 6 on at least one of the rolls. The events are defined in the following way:

Event A: 6 on the first roll: p(A) = 1/6

Event B: 6 on the second roll: p(B) = 1/6

P(A and B) = 1/6 x 1/6

P(A or B) = 1/6 + 1/6 - 1/6 x 1/6 = 11/36

The same answer can be computed using the following (admittedly convoluted) approach: Getting a 6 on either roll is the same thing as not getting a number from 1 to 5 on both rolls. This is equal to: 1 - P(1 to 5 on both rolls).

The probability of getting a number from 1 to 5 on the first roll is 5/6. Likewise, the probability of getting a number from 1 to 5 on the second roll is 5/6. Therefore, the probability of getting a number from 1 to 5 on both rolls is: 5/6 x 5/6 = 25/36. This means that the probability of not getting a 1 to 5 on both rolls (getting a 6 on at least one roll) is: 1-25/36 = 11/36.

Despite the convoluted nature of this method, it has the advantage of being easy to generalize to three or more events. For example, the probability of rolling dice three times and getting a six on at least one of the three rolls is: 1 - 5/6 x 5/6 x 5/6 = 0.421

4. Multiplication Rule for Independent Events

```
    Two events, A and B, are <b>independent</b> if and only if P(A|B) = P(A), or equivalen
```

In more detail, whether or not B occurs will have no effect on the probability of A and vice versa. Thus, there will be no difference between P(A|B) and P(A), and similarly there will be no difference between P(B|A) and P(B).

For example, suppose you flip a coin twice. The event of getting heads on the first flip does not affect the probability of getting heads on the second flip. Therefore, the event of getting heads on the second flip is independent of the event of getting heads on the first flip.

When two events are not independent, they must be **dependent**: the occurrence of one is related to the probability of the occurrence of the other.

If we are trying to forecast one event, information about a dependent event will be useful but information about an independent event will not be useful.

Example 1

If C = {the price of insurance share C goes up} and D = {the price of insurance share D goes up}, then clearly

C and D are dependent events, because the market as a whole might be bullish or the insurance sector alone might be having a good day. Although the increase in share price C might not affect share price D at all, there is clearly a good chance that the two shares will move in the same direction.

Example 2

If A = {a person in Europe drives a red car} and B = {a person in Australia drives a white car}, then events A and B are clearly independent, as the one almost certainly has no bearing upon the other.

Remember that if A and B are independent, Ac and Bc will also be independent.

A and B are two events. If A and B are independent, the probability that events A and B both occur is:

$$P(A \text{ and } B) = P(A) \text{ x } P(B)$$

In other words, the probability of A and B both occurring is the product of the probability of A and the probability of B. This relationship is known as the **multiplication rule for independent events**.

What is the probability that a fair coin will come up with heads twice in a row? Two events must occur: heads on the first toss and heads on the second toss. Since the probability of each event is 1/2, the probability of both events is: 1/2 x 1/2 = 1/4.

Now consider a similar problem: Someone draws a card at random out of a deck, replaces it, and then draws another card at random. What is the probability that the first card is the ace of clubs and the second card is a club (any club)? Since there is only one ace of clubs in the deck, the probability of the first event is 1/52. Since 13/52 = 1/4 of the deck is composed of clubs, the probability of the second event is 1/4. Therefore, the probability of both events is: 1/52 x 1/4 = 1/208.

Similarly, for any number of independent events E_1, E_2.....E_n, the probability that all of them occur is:

$$P(E_1 \text{ and } E_2\text{..... and } E_n) = P(E_1) \text{ x } P(E_2) \text{ x ..... x } P(E_n)$$

Example

In a bullish market, three shares, chosen from different sectors of the market, have probabilities of 0.6, 0.5 and 0.8 that their share prices will rise on any particular day. Let's call the shares K, L and M respectively.

If we make the assumption that prices of shares on successive days are independent, and also that the price movement of one of the shares above is independent of the others, we can then carry out the following calculations.

Note that although this may seem slightly unrealistic, the fact that the shares come from different sectors of the market lends some credence to the assumption, and also simplifies our calculations considerably.

P (share K has a price rise on two consecutive days) = 0.6 x 0.6 = 0.36

P (shares L and M both rise in price on the same day) = 0.5 x 0.8 = 0.4

Thus, the independent assumptions make our work much easier.

To calculate the probability that share M rises in price on four consecutive days, we can use the results above to calculate this probability as: 0.8 x 0.8 x 0.8 x 0.8 = 0.4096 (i.e., 0.8 multiplied four times, once for each day).

Although the chance that the share will rise in price on any one day is 80%, the chance that this will happen for four days in a row is just over half of this total, or 40.96%.

If we wish to calculate the probability that all three shares will rise in price on the same day, we can use the results above to get: 0.6 x 0.5 x 0.8 = 0.24 (i.e., the individual probabilities multiplied together)

Warning: It is important to note that multiplying individual probabilities together can only be done if the events that make up those probabilities are independent. If the events are dependent, this process is not valid.

To calculate the probability that, of the three shares above, none will have a price rise on a particular day, we can multiply the probabilities of the complementary events together to get: 0.4 x 0.5 x 0.2 = 0.04.

Thus, there is only a 4% chance that no shares will have a price rise on a particular day.

You might be wondering, if the chance that all three shares rise in price is 24% and the chance that no shares rise in price is 4%, about what has happened to the remaining 100% - 24% - 4% = 72%. The answer is that this percentage covers situations where some shares rise in price while others don't. You'll see this type of example in more detail in later lessons.

5. The Total Probability Rule

```
    If we have an event or scenario S, the event not-S, called the <b>complement</b> of S,
```

The **total probability rule** explains the unconditional probability of an event in terms of probabilities conditional on the scenarios.

- $P(A) = P(A|S)P(S) + P(A|S^C)P(S^C)$
- $P(A) = P(A|S_1)P(S_1) + P(A|S_2)P(S_2) + ... + P(A|S_n)P(S_n)$

- The first equation is just a special case of the second equation.
- The second equation states the following: the probability of any event [P(A)] can be expressed as a weighted average of the probabilities of the event, given scenarios [terms such as $P(A|S_1)$]; the weights applied to these conditional probabilities are the respective probabilities of the scenarios [terms such as $P(A_1$ multiplying $P(A|S_1)$], and the scenarios must be mutually exclusive and exhaustive.

Suppose there are two events:

- Event A: IBM's revenue will increase.
- Event B: the economy is going into an expansion. P(B) = 0.6, and therefore $P(B^c) = 0.4$.

The probability of an increase in IBM's revenue given an economic expansion is P(A|B) = 0.8.

The probability of an increase in IBM's revenue given no economic expansion is $P(A|B^c) = 0.7$.

Using the total probability rule, we can compute the probability of an increase in IBM's revenue: $P(A) = P(A|B) \times P(B) + P(A|B^c) \times P(B^c) = 0.8 \times 0.6 + 0.7 \times 0.4 = 0.76$.

Typical exam question

An analyst constructs the following probability table for the market and Company X's stock:

1. Compute the total probability of good performance for Company X's stock.

Here we are asked to find the total probability of good performance. This means we have to find the joint probability of one stock outcome. To do this we multiply and add.

We take Î£ (probability of economic state x good conditional probability):

Joint probability = (0.5 x 0.4) + (0.3 x 0.5) + (0.2 x 0.5) = 45%

1. Compute the probability of simultaneously realizing a bull economy and poor stock performance for Company X.

This question asks you to determine the probability of a specific branch. If we follow the branch and multiply the probabilities, we will arrive at the correct answer as follows.

Bull economy: 0.5

Poor stock: 0.3

Probability = 0.5 x 0.3 = 0.15 = 15%

6. Expected Value (Mean), Variance, and Conditional Measures of Expected Value and Variance

```
   The <b>expected value</b> of a random variable is its probability-weighted average of
```

For a random variable X, the expected value of X is denoted E(X).

$$E(X) = P(x_1)\ x_1 + P(x_2)\ x_2 + ... + P(x_n)\ x_n$$

In investment analysis, forecasts are frequently made using expected value, for example, the expected value of earnings per share, dividend per share, rate of return, etc. It represents the *central value* of all possible outcomes.

Example

The organizers of an outdoor event know that the success of the event depends on the weather. It costs $50,000 to stage the event. If the weather is favorable, the organizers will take in $200,000. If the weather is moderate, the organizers will take in $80,000. If the weather is unfavorable, the organizers will be forced to abandon the event, and thus take in $0. The weather bureau forecasts that the chances of favorable, moderate and unfavorable weather are 20%, 30% and 50% respectively. Should the organizers go ahead and stage the event?

We can use expected value to work out what revenue the organizers can expect to generate. Once we have this number, we can compare it with the cost of the event, $50,000, to assess whether the venture is likely to be profitable.

Using the expected value formula, we will multiply each amount by its probability, and add the answers. E(X) = 200,000 x 0.2 + 80,000 x 0.3 + 0 x 0.5 = 40,000 + 24,000 + 0 = $64,000

Thus, the organizers can expect to take in $64,000. Since it costs $50,000 to stage the event, this translates to a profit of $14,000, so they should certainly go ahead with the venture.

It's important to realize that none of the outcomes actually produces an amount of $64,000. This is simply the weighted average of all possible outcomes. Although there is a 50% chance of a loss the big profit that will be made the remaining 50% of the time more than offsets this and creates an overall expected profit.

However, with a one-off concert, there is a major risk involved, particularly in the event of unfavorable weather. An easier way to interpret expected value is as follows: If a number of such concerts were held, the organizers can expect to achieve a profit of $14,000 for each concert. So expected values actually make more sense when viewed over the long run.

The **variance** of a random variable is the expected value (the probability-weighted average) of squared deviations from the random variable's expected value.

$$\text{Ï}\mathit{ƒ}^2(X) = E\{[X - E(X)]^2\}$$

Variance is a number greater than or equal to 0.

- If it is 0, there is no dispersion or risk. The outcome is certain.
- Variance greater than 0 indicates dispersion of outcomes.
- Increasing variance indicates increasing dispersion, if all other factors are equal.

- Variance of X is a quantity in the squared units of X; it is difficult to interpret this variance.

The **standard deviation** is the positive square root of variance.

Variance and standard deviation measure the dispersion of possible outcomes around the expected value of the random variable. If all other factors are equal, increasing variance or standard deviation indicates increasing dispersion of the possible outcomes.

In the example above, we calculated the expected value of revenue to be $64,000. This was before we subtracted the costs. To calculate the variance of the organizers' revenue, we simply take each value, subtract 64,000, square the answer, multiply by the relevant probability in each case, and add.

Var (X) = $[200{,}000 - 64{,}000]^2$ x 0.2 + $[80{,}000 - 64{,}000]^2$ x 0.3 + $[0 - 64{,}000]^2$ x 0.5 = 5824000000

The standard deviation is the square root of this number. So, SD(X) = 76,315.13611.

These numbers are often large, particularly if your original data comprises large numbers, as is the case here. Because the calculations for variance and standard deviation yield big numbers, we can conclude that the values in the data set are extremely variable and scattered fairly far away from the expected value.

Parallel to the total probability rule for stating unconditional probabilities in terms of conditional probabilities, **total probability rule for expected value** states (unconditional) expected values in terms of conditional expected values.

- $E(X) = E(X|S)P(S) + E(X|S^C)P(S^C)$
- $E(X) = E(X|S_1)P(S_1) + E(X|S_2)P(S_2) + ... + E(X|S_n)P(S_n)$
  (where S_1, S_2, ..., S_n are mutually exclusive and exhaustive scenarios or events.)

The general case, equation 2, states that the expected value of X equals the expected value of X given Scenario 1, $E(X|S_1)$, times the probability of Scenario 1, $P(S_1)$, plus the expected value of X given Scenario 2, $E(X|S_2)$, times the probability of Scenario 2, $P(S_2)$, and so on.

In investments, we make use of any relevant information available in making our forecast. When we refine our expectations or forecasts, we are typically making adjustments based on new information or events; in these cases we are using **conditional expected values**. The expected value of a random variable X given an event or scenario S is denoted E(X|S).

Relating the formula to the example above and using the following notation:

X = revenue, F = favorable weather, M = moderate weather, U = unfavorable weather, the formula becomes:

E(X) = E (X|F) x P(F) + E (X|M) x P(M) + E(X|U) x P(U)

Note that the right-hand side has three terms because there are three possible weather scenarios.

The E terms on the right are calculated as follows:

E (X|F) = Expected value (Revenue | Favorable weather) = 200,000, because if the weather is favorable, the revenue will be $200,000.

Similarly, E (X|M) = 80,000 and E(X|U) = 0.

So, E(X) = 200,000 x 0.2 + 80,000 x 0.3 + 0 x 0.5 = 40,000 + 24,000 + 0 = 64,000.

This is the same answer that we calculated before; the formula above is just another way of carrying out the same calculation.

Note that had there been ten different weather scenarios, the right-hand side would contain ten different terms.

The key information is that the different weather scenarios are both mutually exclusive and exhaustive.

Probability Tree

Probability trees are useful for calculating combined probabilities for sequences of events. It helps you to map out the probabilities of many possibilities graphically, without the use of complicated probability formulas.

A probability tree has two main parts: the branches and the ends(sometimes called leaves). The probability of each branch is generally written on the branches, while the outcome is written on the ends of the branches. In general you multiply along the branches and add probabilities down the columns (up to 1).

Example

Suppose a farmer must decide what to do with his land for the next growing season. He can choose to plant corn or soybeans or to not plant anything at all. If he plants nothing at all, the government farm subsidy will pay him \$30 per acre.

If the farmer decides to plant corn or soybeans on his land, there is some risk involved. The yield per acre depends on the amount of rainfall. Too much rain or too little rain will give poorer results than the right amount of rainfall. There is a 40 percent probability that the rainfall will be low; there is a 40 percent probability that the rainfall will be medium; and there is a 20 percent chance that the rainfall will be high.

If the farmer decides to plant corn, the yield per acre will be \$0, \$90, and \$50, respectively, if the rainfall is low, medium, or high. If the farmer decides to plant soybeans, the yield per acre will be \$40, \$70, and \$20, respectively, for low, medium, and high amounts of rainfall.

As shown in the figure, the decision to be made is whether the farmer should plant corn, soybeans, or nothing at all. There are three lines coming out of the decision box to indicate the three choices. Each choice leads to a probabilistic occurrence - how much rainfall will occur.

Each probabilistic occurrence has three possible outcomes - low, medium, or high amounts of rainfall. For each of these events there is an associated payoff. The payoff amount multiplied by the probability of that event occurring is the expected value of each occurrence.

In order to evaluate the decisions, we must add the expected value of each event associated with each decision to get the expected value for each decision. For corn, low rainfall means that no money will be made from the crop. For medium rainfall there is a 40 percent chance and a \$90 yield, giving an expected value of \$36. For high rainfall there is not as much yield per acre at \$50 and there is a 20 percent probability of that occurring. The expected value for high rainfall is thus \$10 per acre. Adding the expected values for the events gives us the expected value for the decision. This is \$46 per acre.

Using the same calculation for the soybeans and for not planting at all, we see that of the three decisions, planting soybeans has the greatest yield.

7. Expected Value, Variance, Standard Deviation, Covariances, and Correlations of Portfolio Returns

```
   Variance and standard deviation measure the dispersion of a single random variable. Of
```

Given two random variables, R_i and R_j, the **covariance** between the two variables is:

Facts about covariance:

- Covariance of returns is negative if, when the return on one asset is above its expected value, the return on the other asset is below its expected value (an average inverse relationship between returns).
- Covariance of returns is 0 if returns on the assets are unrelated.

- Covariance of returns is positive if, when the return on one asset is above its expected value, the return on the other asset is above its expected value (an average positive relationship between returns).
- The covariance of a random variable with itself (own covariance) is its own variance.

Example

Suppose that the future short-term outlook for the economy is favorable with a probability 0.6 and unfavorable with a probability of 0.4. For two stocks, F and G, returns are 0.25 and 0.2, respectively, in favorable conditions, and 0.01 and 0.02, in unfavorable conditions. Calculate cov (R_f, R_g).

We must firstly calculate the expected value of the return on each stock:

E[R_f] = 0.6 x 0.25 + 0.4 x 0.01 = 0.154

E[R_g] = 0.6 x 0.2 + 0.4 x 0.02 = 0.128

cov (R_f, R_g) = E[{R_f - E(R_f)} x {R_g - E(R_g)}] = 0.6 x [{0.25 - 0.154}x {0.2 - 0.128}] + 0.4 x [{0.01 - 0.154}x {0.02-0.128}] = 0.010368

The fact that the answer is positive indicates that the return on both stocks is above (or below) the expected value at the same time. We know that this is the case because both returns are higher in favorable conditions and lower in unfavorable conditions. Had we obtained a negative answer, logic would have told us that we had made an error somewhere.

The **correlation** between two random variables, R_i and R_j, is defined as:

Alternative notations are corr(R_i, R_j) and Ï$_{ij}$.

Properties of correlation:

- Correlation is a number between -1 and +1.
- A correlation of 0 indicates an absence of any linear (straight-line) relationship between the variables.
- Increasingly positive correlation indicates an increasingly strong positive linear relationship (up to 1, which indicates a perfect linear relationship).
- Increasingly negative correlation indicates an increasingly strong negative linear relationship (down to -1, which indicates a perfect inverse linear relationship).

The correlation between two variables represents the degree to which these variables are related. It is important to keep in mind that correlation does not necessarily mean causation. For example, there is a high positive relationship between the number of fire-fighters sent to a fire and the amount of damage done. Does this mean that the fire fighters cause the damage? Or is it more likely that the bigger the fire, the more fire fighters are sent and the more damage is done? In this example, the variable "size of the fire" is the causal variable, correlating with both the number of fire-fighters sent and the amount of damage done.

The relationship among covariance, standard deviation, and correlation:

Using the figures from the previous example, we first need to calculate the two standard deviation terms:

Var(R_f) =[{R_f - E(R_f)}^{2}] x P(R_f) = 0.6 x [{0.25-0.154}^{2}] + 0.4 x [{0.01-0.154}^{2}] = 0.013824. Hence, Ï*ƒ*(R_f) = 0.117576

Var(R_g) = [{R_g - E(R_g)}^{2}] x P(R_g) = 0.6 x [{0.2-0.128}^{2}] + 0.4 x [{0.02-0.128}^{2}] = 0.007776. Hence, Ï*ƒ*(R_g) = 0.088182

Also, we know that cov(R_f,R_g) = 0.010368.

So, correlation = Ï(R_f, R_g)= cov(R_f, R_g) / Ïƒ(R_f) x Ïƒ(R_g) = 0.010368 / (0.117576 x 0.088182) = 0.99999.

This indicates an almost perfect positive linear relationship between R_f and R_g.

Portfolio Expected Return

The expected return on a portfolio of assets is the market-weighted average of the expected returns on the individual assets in the portfolio. The variance of a portfolio's return consists of two components: the weighted average of the variance for individual assets and the weighted covariance between pairs of individual assets.

Ïƒ^2(R_p) = w_1^2Ïƒ^2(R_1) + w_2^2Ïƒ^2(R_2) + 2w_1w_2Cov(R_1, R_2)

You have a portfolio of two mutual funds, A and B, with 75% invested in A.

E(R_A) = 20%; E(R_B) = 12%.

Covariance Matrix:

The values on the main diagonal are the variances and the other values are the covariances.

The expected return on the portfolio is:

E(R_p) = w_A E(R_A) + (1 - w_A) E(R_B) = 0.75 x 20% + 0.25 x 12% = 18%

The correlation matrix:

Ïƒ(R_A) = $(625)^{1/2}$ = 25, Ïƒ (R_B) = $(196)^{1/2}$ = 14

Ï(R_A, R_B) = Cov(R_A, R_B) / [Ïƒ(R_A) x Ïƒ(R_B)] = 120 / (25 x 14) = 0.342857, or 0.34

The variance of the portfolio is:

Ïƒ^2(R_P)

= w_A^2Ïƒ^2(R_A) + w_B^2Ïƒ^2(R_B) + 2w_Aw_BCov(R_A, R_B)

= $(0.75)^2(625) + (0.25)^2(196) + 2(0.75)(0.25)(120)$

= 408.8125

The standard deviation is Ïƒ(R_P) = $(408.8125)^{1/2}$ = 20.22%.

It's also possible that you could be given a correlation matrix, which is simply a matrix that shows the correlation between any two assets in the portfolio. Consider the following correlation matrix for assets A, B and C.

Note that the matrix is symmetrical about its main diagonal (top left to bottom right). The entries on this diagonal are all 1, as the correlation between any variable and itself is obviously 1. Similarly, the correlation between RA and RB is 0.53, the correlation between RA and RC is 0.78, and the correlation between RB and RC is 0.6.

The steps that would now be involved would be:

- Calculate expected values and variances for the return on each asset.

- Square-root your variances in each case to get standard deviations.
- Use the standard deviations together with the correlations from the matrix above to calculate covariances using the link formula.
- Calculate the values of the portfolio weights.
- Now calculate E(Rp) and Var(Rp) using the above formula.

Essentially, the processes are the same. In each case, we need to obtain expected values, variances and covariances, in order to calculate E(Rp) and Var(Rp). How we obtain them depends on how the data are presented to us.

Familiarize yourself with the two different types of matrices, as explained in this section, and know what each term represents in each covariance formula.

8. Covariance Given a Joint Probability Function

```
<p> </p>We can calculate the covariance between two asset returns given the joint prob
```

Suppose we wish to find the variance of each asset and the covariance between the returns of A and B, given that the amount invested in each company is $1,000.

This table is used to calculate the expected returns:

For us to find the covariance, we must calculate the expected return of each asset as well as their variances. The assets weights are: W_A = 1000/2000 = 0.5 and W_B = 1000/2000 = 0.5

Next, we should calculate the individual expected returns:

$E(R_A) = 0.15 \text{ x } 0.40 + 0.60 \text{ x } 0.2 + 0.25 \text{ x } 0.00 = 0.18$

$E(R_B) = 0.15 \text{ x } 0.2 + 0.60 \text{ x } 0.15 + 0.25 \text{ x } 0.04 = 0.13$

Finally, we can compute the covariance between the returns of the two assets:

$Cov(R_{A,\ B}) = 0.15\ (0.40 - 0.18)\ (0.20 - 0.13) + 0.6\ (0.20 - 0.18)\ (0.15 - 0.13) + 0.25\ (0.00 - 0.18)\ (0.04 - 0.13) = 0.0066$

Interpretation: Since covariance is positive, the two returns show some co-movement, though it's a weak one.

9. Bayes' Formula

```
Bayes' formula ties in nicely with the total probability rule. This rule is:<p> </p>
```

$$E(X) = E(X|S)P(S) + E(X|S^C)P(S^C)$$

Thus, with the total probability rule, we calculate an unconditional probability (in this case, the probability of event X) by using the fact that we know conditional probabilities of X given other events (S and S^C) and their respective unconditional probabilities.

Bayes' formula is used when we know that the event whose probability we have just calculated (i.e., event X) has occurred, and we wish to evaluate conditional probabilities based on this fact.

Thus, Bayes' formula allows us to calculate P(S|X) and $P(S^C|X)$, (i.e., conditional probabilities) another way from how they appear in the above formula.

Effectively, we are using Bayes' formula to update our knowledge of a specific event occurring in light of new information received (an event has occurred).

The general formula is:

Updated probability = (probability of the new information given event / unconditional probability of the new information) x prior probability of event.

Typical exam question

An analyst has developed a ratio to identify a company's expectation of experiencing declining PE multiples over time. Research shows that 55% of firms with declining PEs have a negative ratio, while only 25% of firms not experiencing a decline in PEs have a negative ratio. The analyst expects that 15% of all publicly traded companies will experience a decline in PE next year. The analyst randomly selects a company and its ratio is negative. Based on Bayes' theorem, compute the probability that the company will experience a PE decline next year.

This question deals with applying Bayes' theorem to determine the probability that a company will experience a decline in PE.

We need to define the notation to begin with.

X_1: PE will decline. $P(_1) = 0.15$

X_2: PE will not decline. $P(X_1) = 0.85$

$P(B|X_1) = 0.55$: ratio is negative when PE declines

$P(B|X_2) = 0.25$: ratio is negative when PE does not decline

We are looking for the probability that PE will decline given a negative ratio.

Bayes' theorem can be applied as follows:

$P(X_1|B) = [P(X_1) \times P(B|X_1)]$ / Unconditional probability of the new information

$= [P(X_1) \times P(B|X_1)] / [P(X_1) \times P(B| X_1) + P(X_2) \times P(B| X_2)]$

$= [0.15 \times 0.55] / [(0.15 \times 0.55 + 0.85 \times 0.25] = 28\%$

10. Principles of Counting

In some cases, it's relatively easy to list and count all possible outcomes. For examp

If one thing can be done in n_1 ways, and a second thing, given the first, can be done in n_2 ways, and so on for k things, then the number of ways the k things can be done is $n_1 \times n_2 \times n_3 \ldots \times n_k$.

For example, suppose a portfolio manager is making two decisions:

- Which type of instrument to invest in: stocks or bonds?
- Which country to invest in: U.S., Canada, or Germany?

The number of possible ways the manager can make these two decisions is 2 x 3 = 6.

Note that the multiplication rule is applicable if there are two or more groupings. In the preceding example, there are two groups: one for investment instruments and the other for countries. In addition, only one item can be selected from each group.

Suppose that there are n numbers in a group and n slots available. Only one member can be assigned to each slot. The number of ways to assign every number to the n slots is **n factorial**: n! = n x (n - 1) x (n - 2) x (n - 3) ... x 1. Note that by convention 0! = 1.

For example, five equity analysts are assigned to cover five industries. The number of ways to assign them is 5! = 5 x 4 x 3 x 2 x 1 = 120.

Unlike the multiplication rule, factorial involves only a single group. It involves arranging items within a group, and the *order* of the arrangement *does* matter. The arrangement of ABCDE is different from the arrangement of ACBDE.

A **combination** is a listing in which the order of listing does not matter. This describes the number of ways that we can choose r objects from a total of n objects, where the order in which the r objects is listed *does not matter* (The **combination formula**, or the **binomial formula**):

For example, if you select two of the ten stocks you are analyzing, how many ways can you select the stocks? 10! / [(10 - 2)! x 2!] = 45.

An ordered listing is known as a **permutation**, and the formula that counts the number of permutations is known as the permutation formula. The number of ways that we can choose r objects from a total of n objects, where the order in which the r objects is listed *does matter*, is:

For example, if you select two of the ten stocks you are analyzing and invest $10,000 in one stock and $20,000 in another stock, how many ways can you select the stocks? Note that the order of your selection is important in this case. $_{10}P_2$ = 10!/(10 - 2)! = 90

Note that there can never be more combinations than permutations for the same problem, because permutations take into account all possible orderings of items, whereas combinations do not.

Quantitative Methods (2)

Common Probability Distributions

1. Introduction and Discrete Random Variables

```
   A <b>probability distribution</b> specifies the probabilities of the possible outcomes
```

If you toss a coin 3 times, the possible outcomes are as follows (where H means heads and T means tails): TTT, TTH, THT, HTT, THH, HTH, HHT, HHH.

In total, there are 8 possible outcomes. Of these:

- Only 1 (TTT) has 0 heads occurring.
- Three (TTH, THT and HTT) have 1 heads occurring.
- Three (THH, HTH and HHT) have 2 heads occurring.
- One (HHH) has 3 heads occurring.

Thus, if x = number of heads in 3 tosses of a coin, then x = 0, 1, 2 or 3.

Now, the respective probabilities are 1/8, 3/8, 3/8 and 1/8, as you have just seen. So:

p(0) = p(0 Heads) = 1/8

p(1) = p(1 Head) = 3/8

p(2) = p(2 Heads) = 3/8

p(3) = p(3 Heads) = 1/8

This is a probability distribution; it records probabilities for each possible outcome of the random variable.

Discrete Probability Distribution

A table, graph or rule that associates a probability $P(X=x_i)$ with each possible value x_i that the discrete random variable X can assume is called a discrete probability distribution. It is a theoretical model for the relative frequency distribution of a population.

A **random variable** is a quantity whose future outcomes are uncertain. Depending on the characteristics of the random variable, a probability distribution may be either discrete or continuous.

A **discrete variable** is one that cannot take on all values within the limits of the variable. It can assume only a countable number of possible values. For example, responses to a five-point rating scale can only take on the values 1, 2, 3, 4, and 5. The variable cannot have the value 1.7. The variable "number of correct answers on a 100-point multiple-choice test" is also a discrete variable since it is not possible to get 54.12 problems correct. The number of movies you will see this year, the number of trades a broker will perform next month, and the number of securities in a portfolio are all examples of discrete variables.

A **continuous variable** is one within the limits of variable ranges for which any value is possible. The number of possible values cannot be counted, and, as you will see later, each individual value has zero probability associated with it. For example, the variable "time to solve an anagram problem" is continuous since it could take 2 minutes or 2.13 minutes, etc., to finish a problem. A variable such as a person's height can take on any value as well. The rate of return on an asset is also a continuous random variable since the exact value of the rate of return depends on the desired number of decimal spaces.

Statistics computed from discrete variables are continuous. The mean on a five-point scale could be 3.117 even though 3.117 is not possible for an individual score.

For any random variable, it is necessary to know two things:

- the list of all possible values that the random variable can take on.
- the probability of each value occurring.

These give a probability distribution. The first item on the list is called the range.

With regard to the range of possible outcomes of a specified random variable:

- **Sometimes the possible values of a random variable have both lower and upper bounds.** For example, there are three possible values of the number of heads showing face-up on two tosses of a coin: 0, 1, and 2. Therefore, the lower bound is 0 and the upper bound is 2.

- **Sometimes the lower bound exists, but the upper bound does not.** For example, the lower bound of the price of a stock is 0, since it cannot fall below 0. However, there is no upper bound on the price (at least theoretically).

- **Sometimes the upper bound exists, but the lower bound does not.** Consider the profit or loss of the seller of a call option. Suppose the buyer pays the seller $2 to buy a call option, which gives the buyer the right to buy a stock at $10 by the end of 2006. The maximum profit the seller can make is $2, but the maximum loss the seller may incur is unlimited since there is no upper bound on the possible values of stock prices.

- **In other cases, neither bound is obvious.** Consider the profit or loss of a big company. In a good year, profits could be as high as dozens of billions of dollars, losses could be equivalent in a very bad year.

2. Probability Function

```
   Every random variable is associated with a probability distribution that describes the
```

A probability function has two key properties:

- 0 â‰¤ P(X=x) â‰¤ 1, because probability is a number between 0 and 1.

- Î£P(X=x) = 1. The sum of the probabilities P(X=x) over all values of X equals 1. If there is an exhaustive list of the distinct possible outcomes of a random variable and the probabilities of each are added up, the probabilities must sum to 1.

The following examples will utilize these two properties in order to examine whether they are probability functions.

Example 1

p(x) = x/6 for X = 1, 2, 3, and p(x) = 0 otherwise

- Substituting into p(x): p(1) = 1/6, p(2) = 2/6 and p(3) = 3/6
  Note that it is not necessary to substitute in any other values, as p(x) is only non-zero for X values 1, 2 and 3.
  In all 3 cases, p(x) lies between 0 and 1, as 1/6, 2/6 and 3/6 are all values in the range 0 to 1 inclusive. So, the first property is satisfied.
- Summing the probabilities gives 1/6 + 2/6 + 3/6 = 1, showing the second property is also satisfied.

Example 2

p(x) = (2x - 3)/16 for X = 1, 2, 3, 4 and p(x) = 0 otherwise

Substituting into p(x): p(1) = -1/16

STOP HERE!

It is impossible for any probability to be negative, so it's not necessary to continue. Property 1 is violated, so it can be said straightaway that p(x) is not a probability function.

Note that individual probabilities in a continuous case cannot occur, so P(X = 5), say, is 0 if X is continuous.

In a continuous case, only a range of values can be considered (that is, 0 < X < 10), whereas in a discrete case, individual values have positive probabilities associated with them.

For a **discrete random variable**, the shorthand notation is p(x) = P(X = x). For **continuous random variables**, the probability function is denoted f(x) and called **probability density function (pdf)**, or just the density. This function is effectively the continuous analogue of the discrete probability function p(x).

- The probability density function, which has the symbol f(x), does not give probabilities, despite its name. Instead, it is the area between the graph and the horizontal axis that gives probabilities. Because of this, the height of f(x) is not restricted to the range 0 to 1, and the graph, which in itself is not a probability, is unrestricted as far as its height is concerned.

- From this information, it follows that the area under the entire graph (i.e., between the graph and the x-axis) must equal 1, because this area encapsulates all the probability contained in the random variable. Recall that for discrete distributions, the probabilities add up to 1.

- Because continuous random variables are concerned with a range of values, individual values have no probabilities, because there is no area associated with individual values. Rather, probabilities are calculated over a range of values. Another way of saying this is that p(x) = 0 for every individual X.

- If a discrete random variable has many possible outcomes, then it can be treated as a continuous random variable for conciseness, and ranges of values can be considered in determining probabilities.

3. Cumulative Distribution Function

```
Analysts are often interested in finding the probability of a range of outcomes rather
```

The two characteristics are:

- The cumulative distribution function lies between 0 and 1 for any x: 0 â‰¤ F(x) â‰¤ 1.
- As we increase x, the cdf either increases or remains constant.

Given the cumulative distribution function, the probabilities for the random variable can also be calculated. In general:

$$\mathbf{P(X = x_n) = F(X_n) - F(X_{n-1})}$$

A **cumulative frequency distribution** is a plot of the number of observations falling in or below an interval. It can show either the actual frequencies at or below each interval (as shown here) or the percentage of the scores at or below each interval. The plot can be a histogram as or a polygon.

Example

Consider a probability function: p(X) = X/6 for X = 1, 2, 3 and p(X) = 0 otherwise. In a previous example it was shown that p(1) = 1/6, p(2) = 2/6, and p(3) = 3/6.

- F(1) indicates the probability that has been accumulated up to and including the point X = 1. Clearly, 1/6 of probability has been accumulated up to this point, so F(1) = 1/6.
- F(2) indicates the probability that has been accumulated up to and including the point X = 2. When X = 2 is reached, the accumulation of 1/6 is taken from X = 1 and 2/6 from X = 2; in total accumulation is 1/6 + 2/6 = 3/6 or, of the probability, so F(2) = 3/6.
- F(3) indicates the probability that has been accumulated up to and including the point X = 3. By the time X = 3 is reached, all the probability has been accumulated: 1/6 from X = 1, 2/6 from X = 2 and 3/6 from X = 3. Thus, 1/6 + 2/6 + 3/6 = 1. Therefore, F(3) = 1.

It is also possible to calculate F(X) for intermediate values. F(0) = 0, as no probability has been accumulated up to the point X = 0; F(1.5) = 1/6, as by the time X = 1.5 is reached, 1/6 of probability has been accumulated from X = 1; F(7) = 1, as by the time 7 is reached, all possible probability from X = 1, 2 and 3 has been collected.

4. Discrete and Continuous Uniform Distribution

```
A <b>uniform distribution</b> is one for which the probability of occurrence is the sa
```

The **discrete uniform distribution** is the simplest of all probability distributions. This distribution has a finite number of specified outcomes, and each outcome is equally likely. Mathematically, suppose that a discrete uniform random variable, X, has n possible outcomes: x_1, x_2, ..., x_{n-1}, and x_n.

- $p(x_1) = p(x_2) = p(x_3) = ... = p(x_{n-1}) = p(x_n) = p(x)$. That is, the probabilities for all possible outcomes are equal.
- $F(x_k) = kp(x_k)$. That is, the cumulative distribution function for the k^{th} outcome is k times of the probability of the k^{th} outcome.
- If there are k possible outcomes in a particular range, the probability for that range of outcomes is kp(X).

For example, the possible outcomes are the integers 1 to 8 (inclusive), and the probability that the random variable takes on any of those possible values is the same for all outcomes (i.e., it is uniform).

If a continuous random variable is equally likely to fall at any point between its maximum and minimum values, it is a **continuous uniform random variable**, and its probability distribution is a **continuous probability distribution**.

- The probability density function is: f(x) = 1/(b - a) for a â‰¤ x â‰¤ b; or 0 otherwise.
- The cumulative density function is: F(x) = 0 for x â‰¤ a; (x - a)/(b - a) for a â‰¤ x â‰¤ b; 1 for x â‰¥ b.

- The probability density function is a horizontal line with a height of 1/(b-a) over a range of values from a to b.
- The cumulative density function is a sloped line with a height of 0 to 1 over a range of values from a to b, and is a horizontal line with a height of 1 when the value of the variable equals or exceeds b.

For example, with a = 0 and b = 8, f(x) = 0.125. If this density is graphed, it will plot as a horizontal line with a value of 0.125.

To find the probability F(3) = P(X â‰¤ 3), find the area under the curve graphing the probability density function, between 0 and 3 on the x-axis. The middle line of the expression for the cumulative probability function is:

F(x) = 0 for x â‰¤ a; (x - a)/(b - a) for a < x < b; 1 for x â‰¥ b

For a continuous uniform random variable, the mean is given by Î¼ = (a + b)/2 and the variance is given by Ïƒ^2 = (b - a)^2/12.

5. Binomial Distribution

```
When a coin is flipped, the outcome is either heads or tails. When a magician guesses
```

A **Bernoulli trial** is an experiment with two outcomes, which can represent success or failure, up move or down move, or another binary outcome. As one of these two outcomes must definitely occur, that is, they are exhaustive, and also mutually exclusive, it follows immediately that the sum of the probabilities of a "success" and a "failure" is 1.

A **binomial random variable** X is defined as the number of successes in n Bernoulli trials. The assumptions are:

- The probability (p) of success is constant for all trials. Similarly, the failure probability 1 - p stays constant throughout the experiment.
- The trials are independent. Thus, the outcome of one trial does not in any way affect the outcome of any subsequent trial.
- The sampling is done with replacement. This means that once an outcome has occurred, it is not precluded from occurring again.

The binomial probability for obtaining r successes in n trials is:

where p(r) is the probability of exactly r successes, n is the number of events, and p is the probability of success on any one trial. This formula assumes that the events are:

- dichotomous (fall into only two categories)
- mutually exclusive
- independent
- randomly selected

To remember the formula, note that there are three components:

- n!/[(n-r)! x r!]. This indicates the number of ways r successes can be achieved and n - r failures in n trials, where the order of success or failure does not matter. This is the combination formula.
- p^r. This is the probability of getting r consecutive success.
- $(1 - p)^{n-r}$. This is the probability of getting n - r consecutive failures.

The values for n and p will always be given to you in a question; their values will never have to be guessed.

Consider this simple application of the binomial distribution. What is the probability of obtaining exactly 3 heads if a coin is flipped 6 times? For this problem, n = 6, r = 3, and p = 0.5, =>p(3) = {6!/[(6 - 3)! x 3!]}$0.5^3(1 - 0.5)^{6-3}$ = 0.3125.

Often the cumulative form of the binomial distribution is used. To determine the probability of obtaining 3 or more successes with n = 6 and p = 0.3, compute p(3) + p(4) + p(5) + p(6).

For a single Bernoulli random variable, Y, which takes on the value 1 with probability p and the value 0 with probability 1 - p, the mean is p and the variance is p(1 - p).

Every random variable has a mean and a variance associated with it. A general binomial random variable, B(n, p), is the sum of n Bernoulli random variables, and so the mean of a B(n, p) random variable is np. Given that a B(1, p) variable has variance p(1 - p), the variance of a B(n, p) random variable is n times that value, or np(1 - p), using the independent assumption.

For example, for a B(n = 5, p = 0.10) random variable, the expected number of successes is 5 x 0.1 = 0.5 with a standard deviation of $(5 \times 0.1 \times 0.9)^{1/2}$ = 0.67.

6. Normal Distribution

```
<b>Normal distributions</b> are a family of distributions that have the same general s
```

- They are symmetrical with scores more concentrated in the middle than in the tails.
- Normal distributions are sometimes described as *bell-shaped* with a single peak at the exact center of the distribution.
- The tails of the normal curve extends indefinitely in both directions. That is, possible outcomes of a normal distribution lie between - âˆž and + âˆž.
- Normal distributions may differ in how spread-out they are.

The graph looks like this:

The key properties are:

- The normal distribution is completely described by two parameters: the mean (Î¼) and the standard deviation (Ïƒ).
- The normal distribution is symmetrical: it has a skewness of 0, a kurtosis (it measures the peakedness of a distribution) of 3, and an excess kurtosis (which equals kurtosis less 3) of 0. As a consequence, the mean, median, and mode are all equal for a normal random variable.
- A linear combination of two or more normal random variables is also normally distributed.

One reason the normal distribution is important is that many psychological, educational, and financial variables are distributed approximately normally. Measures of reading ability, introversion, job satisfaction, and memory are among the many psychological variables approximately normally distributed. Although the distributions are only approximately normal, they are usually quite close.

A second reason is that it is easy for mathematical statisticians to work with. Many kinds of statistical tests can be derived for normal distributions. Almost all statistical tests discussed in the textbook assume normal distributions. Fortunately, these tests work very well even if the distribution is only approximately normally distributed. Some tests work well even with very wide deviations from normality.

Finally, if the mean and standard deviation of a normal distribution are known, it is easy to convert back and forth from raw scores to percentiles.

For example, normal distribution is an approximate model for asset returns. The price of any asset can only drop to 0. Therefore, the lowest return on an asset is -100% (i.e., all investment in the asset is lost). Since the normal distribution extends to negative infinity without limit, it is not an accurate model for asset returns. However, for the normal distribution, the probability of outcomes below -100% is very small. Therefore, the normal distribution can be considered an approximate model for returns. However, the normal distribution tends to underestimate the probability of extreme returns.

Confidence Intervals for a Normally Distributed Random Variable.

Analysts can use the sample mean to estimate the population mean, and the sample standard deviation to estimate the population standard deviation. The sample mean and sample standard deviation are **point estimates**.

Probability statements about a random variable are often framed using **confidence intervals** built around point estimates. In investment work, confidence intervals for a normal random variable in relation to its estimated mean are often used.

Confidence intervals use point estimates to make probability statements about the dispersion of the outcomes of a normal distribution. A confidence interval specifies the percentage of all observations that fall in a particular interval.

The exact confidence intervals for a normal random variable X:

- 90% confidence interval for X is: x-bar - 1.645Ïƒ to x-bar + 1.645Ïƒ: this means that 10% of the observations fall outside the 90% confidence interval, with 5% on each side.

- 95% confidence interval for X is: x-bar - 1.96Ïƒ to x-bar + 1.96 Ïƒ: this means that 5% of the observations fall outside the 95% confidence interval, with 2.5% on each side.

- 99% confidence interval for X is: x-bar - 2.58 Ïƒ to x-bar + 2.58 Ïƒ: this means that 1% of the observations fall outside the 99% confidence interval, with 0.5% on each side.

Hint: memorize these numbers (1.645, 1.96 and 2.58) to quickly solve relevant problems. For details about confidence intervals, refer to Reading 5 - Sampling and Estimation.

The Univariate and Multivariate Distributions

To this point, the focus has been on distributions that involve only one variable, such as the binomial, uniform, and normal distributions. A **univariate distribution** describes a single random variable. For example, suppose that you would like to model the distribution of the return on an asset. Such a distribution is a univariate distribution.

A **multivariate distribution** specifies the probabilities for a group of related random variables. It is used to describe the probabilities of a group of continuous random variables if all of the individual variables follow a normal distribution.

Each individual normal random variable would have its own mean and its own standard deviation, and hence its own variance. When you are dealing with two or more random variables in tandem, the strength of the relationship between (or among) the variables assumes huge importance. You will recall that the strength of the relationship between two random variables is known as the correlation.

When there is a group of assets, the distribution of returns on each asset can either be modeled individually or on the assets as a group. A multivariate normal distribution for the returns on n stocks is completely defined by three lists of parameters:

- The list of the mean returns on the individual securities (n means in total).
- The list of the securities' variances of return (n variances in total).
- The list of all the distinct pairwise return correlations (n(n-1)/2 distinct correlations in total).

The higher the correlation values, the higher the variance of the overall portfolio. In general, it is better to build a portfolio of stocks whose prices are not strongly correlated with each other, as this lowers the variance of the overall portfolio.

It is the correlation values that distinguish a multivariate normal distribution from a univariate normal distribution. Consider a portfolio consisting of 2 assets (n = 2). The multivariate normal distribution can be defined with 2 means, 2 variances, and 2 x (2-1)/2 = 1 correlation. If an analyst has a portfolio of 100 securities, the multivariate normal distribution can be defined with 100 means, 100 variances, and 100 x (100 - 1)/2 = 4950 correlations. Portfolio return is a weighted average of the returns on the 100 securities. A weighted average is a linear combination. Thus, portfolio return is normally distributed if the individual security returns are (joint) normally distributed. In order to specify the normal distribution for portfolio return, analysts need means, variances, and the distinct pairwise correlations of the component securities.

7. The Standard Normal Distribution

```
   The problem with working with a normal distribution is that its formula is very compli
```

The way to get around this problem is to standardize a normal random variable, which involves converting it to a general scale for which probability tables exist.

The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1. It is denoted as N(0,1). Below are some confidence intervals for the standard normal distribution:

There is an unlimited number of normal distributions, each with a different mean or standard deviation. Therefore, it's impractical to provide a table of probabilities for each combination of mean and standard deviations. However, the actual distribution for a normal random variable to a standard normal distribution can be standardized. Normal distributions can be transformed to standard normal distributions by the formula:

where X is a score from the original normal distribution, Î¼ is the mean of the original normal distribution, and Ïƒ is the standard deviation of original normal distribution.

The standard normal distribution is sometimes called the **z distribution**. A **z score** (also called **z-value** or **z-statistic**) is the distance between a selected value (X) and the population mean, divided by the population standard deviation. It is in fact a standard normal random variable. For instance, if a person scored 70 on a test with a mean of 50 and a standard deviation of 10, that person scored 2 standard deviations above the mean. Converting the test scores to z scores, an X of 70 would be: z = (70 - 50) / 10 = 2.

So, a z score of 2 means the original score was 2 standard deviations above the mean. Note that the z distribution will only be a normal distribution if the original distribution (X) is normal.

Example

The rates of return on the assets in a portfolio are normally distributed, with a mean of 20% and a standard deviation of 12%. What's the probability that the return on an asset falls between -3.52% and 50.96%?

Solution:

- If the return on the asset is -3.52%, its z-value is z = (-3.52% - 20%)/12% = -1.96.
- If the return on the asset is 50.96%, its z-value is (50.96% - 20%)/12% = 2.58.
- Recall that a normal distribution is symmetrical. The 95% confidence interval is -1.96 to 1.96. The 99% confidence interval is -2.58 to 2.58. The probability of z-value falling between -1.96 and 2.58 is 95%/2 + 99%/2 = 97%.

How to Use the Z-Table

The z-table gives cumulative probability values for a Z-graph (or standard normal distribution). In other words, by looking up a particular z-value, the probability of all values smaller than this value can be found.

Recall that a cumulative distribution function gives probabilities of the form: P (X < x), that is, probabilities that are less than the value being examined.

In the same way, the table gives probabilities of the form: P(Z < z), that is, probabilities that are less than the z-value being looked up or calculated.

The first column, headed z, ranges from 0.0 to 4.0. Across the top, the values range from 0.00 to 0.09. These values allow a person to look up any z-value to two decimal places. The vertical column gives the value to one decimal place and the horizontal row gives the values in the second decimal position. So, for example, if you want to look up z = 0.26, you would go to the row that starts 0.2 and go to the column headed 0.06. You can then read the value 0.6026, which means that 60.26% of z-values or z-scores are less than 0.26.

So, the body of the table (the middle part) gives probabilities (or areas under the graph); hence; all the values in the body of the table lie between 0 and 1. Actually, they lie between 0.5 and 1, because the table starts considering z-values from the point 0.

You can see that 60.26% of the area under the graph lies to the left of the point z = 0.26. Therefore, 39.74% of the area lies to the right of this point. Because areas under continuous graphs give probabilities, you can also say that: P(Z < 0.26) = 0.6026.

At this point, you might have a question. Z-scores generally range from -4 to 4, but the table has no negative values. Why not?

To answer this, recall that a normal distribution is symmetrical about its mean; this fact can be used to calculate probabilities for negative z-values very easily.

Example

Take the point z = 1.83. The table shows the area to the left of 1.83 as 0.9664. The area to the right of 1.83 is therefore 1- 0.9664 = 0.0336. Because of symmetry, the area to the right of 1.83 is the same as the area to the left of -1.83. So, if the area to the left of a negative value is required, simply take 1 - the area to the left of the equivalent positive value. Problem solved.

Continuing with this example, the area to the left of -1.83 is 0.0336, which is the probability of obtaining a z-score smaller than -1.83.

A portion of the z-table is presented below. (N(z) represents the cumulative probability distribution for a standard random variable Z):

Example

Suppose that you want to find the probability that a standard normal variable is less than or equal to 1.01. In the z-table, go down the column headed by the letter z to 1.00. Move to the right and read the entry under the column headed 0.01. It's 0.8438. That is, P(Z â‰¤ 1.01) = 0.8438 or 84.38%, and P(Z â‰¥ 1.01) = 15.62%.

A standard normal distribution is symmetrical with a mean of 0. Therefore, P(Z â‰¤ 0) = 50% ==> P(0 â‰¤ Z â‰¤ 1.01) = 0.8438 - 0.5 = 34.38%.

Comprehensive Example

The heights of people in a city are normally distributed with mean 170 cm and standard deviation 10 cm. Calculate the probability that a randomly chosen person from this population has a height that is:

- less than 180 cm
- less than 150 cm
- greater than 175 cm
- between 160 cm and 185 cm
- greater than 150 cm

Because there is a normal distribution that is not standardized, it is important to standardize each value before looking up answers in a z-table. Let X = the heights of people in the population.

1. You want P(X < 180). Standardizing gives P(Z < (180-170)/10) = P(Z < 1). From tables, look up 1.00 and you get 0.8413. So, the answer is 0.8413.

2. You want P(X < 150). Standardizing gives P(Z < (150-170)/10) = P(Z < -2). Recall that this probability is the same as P(Z > 2), due to the symmetrical nature of the normal distribution. From tables, look up 2.00 and you get 0.9772. This is the probability that Z<2, so 1 - 0.9772 = 0.0228 is the probability that we want. So, the answer is 0.0228.

3. You want P(X > 175). Standardizing gives P(Z > (175-170)/10) = P(Z > 0.5). From tables, look up 0.50 and you get 0.6915. This is the probability that Z < 0.5, so 1-0.6915 = 0.3085 is the probability that we want. So, the answer is 0.3085.

4. You want P(160 < X < 185). Standardizing gives P(-1 < Z < 1.5). In order to find a region like this, note that this region can be written as P(Z < 1.5) - P(Z < -1). From tables, look up 1.50 and you get 0.9332. This is the probability that Z < 1.5. To find P(Z < -1), look up 1.00 and you get 0.8413. 1 - 0.8413 = 0.1587 is therefore the probability you want. So, the answer is 0.9332 - 0.1587 = 0.7745.

5. You want P(X > 150). Standardizing gives P(Z > (150-170)/10) = P(Z > -2). Using symmetry, P(Z > -2) = P(Z < 2). From tables, look up 2.00 and you get 0.9772. So, the answer is 0.9772.

Notes:

- These examples encompass the different types of regions you can get. Familiarize yourself with how each region works. If you can do these five examples, you know that you understand this process well.
- Always standardize before going to tables.
- As a check, if the region you are looking for is bigger than half the area under the graph, the answer will be bigger than 0.5 and vice versa. Verify this factor to make sure that you haven't made a silly mistake.
- It is possible to work in reverse. For example, a standardized value can be converted back to an X value using X = ÏƒZ + Î¼. So, in the height example above, a standardized value of 3 is equivalent to an X value of 10 x 3 + 170 = 200 cm.

8. Shortfall Risk and Roy's Safety-First Criterion

```
   The focus of this section is assessing risks in a portfolio, a process that allows us
```

Shortfall risk is the risk that portfolio value will fall below some minimum acceptable level over some time horizon. The risk that assets in a defined benefit plan will fall below plan liabilities is an example of a shortfall risk. Therefore, shortfall risk is a **downside risk**. In contrast, when a risk-averse investor makes portfolio decisions in terms of the mean return and the variance (or standard deviation) of return, both upside and downside risks are considered. For example, portfolios A and B have the same mean return of 20%. The standard deviations of the returns on A and B are 5% and 8% respectively. Portfolio B has a higher risk because its standard deviation is higher. However, though the return on portfolio B is more likely to fall below 20%, it's also more likely to exceed 20%.

Safety-first rules focus on shortfall risk.

Suppose an investor views any return below a level of R_L as unacceptable. Roy's safety-first criterion states that the optimal portfolio minimizes the probability that portfolio return, R_P, falls below the threshold level, R_L. In symbols, the investor's objective is to choose a portfolio that minimizes $P(R_P < R_L)$. When portfolio returns are normally distributed, the investor can calculate $P(R_P < R_L)$ using the number of standard deviations R_L lies below the expected portfolio return, $E(R_P)$. The portfolio for which $E(R_P) - R_L$ is largest in terms of units of standard deviations minimizes $P(R_P < R_L)$. Thus, if returns are normally distributed, the safety-first optimal portfolio maximizes the safety-first ratio (SFRatio):

The quantity $E(R_P) - R_L$ is the distance from the mean return to the shortfall level. It measures the excess return over the threshold return level. SFRatio gives the distance in units of standard deviation: it measures the excess return over the threshold level per unit of risk.

For example, the expected return on a portfolio is 20%, and the standard deviation of the portfolio return is 15%. Suppose the minimum acceptable return level is 10%. SFRatio = (20% - 10%)/15% = 0.67.

Note that the SFRatio is similar to the Sharpe ratio $(E(R_p) - R_f)$/standard deviation, where R_f is the risk-free rate. The SFRatio becomes the Sharpe ratio if the risk-free rate is substituted for the threshold return R_L. Therefore, the safety-first criterion focuses on the excess return over the threshold return, while the Sharpe ratio focuses on the excess return over the risk-free rate.

Roy's safety-first criterion states that the optimal portfolio should minimize the probability that the rate of return of the portfolio (R_P) will fall below a stated threshold level (R_L). If returns are normally distributed, it states that the optimal portfolio maximizes the safety-first ratio. Therefore, assuming that returns are normally distributed, the safety-first optimal portfolio can be selected using one of the following two criteria:

- Lowest probability of $R_P < R_L$
- Highest safety-first ratio.

There are three steps in choosing among portfolios using Roy's criterion (assuming normality):

- Calculate the portfolio's SFRatio.
- Evaluate the standard normal cdf at the value calculated for the SFRatio; the probability that return will be less than R_L is N(-SFRatio).
- Choose the portfolio with the lowest probability.

Example

Suppose that a certain fund has reached a value of $500,000. At the end of the next year, the fund managers wish to withdraw $20,000 for additional funding purposes, but do not wish to tap into the original $500,000.

There are three possible investment options:

Which option is most preferable? (You may assume normally distributed returns throughout.)

Answer

First, note that since the managers do not want to tap into the original fund, a return of 20000/500000 = 0.04 is the minimum acceptable return; this is the threshold return, R_L.

You now need to calculate the SFRatio in each case:

A: (10-4)/15 = 0.4

B: (8-4)/12 = 0.33

C: (9-4)/14 = 0.36

You can conclude that portfolio A, with the highest SFRatio of the three, is the most preferable.

You can also take this a step further and calculate the probability that the portfolio return will fall below the threshold return, that is, $P(R_P < R_L)$. To do this, we take the negative of the SFRatio in each case and find the cdf of the standard normal distribution for this value.

In symbols, $P(R_P < R_L)$ = F(-SFRatio), where F is the cdf of the standard normal distribution.

From normal tables, F(-0.4) = 0.3446, F(-0.33) = 0.3707 and F(-0.36) = 0.3594.

This indicates that, for portfolio A, the chance of obtaining a return below R is 0.3446, with corresponding values for portfolios B and C of 0.3707 and 0.3594 respectively.

Since the chance of not exceeding the threshold return is lowest for portfolio A, this is again the best option.

9. The Lognormal Distribution

```
   The key properties of the normal distribution have been presented in the LOS above. As
```

- It is symmetrical about the mean.
- It has zero skewness.
- It has a kurtosis of 3.

A random variable, Y, follows a **lognormal distribution** if its natural logarithm, lnY, is normally distributed. You can think of the term lognormal as "the log is normal." For example, suppose X is a normal random variable, and $Y = e^X$. Therefore, $LnY = Ln(e^X) = X$. Because X is normally distributed, Y follows a lognormal distribution.

- Like the normal distribution, the lognormal distribution is completely described by two parameters: mean and variance.
- Unlike the normal distribution, the lognormal distribution is defined in terms of the parameters of the associated normal distribution. Note that the mean of Y is not equal to the mean of X, and the variance of Y is not equal to the variance of X. In contrast, the normal distribution is defined by its own mean and variance.
- The lognormal distribution is bounded below by 0. In contrast, the normal distribution extends to negative infinity without limit.
- The lognormal distribution is skewed to the right (i.e., it has a long right tail). In contrast, the normal distribution is bell-shaped (i.e., it is symmetrical).

The reverse is also true; if a random variable Y follows a lognormal distribution, then its natural logarithm, lnY, is normally distributed.

10. Continuously Compounded Rates of Return

```
   A <b>discretely compounded rate of return</b> measures the rate of changes in the valu
```

The **continuously compounded rate of return** measures the rate of change in the value of an asset associated with a holding period under the assumption of continuously compounding. It is the natural logarithm of 1 plus the holding period return, or equivalently, the natural logarithm of the ending price over the beginning price.

From t to t + 1:

S: stock price. $R_{t,\, t+1}$: the rate of return from t to t + 1

Example 1

S_0 = \$30, S_1 = \$34.50. ==> $R_{t,\, t+1}$ = \$34.50/\$30 - 1 = 0.15, and $r_{0,\, 1}$ = 0.139762. The continuously compounded return is smaller than the associated holding period return.

Example 2

Assume that a particular stock has a price of \$200 at the start of a period and a price of \$250 at the end of that period. That is, S_0 = 200, and S_1 = 250.

Hence, R, the holding period return, is: R = [S_1/ S_0] - 1 = [250/200] - 1 = 25%.

Using the formula for the continuously compounded rate of return gives: ln(1+R) = ln(S_1/ S_0) = ln(1.25) = 0.223 or 22.3%.

In order to see why the latter is preferable, consider the following:

Suppose that the stock now falls from \$250 to \$200. Then, R = [200/250] - 1 = -0.2 or -20%. So, effectively, the stock has returned to its original price, but combining the two rates of return and averaging them gives [(0.25+ (-0.2)] / 2 = 0.05 / 2 = 0.025 or 2.5%, which is misleading, as in actual fact the stock has returned to its original price, and hence the return is effectively 0%.

However, ln(1+R) = ln[(1 + (-0.2)] = ln(0.8) = -0.223 or -22.3%, which is exactly the negative of the original return. Averaging these two rates gives [(0.223 + (-0.223)] / 2 = 0 / 2 = 0 or 0%, which is the true rate of return for the period.

Thus, a continuously compounded return gives a more accurate account of the true picture of the rate of return over a period.

11. Student's t-, Chi-Square, and F-Distributions

```
   No text
```

12. Monte Carlo Simulation

```
   When a system is too complex to be analyzed using ordinary methods, investment analyst
```

After generating the data, quantities such as the mean and variance of the generated numbers can be used as estimates of the unknown parameters of the population (parameters are too complex to find through normal methods).

The term "Monte Carlo simulation" derives from the generation of a large number of random samples, such as might occur in the Monte Carlo Casino.

- It allows us to experiment with a proposed policy and assess the risks before actually implementing it. For example, it is used to simulate the interaction of pension assets and the liabilities of defined benefit pension plans.
- It is widely used to develop estimates of **Value at Risk** (VAR). VAR involves estimating the probability that portfolio losses exceed a predefined level.
- It is used to value complex securities such as European options, mortgage-backed securities with complex embedded options.
- Researchers use it to test their models and tools.

Limitations of Monte Carlo Simulation:

- It is a complement to analytical methods. It provides only statistical estimates, not exact results.
- It does not directly provide precise insights as analytical methods do. For example, it cannot reveal cause-and-effect relationships.

Historical simulation samples from a historical record of returns (or other underlying variables) are used to simulate a process because the historical record provides the most direct evidence on distributions (and that past applies to the future). In contrast, Monte Carlo simulation uses a random number generator with a specified distribution. A drawback is that any risk not represented in the time period selected will not be reflected in the simulation. For example, if a stock market crash did not take place in the sample period, such a risk will not be reflected in the simulation. In addition, this method does not lend itself to "what-if" analysis.

Sampling and Estimation

1. Sampling Methods

```
<p> </p>In investment analysis, it is often impossible to study every member of a popu
```

Sampling is the process of obtaining a sample.

Simple Random Sampling

A **simple random sample** is a sample obtained in such a way that each element of a population has an equal probability of being selected. The selection of any one element has no impact on the chance of selecting another element.

A sample is random if the method for obtaining the sample meets the criterion of randomness (each element having an equal chance at each draw). "Simple" indicates that this process is not difficult, and "random" indicates that you don't know in advance which observations will be selected in the sample. The actual composition of the sample itself does not determine whether or not it's a random sample.

Example

Suppose that a company has 30 directors, and you wish to choose 10 of them to serve on a committee. You could place the names of the 30 directors on separate pieces of paper and draw them out one by one until you have drawn a sample of 10.

Note that the conditions for simple random sampling have been satisfied in that every one of the 30 directors has an equal (non-zero) chance of being selected in the sample.

In this example, it makes no sense to sample with replacement, as this would mean that once you have drawn a name, that name goes back into the hat (it is replaced), and can be drawn again. If the same person's name is drawn more than once, you won't end up with a sample size 10 if you draw 10 names; this experiment should therefore be done without replacement.

Sampling Error

The sample taken from a population is used to infer conclusions about that population. However, it's unlikely that the sample statistic would be identical to the population parameter. Suppose there is a class of 100 students and a sample of 10 from that class is chosen. If, by chance, most of the brightest students are selected in this sample, the sample will provide a misguided idea of what the population looks like (because the sample mean x-bar will be much higher than the population mean in this case). Equally, a sample comprising mainly weaker students could be chosen, and then the opposite applicable characteristics would apply. The ideal is to have a sample which comprises a few bright students, a few weaker students, and mainly average students, as this selection will give a good idea of the composition of the population. However, because which items go into the sample cannot be controlled, you are dependent to some degree on chance as to whether the results are favorable (indicative of the population) or not.

Sampling error (also called **error of estimation**) is the difference between the observed value of a statistic and the quantity it is intended to estimate. For example, sampling error of the mean equals sample mean minus population mean.

Sampling error can apply to statistics such as the mean, the variance, the standard deviation, or any other values that can be obtained from the sample. The sampling error varies from sample to sample. A good estimator is one whose sample error distribution is highly concentrated about the population parameter value.

Sampling error of the mean would be: Sample mean - population mean = x-bar - Î¼.

Sampling error of the standard deviation would be: Sample standard deviation - population standard deviation = s - Ïƒ.

Note that sampling error is not a bias. A **biased sample** is one in which the method used to create the sample results in samples that are *systematically different* from the population. For instance, consider a research project on attitudes toward sex. Collecting the data by publishing a questionnaire in a magazine and asking people to fill it out and send it in would produce a biased sample. People interested enough to spend their time and energy filling out and sending in the questionnaire are likely to have different attitudes toward sex than those not taking the time to fill out the questionnaire.

It is important to realize that it is the method used to create the sample, not the actual makeup of the sample, that defines the bias. A random sample that is very different from the population is not biased: it is by definition not systematically different from the population. It is randomly different.

Sampling Distribution

A sample statistic itself is a random variable, which varies depending upon the composition of the sample. It therefore has a probability distribution. The **sampling distribution** of a statistic is the distribution of all the distinct possible values that the statistic can assume when computed from samples of the same size randomly drawn from the same population. The most commonly used sample statistics include mean, variance, and standard deviation.

If you compute the mean of a sample of 10 numbers, the value you obtain will not equal the population mean exactly; by chance, it will be a little bit higher or a little bit lower. If you sampled sets of 10 numbers over and over again (computing the mean for each set), you would find that some sample means come much closer to the population mean than others. Some would be higher than the population mean and some would be lower. Imagine sampling 10 numbers and computing the mean over and over again, say about 1,000 times, and then constructing a relative frequency distribution of those 1,000 means. This distribution of means is a very good approximation to the sampling distribution of the mean. The sampling distribution of the mean is a theoretical distribution that is approached as the number of samples in the relative frequency distribution increases. With 1,000 samples, the relative frequency distribution is quite close; with 10,000, it is even closer. As the number of samples approaches infinity, the relative frequency distribution approaches the sampling distribution.

The sampling distribution of the mean for a sample size of 10 is just an example; there is a different sampling distribution for other sample sizes. Also, keep in mind that the relative frequency distribution approaches the sampling distribution as the number of samples increases, not as the sample size increases, since *there is a different sampling distribution for each sample size.*

A sampling distribution can also be defined as the relative frequency distribution that would be obtained if all possible samples of a particular sample size were taken. For example, the sampling distribution of the mean for a sample size of 10 would be constructed by computing the mean for each of the possible ways in which 10 scores could be sampled from the population and creating a relative frequency distribution of these means.

Although these two definitions may seem different, they are actually the same: Both procedures produce exactly the same sampling distribution.

Statistics other than the mean have sampling distributions too. The sampling distribution of the median is the distribution that would result if the median instead of the mean were computed in each sample.

Sampling distributions are very important since almost all inferential statistics are based on sampling distributions.

Stratified Random Sampling

In **stratified random sampling**, the population is subdivided into subpopulations (strata) based on one or more classification criteria. Simple random samples are then drawn from each stratum (the sizes of the samples are proportional to the relative size of each stratum in the population). These samples are then pooled.

It is important to note that the size of the data in each stratum does not have to be the same or even similar, and frequently isn't.

Stratified random sampling guarantees that population subdivisions of interest are represented in the sample. The estimates of parameters produced from stratified sampling have greater precision (i.e., smaller variance or dispersion) than estimates obtained from simple random sampling.

For example, investors may want to fully duplicate a bond index by owning all the bonds in the index in proportion to their market value weights. This is known as **pure bond indexing**. However, it's difficult and costly to implement because a bond index typically consists of thousands of issues. If simple sampling is used, the sample selected may not accurately reflect the risk factors of the index. Stratified random sampling can be used to replicate the bond index.

- Divide the population of index bonds into groups with similar risk factors (e.g., issuer, duration/maturity, coupon rate, credit rating, call exposure, etc.). Each group is called a stratum or cell.
- Select a sample from each cell proportional to the relative market weighting of the cell in the index.

A stratified sample will ensure that at least one issue in each cell is included in the sample.

Cluster Sampling

Cluster sampling is a process of dividing a population into multiple groups/clusters. The clusters must be representative of the total population. A simple random sample of the cluster is chosen and the elements in each of these clusters are then sampled.

There are single-stage, two-stage, or multiple-stage sampling methods in cluster sampling. These methods depend upon the number of steps required to create the desired sample.

Cluster Sampling versus Stratified Sampling

- In Cluster Sampling, the sampling is done on a population of clusters therefore, cluster/group is considered a sampling unit. In Stratified Sampling, elements within each stratum are sampled.
- In Cluster Sampling, only selected clusters are sampled. In Stratified Sampling, from each stratum, a random sample is selected.

Cluster sampling is advantageous when a large population is in need of a survey because it is less costly.

Non-Probability Sampling

Non-probability sampling is defined as a sampling technique in which the researcher selects samples based on the subjective judgment of the researcher rather than random selection. Researchers use this method in studies where it is impossible to draw random probability sampling due to time or cost considerations.

Convenience sampling is a non-probability sampling technique where samples are selected from the population only because they are conveniently available to the researcher.

In the **judgmental sampling method**, researchers select the samples based purely on the researcherâ€™s knowledge and credibility. In other words, researchers choose only those people who they deem fit to participate in the research study.

2. The Central Limit Theorem

```
    The <b>central limit theorem</b> states that, given a distribution with a mean Î¼ and
```

The amazing and counter-intuitive thing about the central limit theorem is that no matter the shape of the original distribution, x-bar approaches a normal distribution.

- If the original variable X has a normal distribution, then x-bar will be normal regardless of the sample size.
- If the original variable X does not have a normal distribution, then x-bar will be normal only if N â‰¥ 30. This is called a distribution-free result. This means that no matter what distribution X has, it will still be normal for sufficiently large n.

Keep in mind that N is the sample size for each mean and not the number of samples. Remember that in a sampling distribution the number of samples is assumed to be infinite. The sample size is the number of scores in each sample; it is the number of scores that goes into the computation of each mean.

Two things should be noted about the effect of increasing N:

- The distributions become more and more normal.
- The spread of the distributions decreases.

Based on the central limit theorem, when the sample size is large, you can:

- use the sample mean to infer the population mean.
- construct confidence intervals for the population mean based on the normal distribution.

Note that the central limit theorem does not prescribe that the underlying population must be normally distributed. Therefore, the central limit theorem can be applied to a population with any probability distribution.

3. Standard Error of the Sample Mean

```
    The <b>standard error</b> of a statistic is the standard deviation of the sampling dis
```

Standard errors are important because they reflect how much sampling fluctuation a statistic will show. The inferential statistics involved in the construction of confidence intervals and significance testing are based on standard errors. The standard error of a statistic depends on the sample size. In general, the larger the sample size, the smaller the standard error. The standard error of a statistic is usually designated by the Greek letter sigma (Ïƒ) with a subscript indicating the statistic.

The **standard error of the mean** is designated as: Ïƒ$_{\text{m}}$. It is the standard deviation of the sampling distribution of the mean. The formula for the standard error of the mean is:

where Ïƒ is the standard deviation of the original distribution and N is the sample size (the number of scores each mean is based upon).

This formula does not assume a normal distribution. However, many of the uses of the formula do assume a normal distribution. The formula shows that the larger the sample size, the smaller the standard error of the mean. More specifically, the size of the standard error of the mean is inversely proportional to the square root of the sample size.

Example 1

Suppose that the mean grade of students in a class is 62%, with a standard deviation of 10%. A sample of 30 students is taken from the class. Calculate the standard error of the sample mean and interpret your results.

You are given that Î¼ = 62, and Ïƒ =10. Since n = 30, the standard error of the sample mean is: Ïƒ$_{m}$ = 10/30$^{1/2}$ = 1.8257. This means that if you took all possible samples of size 30 from the class, the mean of all those samples would be 62 and the standard error would be 1.8257.

Note that if you took a sample size of 50, the standard error would then be: Ïƒ$_{m}$ = 10 / 50$^{1/2}$ = 1.4142.

The standard error would drop as the sample size increased, which agrees with the information above.

When sample standard deviation (s) is used as an estimate of Ïƒ (when it is unknown), the estimated standard error of the mean is s/N$^{1/2}$. In most practical applications analysts need to use this formula because the population standard deviation is almost never available.

Example 2

Suppose that the mean grade of students in a class is unknown, but a sample of 30 students is taken from the class and the mean from the sample is found to be 60%, with a standard deviation of 9%. Calculate the standard error of the sample mean and interpret your results.

Now, Î¼ and Ïƒ are unknown, but m is given as 60 and s is given as 9. Since n = 30, you can estimate the standard error of the sample mean as: 9/30$^{1/2}$ = 1.6432. This means that if you took all possible samples of 30 from the class, you would estimate the standard error to be 1.6432.

It is important to note that when you have Ïƒ, you must use it; when you don't, you use its sample equivalent, s.

4. Point Estimates of the Population Mean

```
   Very often, there are a number of different estimators that can be used to estimate ur
```

- **unbiasedness**

An estimator's expected value (the mean of its sampling distribution) equals the parameter it is intended to estimate. For example, the sample mean is an unbiased estimator of the population mean because the expected value of the sample mean is equal to the population mean.

- **efficiency**

An estimator is efficient if no other unbiased estimator of the sample parameter has a sampling distribution with smaller variance. That is, in repeated samples, analysts expect the estimates from an efficient estimator to be more tightly grouped around the mean than estimates from other unbiased estimators. For example, the sample mean is an efficient estimator of the population mean, and the sample variance is an efficient estimator of the population variance.

- **consistency**

A consistent estimator is one for which the probability of accurate estimates (estimates close to the value of the population parameter) increases as sample size increases. In other words, a consistent estimator's sampling distribution becomes concentrated on the value of the parameter it is intended to estimate as the sample size approaches infinity. As the sample size increases to infinity, the standard error of the sample mean declines to 0 and the sampling distribution concentrates around the population mean. Therefore, the sample mean is a consistent estimator of the population mean.

The single estimate of an unknown population parameter calculated as a sample mean is called a **point estimate** of the mean. The formula used to compute the point estimate is called an **estimator**. The specific value calculated from sample observations using an estimator is called an **estimate**. For example, the sample mean is a point estimate of the population mean. Suppose two samples are taken from a population and the sample means are 16 and 21 respectively. Therefore, 16 and 21 are two estimates of the population mean. Note that an estimator will yield different estimates as repeated samples are taken from the sample population.

A **confidence interval** is an interval for which one can assert with a given probability 1 - Î±, called the **degree of confidence**, that it will contain the parameter it is intended to estimate. This interval is often referred to as the (1 - Î±)% confidence interval for the parameter, where Î± is referred to as the level of significance. The end points of a confidence interval are called the lower and upper **confidence limits**.

For example, suppose that a 95% confidence interval for the population mean is 20 to 40. This means that:

- There is a 95% probability that the population mean lies in the range of 20 to 40.
- "95%" is the degree of confidence.
- "5%" is the level of significance.
- 20 and 40 are the lower and higher confidence limits, respectively.

5. Confidence Intervals for the Population Mean and Selection of Sample Size

```
Confidence intervals are typically constructed using the following structure:<p> </p>
```

Confidence Interval = Point Estimate Â± Reliability Factor x Standard Error

- The point estimate is the value of a sample statistic of the population parameter.
- The reliability factor is a number based on the sampling distribution of the point estimate and the degree of confidence (1 - Î±).
- Standard error refers to the standard error of the sample statistic that is used to produce the point estimate.

Whatever the distribution of the population, the sample mean is always the point estimate used to construct the confidence intervals for the population mean. The reliability factor and the standard error, however, may vary depending on three factors:

- Distribution of population: normal or non-normal
- Population variance: known or unknown
- Sample size: large or small

z-Statistic: a standard normal random variable

If a population is normally distributed with a known variance, a z-statistic is used as the reliability factor to construct confidence intervals for the population mean.

In practice, the population standard deviation is rarely known. However, learning how to compute a confidence interval when the standard deviation is known is an excellent introduction to how to compute a confidence interval when the standard deviation has to be estimated.

Three values are used to construct a confidence interval for Î¼:

- the sample mean (m)
- the value of z (which depends on the level of confidence)
- the standard error of the mean (Ïƒ)$_{\text{m}}$

The confidence interval has m for its center and extends a distance equal to the product of z and in both directions. Therefore, the formula for a confidence interval is:

$$m - z\ Ïƒ_{m} <= Î¼ <= m + z\ Ïƒ_{m}$$

For a (1 - Î±)% confidence interval for the population mean, the z-statistic to be used is $z_{Î±/2}$. $z_{Î±/2}$ denotes the points of the standard normal distribution such that Î±/2 of the probability falls in the right-hand tail.

Effectively, what is happening is that the (1 - Î±)% of the area that makes up the confidence interval falls in the center of the graph, that is, symmetrically around the mean. This leaves Î±% of the area in both tails, or Î±/2 % of area in each tail.

Commonly used reliability factors are as follows:

- 90% confidence intervals: $z_{0.05}$ = 1.645. Î± is 10%, with 5% in each tail
- 95% confidence intervals: $z_{0.025}$ = 1.96. Î± is 5%, with 2.5% in each tail
- 99% confidence intervals: $z_{0.005}$ = 2.575. Î± is 1%, with 0.5% in each tail

Example

Assume that the standard deviation of SAT verbal scores in a school system is known to be 100. A researcher wishes to estimate the mean SAT score and compute a 95% confidence interval from a random sample of 10 scores.

The 10 scores are: 320, 380, 400, 420, 500, 520, 600, 660, 720, and 780. Therefore, m = 530, N = 10, and $Ïƒ_{m}$= 100 / $10^{1/2}$ = 31.62. The value of z for the 95% confidence interval is the number of standard deviations one must go from the mean (in both directions) to contain .95 of the scores.

It turns out that one must go 1.96 standard deviations from the mean in both directions to contain .95 of the scores. The value of 1.96 was found using a z table. Since each tail is to contain .025 of the scores, you find the value of z for which 1 - 0.025 = 0.975 of the scores are below. This value is 1.96.

All the components of the confidence interval are now known: m = 530, $Ïƒ_{m}$ = 31.62, z = 1.96.

Lower limit = 530 - (1.96)(31.62) = 468.02

Upper limit = 530 + (1.96)(31.62) = 591.98

Therefore, 468.02 â‰¤ Î¼ â‰¤ 591.98. This means that the experimenter can be 95% certain that the mean SAT in the school system is between 468 and 592. This also means if the experimenter repeatedly took samples from the population and calculated a number of different 95% confidence intervals using the sample information, on average 95% of those intervals would contain Î¼. Notice that this is a rather large range of scores. Naturally, if a larger sample size had been used, the range of scores would have been smaller.

The computation of the 99% confidence interval is exactly the same except that 2.58 rather than 1.96 is used for z. The 99% confidence interval is: 448.54 <= Î¼ <= 611.46. As it must be, the 99% confidence interval is even wider than the 95% confidence interval.

Summary of Computations

- Compute m = âˆ‘X/N
- Compute $Ïƒ_{m}$ = Ïƒ/$N^{1/2}$
- Find z (1.96 for 95% interval; 2.58 for 99% interval)
- Lower limit = m - z $Ïƒ_{m}$
- Upper limit = m + z $Ïƒ_{m}$
- Lower limit <= Î¼ <= Upper limit

Assumptions:

- Normal distribution
- Ïƒ is known
- Scores are sampled randomly and are independent

There are three other points worth mentioning here:

- The point estimate will always lie exactly at the midway mark of the confidence interval. This is because it is the "best" estimate for Î¼, and so the confidence interval expands out from it in both directions.
- The higher the percentage of confidence, the wider the interval will be. As the percentage is increased, a wider interval is needed to give us a greater chance of capturing the unknown population value within that interval.
- The width of the confidence interval is always twice the part after the positive or negative sign, that is, twice the reliability factor x standard error. The width is simply the upper limit minus the lower limit.

It is very rare for a researcher wishing to estimate the mean of a population to already know its standard deviation. Therefore, the construction of a confidence interval almost always involves the estimation of both Î¼ and Ïƒ.

Students' t-Distribution

When Ïƒ is known, the formula m - z Ïƒ$_{m}$ <= Î¼ <= m + z Ïƒ$_{m}$ is used for a confidence interval. When Ïƒ is not known, Ïƒ$_{m}$ = s/N$^{1/2}$ (N is the sample size) is used as an estimate of Ïƒ and Î¼. Whenever the standard deviation is estimated, the t rather than the normal (z) distribution should be used. The values of t are larger than the values of z, so confidence intervals when Ïƒ is estimated are wider than confidence intervals when Ïƒ is known. The formula for a confidence interval for Î¼ when Ïƒ is estimated is:

m - t s$_{m}$ <= Î¼ <= m + t s$_{m}$

where m is the sample mean, s$_{m}$ is an estimate of Ïƒ$_{m}$, and t depends on the degrees of freedom and the level of confidence.

The t-distribution is a symmetrical probability distribution defined by a single parameter known as **degrees of freedom (df)**. Each value for the number of degrees of freedom defines one distribution in this family of distributions. Like a standard normal distribution (e.g., a z-distribution), the t-distribution is symmetrical around its mean. Unlike a standard normal distribution, the t-distribution has the following unique characteristics.

- It is an estimated standardized normal distribution. When n gets larger, t approximates z (s approaches Ïƒ).
- The mean is 0 and the distribution is bell-shaped.
- There is not one t-distribution, but a family of t-distributions. All t-distributions have the same mean of 0. Standard deviations of these t-distributions differ according to the sample size, n.
- The shape of the distribution depends on degrees of freedom (n - 1). The t-distribution is less peaked than a standard normal distribution and has fatter tails (i.e., more probability in the tails).
- t$_{\text{Î±/2}}$ tends to be greater than z$_{\text{Î±/2}}$ for a given level of significance, Î±.
- Its variance is v/(v-2) (for v > 2), where v = n-1. It is always larger than 1. As v increases, the variance approaches 1.

The value of t can be determined from a t-table. The degrees of freedom for t are equal to the degrees of freedom for the estimate of Ïƒ$_{m}$, which is equal to N-1.

A portion of a t-table is presented below:

Suppose the sample size (n) is 30 and the level of significance (Î±) is 5%. df = n - 1 = 29. $t_{Î±/2} = t_{0.025} = 2.045$ (Find the 29 df row, and then move to the 0.05 column.)

Example

Assume a researcher is interested in estimating the mean reading speed (number of words per minute) of high school graduates and computing the 95% confidence interval. A sample of 6 graduates was taken; reading speeds were: 200, 240, 300, 410, 450, and 600. For these data,

- m = 366.6667
- s_m = 60.9736
- df = 6-1 = 5
- t = 2.571

Therefore, the lower limit is: m - (t) (s_m) = 209.904 and the upper limit is: m + (t) (s_m) = 523.430. Therefore, the 95% confidence interval is: 209.904 <= Î¼ <= 523.430.

Thus, the researcher can be 95% sure that the mean reading speed of high school graduates is between 209.904 and 523.430.

Summary of Computations

- Compute m = âˆ‘X/N
- Compute s
- Compute Ïƒ$_m$ = s/N$^{1/2}$
- Compute df = N-1
- Find t for these df using a t table
- Lower limit = m - t s_m
- Upper limit = m + t s_m
- Lower limit <= Î¼ <= Upper limit

Assumptions:

- Normal distribution
- Scores are sampled randomly and are independent.

Discuss the issues surrounding selection of the appropriate sample size

It's all starting to become a little confusing. Which distribution do you use?

When a large sample size (generally larger than 30 samples) is used, a z-table can always be used to construct the confidence interval. It does not matter if the population distribution is normal or if the population variance is known. This is because the central limit theorem assures us that when the sample is large, the distribution of the sample mean is approximately normal. However, the t-statistic is more conservative because it tends to be greater than the z-statistic; therefore, using a t-statistic will result in a wider confidence interval.

If there is only a small sample size, a t-table has to be used to construct the confidence interval when the population distribution is normal and the population variance is not known.

If the population distribution is not normal, there is no way to construct a confidence interval from a small sample (even if the population variance is known).

Therefore, if all other factors are equal, you should try to select a sample larger than 30. The larger the sample size, the more precise the confidence interval.

In general, at least one of the following is needed:

- a normal distribution for the population
- a sample size that is greater than or equal to 30

If one or both of the above occur, a z-table or t-table is used, dependent upon whether Ïƒ is known or unknown. If neither of the above occurs, then the question cannot be answered.

A summary of the situation is as follows:

- If the population is normally distributed and the population variance is known, use a z-score (irrespective of sample size).
- If the population is normally distributed and the population variance is unknown, use a t-score (irrespective of sample size).
- If the population is not normally distributed, and the population variance is known, use a z-score only if n >= 30; otherwise, it cannot be calculated.
- If the population is not normally distributed and the population variance is unknown, use a t-score only if n >= 30; otherwise, it cannot be calculated.

6. Resampling

```
<p> </p><b>Resampling</b> is a way to reuse data to generate new, hypothetical samples
```

- You don't know the underlying distribution for the population,
- Traditional formulas are difficult or impossible to apply,
- As a substitute for traditional methods.

The **bootstrap method** is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement.

Both bootstrapping and traditional methods use samples to draw inferences about populations. Both can estimate sampling distributions. A primary difference is how they estimate sampling distributions.

Traditional methods use properties of the sample data, the experimental design, a test statistic and need to satisfy the assumptions.

The bootstrap method involves iteratively resampling a dataset with replacement. This method takes the sample data that a study obtains, and then resamples it over and over to create many simulated samples. Each of these simulated samples has its own properties, such as the mean. When you graph the distribution of these means on a histogram, you can observe the sampling distribution of the mean. You donâ€™t need to worry about test statistics, formulas, and assumptions.

Example

Suppose a study collects five data points and creates four bootstrap samples:

The resampled datasets are the same size as the original dataset and only contain values that exist in the original set. Furthermore, these values can appear more or less frequently in the resampled datasets than in the original dataset. Finally, the resampling process is random and could have created a different set of simulated datasets.

keep in mind that bootstrapping does not create new data. Instead, it treats the original sample as a proxy for the real population and then draws random samples from it. Consequently, the central assumption for bootstrapping is that the original sample accurately represents the actual population. As the sample size increases, bootstrapping converges on the correct sampling distribution under most conditions.

The **Jackknife** works by sequentially deleting one observation in the data set, then recomputing the desired statistic. It is computationally simpler than bootstrapping. The main application is to reduce bias and evaluate variance for an estimator.

The jackknife requires n repetitions for a sample of n (for example, if you have 10,000 items then you'll have 10,000 repetitions), while the bootstrap requires "B" repetitions. This leads to a choice of B, which isn't always an easy task.

7. Data Snooping Bias, Sample Selection Bias, Look-Ahead Bias, and Time-Period Bias

As has already been mentioned, if there are problems with the choice of sample, then t

There are a number of different types of bias that can creep into samples. It is important to be aware of them and have the ability to comment on their possible appearance in the data where appropriate.

Data-snooping bias is the bias in the inference drawn as a result of prying into the empirical results of others to guide your own analysis.

Finding seemingly significant but in fact spurious patterns in data is a serious problem in financial analysis. Although it afflicts all non-experimental sciences, data-snooping is particularly problematic for financial analysis because of the large number of empirical studies performed on the same datasets. Given enough time, enough attempts, and enough imagination, almost any pattern can be teased out of any dataset. In some cases, these spurious patterns are statistically small, almost unnoticeable in isolation. But because small effects in financial calculations can often lead to very large differences in investment performance, data-snooping biases can be surprisingly substantial.

For example, after examining the empirical evidence from 1986 to 2002, Professor Minard concludes that a growth investment strategy produces superior investment performance. After reading about Professor Minard's study, Monica decides to conduct research of growth versus value investing based on the same or related historical data used by Professor Minard. Monica's research is subject to data-snooping bias because, among other things, the data used by Professor Minard may be spurious.

The best way to avoid data-snooping bias is to examine new data. However, data-snooping bias is difficult to avoid because investment analysis is typically based on historical or hypothesized data.

Data-snooping bias can easily lead to data-mining bias.

Data-mining is the practice of finding forecasting models by extensive searching through databases for patterns or trading rules (i.e., repeatedly "drilling" in the same data until you find something). It has a very specific definition: continually mixing and matching the elements of a database until one "discovers" two more or more data series that are highly correlated. Data-mining also refers more generically to any of a number of practices in which data can be tortured into confessing anything.

Two signs may indicate the existence of data-mining in research findings about profitable trading strategies:

- Many of the variables actually used in the research are not reported. These terms may indicate that the researchers were searching through many unreported variables.
- There is no plausible economic theory available to explain why these strategies work.

To avoid data-mining, analysts should use out-of-sample data to test a potentially profitable trading rule. That is, analysts should test the trading rule on a data set other than the one used to establish the rule.

Sample selection bias occurs when data availability leads to certain assets being excluded from the analysis. The discrete choice has become a popular tool for assessing the value of non-market goods. Surveys used in these studies frequently suffer from large non-response numbers, which can lead to significant bias in parameter estimates and in the estimate of mean.

Survivorship bias is the most common type of sample selection bias. It occurs when studies are conducted on databases that have eliminated all companies that have ceased to exist (often due to bankruptcy). The findings from such studies most likely will be upwardly biased, since the surviving companies will look better than those that no longer exist. For example, many mutual fund databases provide historical data about only those funds that are currently in existence. As a result, funds that have ceased to exist due to closure or merger do not appear in these databases. Generally, funds that have ceased to exist have lower returns relative to the surviving funds. Therefore, the analysis of a mutual fund database with survivorship bias will overestimate the average mutual fund return because the database only includes the better-performing funds. Another example is the return data on stocks listed on an exchange, as it is subject to survivorship bias; it's difficult to collect information on delisted companies and these companies often have poor performances.

Look-ahead bias exists when studies assume that fundamental information is available when it is not. For example, researchers often assume that a person had annual earnings data in January; in reality, the data might not be available until March. This usually biases results upwards.

Time period bias occurs when a test design is based on a time period that may make the results time-period specific. Even the worst performers have months or even years in which they look wonderful. After all, stopped clocks are right twice a day. To eliminate strategies that have just been lucky, research must encompass many years. However, if the time period is too long, the fundamental economic structure may have changed during the time frame, resulting in two data sets that reflect different relationships.

Hypothesis Testing

1. Introduction

```
   A <b>hypothesis</b> is a statement about a population created for the purpose of stati
```

Examples of hypotheses made about a population parameter are:

- The mean monthly income for a financial analyst is $5,000.
- 25% of R&D costs are ultimately written off.
- 19% of net income statements are later found to be materially incorrect.

Testing a Hypothesis

- **State the Null Hypothesis and the Alternate Hypothesis**
  The first step is to state the **null hypothesis** (designated: H_0), which is the statement that is to be tested. The null hypothesis is a statement about the value of a population. The null hypothesis will either be rejected or fail to be rejected.

  The **alternate hypothesis** is the statement that is accepted if the sample data provides sufficient evidence that the null hypothesis is false. It is designated as H_1 and is accepted if the sample data provides sufficient statistical evidence that H_0 is false.

- **Determine the Appropriate Test Statistic and its Probability Distribution**
  A **test statistic** is simply a number, calculated from a sample, whose value, relative to its probability distribution, provides a degree of statistical evidence against the null hypothesis. In many cases, the test statistic will not provide evidence sufficient to justify rejecting the null hypothesis. However, sometimes the evidence will be strong enough so that the null hypothesis is rejected and the alternative hypothesis is accepted instead.
  Typically, the test statistic will be of the general form:
  test statistic = (sample statistic - parameter value under H_0) / standard error of sample statistic

- **Select the Level of Significance**
  After setting up H_0 and H_1, the next step is to state **the level of significance**, which is the probability of rejecting the null hypothesis when it is actually true. **Alpha** is used to represent this probability. The idea behind setting the level of significance is to choose the probability that a decision will be subject to a Type I error. There is no one level of significance that is applied to all studies involving sampling. A decision by the researcher must be made to use the 0.01 level, 0.05 level, 0.10 level, or any other level between 0 and 1. A lower level of significance means that there is a lower probability that a Type I error will be made.

- **Formulate the Decision Rule**
  A **decision rule** is a statement of the conditions under which the null hypothesis will be rejected and under which it will not be rejected.
  The **critical value** (or rejection point) is the dividing point between the region where the null hypothesis is rejected and the region where it is not rejected. The region of rejection defines the location of all those values that are so large (in absolute value) that the probability of their occurrence is low if the null hypothesis is true.

- **Collect the Data, Perform Calculations**
  Once the data is collected, the test statistic should be calculated.

- **Making a Decision**
  If the test statistic is greater than the higher critical value (or, in a two-tailed test, less than the lower critical value), then the null hypothesis is rejected in favor of the alternative hypothesis.

- **Making an Investment Decision**
  Making an investment decision (or economic decision) entails not just the use of statistical decision algorithms but economic considerations as well.

2. Null Hypothesis and Alternative Hypothesis

```
    The <b>null hypothesis</b> (designated H<sub>0</sub>) is the statement that is to be t
```

- For example, the null hypothesis could be: "The mean monthly return for stocks listed on the Vancouver Stock Exchange is not significantly different from 1%." Note that this is the same as saying the mean (Î¼) monthly return on stocks listed on the Vancouver Stock Exchange is equal to 1%. This null hypothesis, H_0, would be written as: H_0: Î¼ = 1%.

- As another example, if a null hypothesis is stated as "There is no difference in the revenue growth rate for satellite TV dishes before and after the negative TV advertising campaign aired by the cable industry," then the null hypothesis could be written to show that two rates are equal: H_0: $r_1 = r_2$.

It is important to point out that accepting the null hypothesis does not prove that it is true. It simply means that there is not sufficient evidence to reject it.

Note that it makes no sense to hypothesize about known sample values, for the simple reason that they are known, just like it makes no sense to construct confidence intervals or obtain point estimates for known values. Hypothesis tests are carried out on unknown population parameters.

The **alternate hypothesis** is the statement that is accepted if the sample data provides sufficient evidence that the null hypothesis is false. It is designated as H_1 and is accepted if the sample data provides sufficient statistical evidence that H_0 is false.

The following example clarifies the difference between the two hypotheses. Suppose the mean time to market for a new pharmaceutical drug is thought to be 3.9 years. The null hypothesis represents the current or reported condition and would therefore be H_0: Î¼ = 3.9. The alternate hypothesis is that this statement is not true, that is, H_1: Î¼ â‰ 3.9. The null and alternative hypotheses account for all possible values of the population parameter.

There are three basic ways of formulating the null hypothesis.

- H_0: Î¼ = Î¼$_0$ versus H_1: Î¼ â‰ Î¼$_0$. This hypothesis is two-tailed, which means that you are testing evidence that the actual parameter may be statistically greater or less than the hypothesized value.

- H_0: â‰¤ Î¼$_0$ versus H_1: Î¼ > Î¼$_0$. This hypothesis is one-tailed; it tests whether there is evidence that the actual parameter is significantly greater than the hypothesized value. If there is, the null hypothesis is rejected. If there is not, the null hypothesis is accepted.

- H_0: Î¼ â‰¥ Î¼$_0$ versus H_1: Î¼ < Î¼$_0$. This hypothesis is one-tailed; it tests whether there is evidence that the actual parameter is significantly less than the hypothesized value. If there is, the null hypothesis is rejected. If there is not, the null hypothesis is accepted.

The question most likely to be raised at this point is how do you know if a test is one-sided or two-sided? The general rule is as follows:

- If a question makes it clear that only one direction is to be examined, use a one-sided test.
- If there is no clue in the question as to which direction should be examined, use a two-sided test.

Normally, there is little ambiguity; the question will make it clear which test should be used. A question often asked involves testing whether a population mean is greater than or less than a specific number. In this case, use a one-tailed test. If the question asks you to test whether a population mean is different from a specific number, use a two-tailed test.

With practice, you'll see that this issue is not really a huge problem.

3. Test Statistic and Significance Level

```
    A <b>test statistic</b> is simply a number, calculated from a sample, whose value, rel
```

The value of the test statistic is the focal point of assessing the validity of a hypothesis. Typically, the test statistic will be of the general form:

For example, a test statistic for the mean of a distribution (such as the mean monthly return for a stock index) often follows a standard normal distribution. In such a case, the test statistic requires use of the z-test, P(Z <= test statistic = z). This is shown as:

X-bar = sample mean

Î¼$_0$ = hypothesized value

Ïƒ = sample standard deviation

n = sample size

Note that this assumes the population variance (and, therefore, the population standard deviation) is unknown and can only be estimated from the sample data.

There are other probability distributions as well, such as the t-distribution, the chi-square distribution, and the F-distribution. Depending on the characteristics of the population and the sample, a test statistic may follow one of these distributions, which will be discussed later.

After setting up H_0 and H_1, the next step is to state the level of significance, which is the probability of rejecting the null hypothesis when it is actually true. Alpha (Î±) is used to represent this probability. The idea behind setting the level of significance is to choose the probability that any decision will be subject to a Type I error. There is no one level of significance that is applied to all studies involving sampling. A decision by the researcher must be made to use the 0.01 level, 0.05 level, 0.10 level, or any other level between 0 and 1. A lower level of significance means that there is a lower probability that a Type I error will be made.

4. Type I and Type II Errors in Hypothesis Testing

```
    Because hypothesis tests are heavily dependent on the samples used as "evidence," it i
```

When a hypothesis is tested, there are four possible outcomes:

- Reject the null hypothesis when it's false. This is a correct decision.
- Incorrectly reject the null hypothesis when it's correct. This is known as a **Type I error**. The probability of a Type I error is designated by the Greek letter alpha (Î±) and is called the Type I error rate.
- Don't reject the null hypothesis when it's true. This is a correct decision.
- Don't reject the null hypothesis when it's false. This is known as a **Type II error**. The probability of a Type II error (the Type II error rate) is designated by the Greek letter beta (Î²).

A Type II error is only an error in the sense that an opportunity to reject the null hypothesis correctly was lost. It is not an error in the sense that an incorrect conclusion was drawn, since no conclusion is drawn when the null hypothesis is not rejected. It has nothing to do with Î±, other than the fact that it moves in the opposite direction from Î± (that is, the bigger the one, the smaller the other).

A Type I error, on the other hand, is an error in every sense of the word. A conclusion is drawn that the null hypothesis is false when, in fact, it is true. Therefore, Type I errors are generally considered more serious than Type II errors. The probability of a Type I error (Î±) is called the **significance level**; it is set by the experimenter. For example, a 5% level of significance means that there is a 5% probability of rejecting the null hypothesis when it is true.

There is a tradeoff between Type I and Type II errors. The more an experimenter protects him or herself against Type I errors by choosing a low level, the greater the chance of a Type II error. Requiring very strong evidence to reject the null hypothesis makes it very unlikely that a true null hypothesis will be rejected. However, it increases the chance that a false null hypothesis will not be rejected, thus lowering its power. The Type I error rate is almost always set at 0.05 or at 0.01, the latter being more conservative (since it requires stronger evidence to reject the null hypothesis at the 0.01 level then at the 0.05 level).

To reduce the probabilities of both types of errors simultaneously, the sample size n must be increased.

5. The Power of a Test

```
<b>Power</b> is the probability of correctly rejecting a false null hypothesis. Power
```

Sometimes more than one test statistic is used to conduct a hypothesis test. In this case the relative power of the test needs to be computed for the competing statistics; the test statistic that is *most powerful* must be selected.

If you want to know more about "power of a test," read the following (not required for Level I candidates):

Consider a hypothetical experiment designed to test whether rats brought up in an enriched environment can learn mazes faster than rats brought up in the typical laboratory environment (the control condition). Two groups of 12 rats are tested. Although the experimenter does not know it, the population mean number of trials it takes to learn the maze is 20 for rats from the enriched environment and 32 for rats from the control condition. The null hypothesis that the enriched environment makes no difference is therefore false.

The question is, *What is the probability that the experimenter is going to be able to demonstrate that the null hypothesis is false by rejecting it at the 0.05 level?* This is the same as asking, *What is the power of the test?* Before the power of the test can be determined, the standard deviation (s) must be known. If s = 10, then the power of the significance test is 0.80. This means that there is a 0.80 probability that the experimenter will be able to reject the null hypothesis. Since power = 0.80, b = 1 - 0.80 = 0.20.

It is important to keep in mind that power is not about whether or not the null hypothesis is true (it is assumed to be false). It is the probability the data gathered in an experiment will be sufficient to reject the null hypothesis. The experimenter does not know that the null hypothesis is false. The experimenter asks the question: *If the null hypothesis is false with specified population means and standard deviation, what is the probability that the data from the experiment will be sufficient to reject the null hypothesis?*

If the experimenter discovers that the probability of rejecting the null hypothesis is low (power is low), even if the null hypothesis is false to the degree expected (or hoped for), then it is likely that the experiment should be redesigned. Otherwise, considerable time and expense will go into a project that has little chance of being conclusive even if the theoretical ideas behind it are correct.

6. The Decision Rule

```
   A <b>decision rule</b> is a statement of the conditions under which the null hypothesi
```

The **critical value** (or **rejection point**) is the dividing point between the region where the null hypothesis is rejected and the region where it is not rejected. The region of rejection defines the location of all those values that are so large (in absolute value) that the probability of their occurrence is low if the null hypothesis is true.

In general, the decision rule is that:

- If the magnitude of the calculated test statistic *exceeds* the rejection point(s), the result is considered *statistically significant* and the null hypothesis (H_0) should be rejected.
- Otherwise, the result is considered *not statistically significant* and the null hypothesis (H_0) should *not be rejected.*

The specific decision rule varies depending on two factors:

- The distribution of the test statistic.
- Whether the hypothesis test is one-tailed or two-tailed.

For the following discussion, you assume that the test statistic follows a standard normal distribution. Therefore:

- The calculated value of the test statistic is a standard normal variable, denoted by z.
- Rejection points are determined from the standard normal distribution (z-distribution).

The Decision Rule for a Two-Tailed Test

A two-tailed hypothesis test for the population mean (Î¼) is structured as H_0: Î¼ = Î¼$_0$ vs. H_1: Î¼ â‰ Î¼$_0$. For a given level of significance (Î±), the total probability of a Type I error must sum to Î±, with Î±/2 in each tail. Therefore, the decision rule for a two-tailed test is:

Reject H_0 if z < -$z_{\text{Î±/2}}$ or z > $z_{\text{Î±/2}}$

where $z_{\text{Î±/2}}$ is chosen such that the probability of z > $z_{\text{Î±/2}}$ is Î±/2.

For a two-sided test at the 5% significance level, the total probability of a Type I error must sum to 5%, with 5%/2 = 2.5% in each tail. As a result, the two rejection points are -$z_{0.025}$ = -1.96 and $z_{0.025}$ = 1.96. These values are obtained from the z-table. Therefore, the null hypothesis can be rejected if z < -1.96 or z > 1.96, where z is the calculated value of the test statistic. If the calculated value of the test statistic falls within the range of -1.96 and 1.96, the null hypothesis cannot be rejected.

The Decision Rule for a One-Tailed Test

A one-tailed hypothesis test for the population mean (Î¼) is structured as H_0: Î¼ â‰¤ Î¼$_0$ vs. H_1: Î¼ > Î¼$_0$. For a given level of significance (Î±), the probability of a Type I error equals Î±, all in the right tail. Therefore, the decision rule for a one-tailed test is:

Reject H_0 if z $z_{\text{Î±}}$

>

where z$_{Î±}$ is chosen such that the probability of z > z$_{Î±}$ is Î±.

For a test of H_0: Î¼ â‰¤ Î¼$_0$ vs. H_1: Î¼ > Î¼$_0$ at the 5% significance level, the total probability of a Type I error equals 5%, all in the right tail. As a result, the rejection point is $z_{0.05}$ = 1.645. Therefore, the null hypothesis will be rejected if z > 1.645, where z is the calculated value of the test statistic. If z â‰¤ 1.645, the null hypothesis cannot be rejected.

If a one-tailed hypothesis test for the population mean (Î¼) is structured as H_0: Î¼ â‰¥ Î¼$_0$ vs. H_1: Î¼ < Î¼$_0$, similar analysis can be performed (i.e., reject H_0 if z < -z$_{Î±}$).

Statistical Decision vs. Economic Decision.

Making a statistical decision involves using the stated decision rule to determine whether or not to reject the null hypothesis. A **statistical decision** is based solely on statistical analysis of the sample information. The **economic** or **investment decision** takes into consideration not only the statistical decision, but also all economic issues pertinent to the decision. Slight differences from a hypothesized value may be statistically significant but not economically meaningful (taking into account transaction costs, taxes, and risk).

For example, it may be that a particular strategy has been shown to be statistically significant in generating value-added returns. However, the costs of implementing that strategy may be such that the added value is not sufficient to justify the costs required. Thus, while statistical significance may suggest a particular course of action is optimal, the economic decision also takes into account the costs associated with the strategy, and whether the expected benefits justify implementing the strategy.

Researchers frequently report the p-value associated with a particular test in order to present a fuller picture.

The P-Value Approach to Hypothesis Testing

Note: This is not required by the study guide.

If you go back to subject a, you will see that seven steps in any hypothesis testing procedure are listed. These seven steps compose the process that needs to be carried out in order to perform a test.

However, there is an alternative and widely used approach to hypothesis testing, which will be discussed here. This is not a different method, but rather a different way of presenting the results of a test. The new approach is known as the p-value approach to hypothesis testing.

Before proceeding with the new approach, a summary of the seven steps encountered earlier:

Step 1: State the hypotheses.

Step 2: Identify the test statistic and its probability distribution.

Step 3: Specify the significance level.

Step 4: State the decision rule.

Step 5: Collect the data in the sample and calculate the necessary value(s) using the sample data.

Step 6: Make a decision regarding the hypotheses.

Step 7: Make a decision based on the test results.

When a p-value approach is used, steps 3 and 4 become redundant. The new steps are as follows:

Step 1: State the hypotheses.

Step 2: Identify the test statistic and its probability distribution.

Step 3: Collect the data in the sample and calculate the necessary value(s) using the sample data.

Step 4: Calculate the p-value for the test.

Step 5: Make a decision regarding the hypotheses.

Step 6: Make a decision based on the test results.

It can thus be seen that the p-value approach is simpler, and also provides useful information about a test statistic. So, by now you are probably anxious to find out: What exactly is a p-value?

P-value is the area from the test statistic to the end of the tail (or tails, in the case of a two-sided test) of interest. The "p" stands for probability, and the p-value is a probability. Effectively, it represents the chance of obtaining a test statistic whose value is at least as extreme as the one just calculated in the test, when H_0 is true.

The term "tail of interest" means the right-hand tail in a > test, the left-hand tail in a < test, and both tails in a â‰ test.

It should be clear to you that if the test statistic is very extreme (that is, close to the tail) then the area from that value to the end of the tail will be very small. By definition, the p-value is thus very small. In this case, the test statistic would certainly fall in the rejection region.

Conversely, if the test statistic is not extreme (that is, it is far from the tail) and would thus fall in the acceptance region, the area from the test statistic to the end of the tail would be fairly large. In this case, the p-value is fairly large.

So, you have just learned how to interpret p-values. A small p-value indicates a rejection of the null hypothesis whereas a large p-value indicates a non-rejection of the null hypothesis.

This is all well and good, but how large is large?

P-value is normally compared with our usual Î± value, so 0.05 is a common reference point. This means that, for p-values smaller than 0.05, H_0 would be rejected, and for p-values larger than 0.05, H_0 would not be rejected.

However, p-value provides not only a reference point but an idea about just how strong or weak this rejection or non-rejection is.

For example, a p-value of 0.0003 would indicate a very strong rejection of H_0 at virtually any a-level, whereas a p-value of 0.048 would indicate a rejection of H_0 when compared with a value of 0.05 but a non-rejection of H_0 when compared with an Î± value of 0.01.

In general, the larger the p-value, the smaller the likelihood that H_0 will be rejected and vice versa.

In the case of a two-sided test, if the area from the test statistic to the end of the right tail is 0.04, for example, then the p-value will be 0.08; doubling the area is necessary because both tails are being used. It's as simple as that.

Examples

Now try the following for yourself. In each case, determine whether or not you would reject H_0 when compared with an Î± value of 0.05.

You are given the following p-values.

A. 0.02

B. 0.456

C. 0.053

D. 0.0001

The correct answers are:

A. Reject H_0.

B. Do not reject H_0: The p-value is very large (over 45%), and therefore you are very far away from a rejection of H_0.

C. Do not reject H_0: The p-value is only just more than 0.05, so you are close to a rejection.

D. Reject H_0: The p-value is very small, so you would reject H_0 in almost all cases.

Now that you understand how p-values work, let us go through the testing procedure again:

- First write down your hypotheses, then determine which test statistic to use. There is no need to have a significance level, and hence, you don't have a specifically demarcated rejection region.
- Calculate the test statistic and determine the resultant p-value.
- The conclusion of the test is based on your p-value. Recall that a small p-value means a rejection of H_0, whereas a large p-value means a non-rejection of H_0.

Note: P-values can be worked out for all continuous statistical distributions, no matter what the shape of the distribution is. So, this method can be used for z-graphs, t-graphs, x_2-graphs and F-graphs. You will encounter the x_2- and F-distributions later on in this chapter.

Confidence Interval and Hypothesis Testing

A **confidence interval** is a range of values within which it is believed that a certain unknown population parameter (often the mean) will fall with a certain degree of confidence. The percentage of confidence is denoted (1 - Î±)%.

A **test of significance** is a test to ascertain whether or not the value of an unknown population parameter is as stated by an individual or institution. This test is carried out at a significance level of Î±, which determines the size of the rejection region.

Note how the confidence interval is related to the test statistic. They are linked by the rejection point(s).

- The confidence interval is:
- Recall that the test statistic is:

The likelihood that the z-value will be less than the test statistic is what is being tested. Setting up this inequality, and rearranging, the test is:

Confidence intervals can be used to test hypotheses. Note that the right side of the equation is the left endpoint of the confidence interval. Essentially, if the confidence interval contains the value of the unknown population parameter as hypothesized under H_0, then H_0 would not be rejected in a two-sided hypothesis test with corresponding Î±; if the confidence interval does not contain the value of the unknown population parameter as hypothesized under H_0, then H_0 would be rejected in a two-sided hypothesis test with corresponding Î±.

The reason for the above determination is as follows:

- If the hypothesized value does not fall in the confidence interval, then there is a very small chance that the value can be a true value for the unknown parameter, so H_0 is almost certainly wrong and will be rejected.

- If the hypothesized value does fall in the confidence interval, then there is a very good chance that the value can be a true value for the unknown parameter, so H_0 is definitely possible and will not be rejected.

This comparison can be used only for two-sided tests, not one-sided tests, because confidence intervals cannot be linked with one-sided tests.

7. Multiple Tests and Interpreting Significance

```
    <p> </p>A type I error is where you incorrectly reject the null hypothesis; In other w
```

The FDR is the expected ratio of the number of false positive classifications (false discoveries) to the total number of positive classifications (rejections of the null). The total number of rejections of the null include both the number of false positives (FP) and true positives (TP). Simply put, FDR = FP / (FP + TP).

For a more humorous (an perhaps understandable) look at the problem, take a look at XKCD's "Jelly Bean Problem."(https://xkcd.com/882) The comic shows a scientist finding a link between acne and jelly beans, when a hypothesis was tested at a 5% significance level. Although there is no link between jelly beans and acne, a significant result was found (in this case, a jelly bean caused acne) by testing multiple times. Testing 20 colors of jelly beans, 5% of the time there is 1 jelly bean that is incorrectly fingered as being the acne culprit. The implications for false discovery in hypothesis testing is that if you repeat a test enough times, youâ€™re going to find an effect, but that effect may not actually exist.

Example

In medical testing, the false discovery rate is when you get a â€œpositiveâ€ test result but you donâ€™t actually have the disease.

Out of 10,000 people given the test, there are 450 true positive results (box at top right) and 190 false positive results (box at bottom right) for a total of 640 positive results. Of these results, 190/640 are false positives so the false discovery rate is 30%.

Adjusting the FDR

If you repeat a test enough times, you will always get a number of false positives. One of the goals of multiple testing is to control the FDR: the proportion of these erroneous results. For example, you might decide that an FDR rate of more than 5% is unacceptable. Note though, that although 5% sounds reasonable, if youâ€™re doing a lot of tests, you'll also get a large number of false positives; for 1000 tests, you could expect to get 50 false positives by chance alone. This is called the multiple testing problem, and the FDR approach is one way to control for the number of false positives.

The FDR approach adjusts the p-value for a series of tests. A p-value gives you the probability of a false positive on a single test; If you're running a large number of tests from small samples, you should use q-values instead.

- A p-value of 5% means that 5% of all tests will result in false positives.
- A q-value of 5% means that 5% of significant results will be false positives.

Although controlling for type I errors sound ideal (why not just set the threshold really low and be done with it?), Type I and Type II errors form an inverse of relationship; when one goes down, the other goes up and vice versa. By decreasing the false positives, you increase the number of false negatives - that's where there is a real effect, but you fail to detect it.

8. Tests Concerning a Single Mean

When testing a hypothesis concerning the value of a population mean, either a t-test c

- A **t-test** is a hypothesis test that uses a t-statistic, which follows a t-distribution.
- A **z-test** is a hypothesis that uses a z-statistic, which follows a z-distribution (a standard normal distribution).

Deciding when to use a t-test or a z-test depends on three factors:

- Distribution of population: normal or non-normal
- Population variance: known or unknown
- Sample size: large or small

When to use the t-test?

This test should be used if the population variance is unknown and either of the following conditions holds:

- The sample size is large (in general, n â‰¥ 30).
- The sample size is small (n â‰¤ 30) but the population is normally distributed or approximately normally distributed.

For a t-test concerning a single population mean (Î¼), the test statistic to be used is the t-statistic with n - 1 degrees of freedom.

How to make the decision to conduct a t-test?

- Calculate the t-statistic with n - 1 degrees of freedom.
- Use a t-table to find the rejection point(s) at the specified level of significance with n - 1 degrees of freedom.

  A portion of a t-table is presented below:

  Note: This t-table is a combination of a one-tailed and a two-tailed table. On the exam, you may be asked to conduct a two-tailed hypothesis test using a one-tailed table or a one-tailed test using a two-tailed table.
- Compare the calculated value of the t-statistic with the rejection point(s) to make the decision.
  - For a **two-tailed** test (H_0: Î¼ = Î¼$_0$ versus H_a: Î¼ â‰ Î¼$_0$), there are **two** rejection points. Reject the null if the t-statistic is greater than the upper rejection point or less than the lower rejection point.
  - For a one-tailed test (H_0: Î¼ â‰¤ Î¼$_0$ versus H_a: Î¼ > Î¼$_0$), there is only **one** rejection point: the **upper** rejection point. Reject the null if the t-statistic is **greater** than the upper rejection point.
  - For a one-tailed test (H_0: Î¼ â‰¥ Î¼$_0$ versus H_a: Î¼ < Î¼$_0$), there is only **one** rejection point: the **lower** rejection point. Reject the null if the t-statistic is **less** than the upper rejection point.

When to use the z-test?

The z-test should be used if the population variance is **normally distributed** with **known** variance.

For a z-test concerning a single population mean (Î¼), the test statistic to be used is the z-statistic with n degrees of freedom.

To conduct a z-test, compare the calculated value of the z-statistic with rejection point(s) at the specified level of significance. Rejection points are obtained using a z-table.

Most Frequently Used Rejection Points for a z-test:

For a population with unknown variance, it is acceptable to use the z-test if the sample size is large.

- The **t-test** should be used if the population variance is **unknown** and the sample size is **large**.
- In this case, it is also acceptable to use the z-test because of the central limit theorem. Recall that according to the central limit theorem, if the sample size is sufficient large, the sampling distribution of the sample mean will be approximately normally distributed.
- The z-statistic is computed below:

If you are interested, there is a detailed explanation of this last point at the end of this subject (the z-test and the central limit theorem).

Summary

- In practice, the population variance is typically unknown.
- The table below summarizes tests concerning the population mean when the population has unknown variance.

Example

The Jones Fund has been in existence for 20 years. Monthly returns over this period are approximately normally distributed. A random sample of 20 monthly returns shows that the fund has achieved a mean monthly return of 2%. The sample standard deviation of monthly returns is 3%. The Jones Fund was expected to have earned a 1.5% mean monthly return over the 20-year period. Using a 5% significance level, determine if the fund's actual performance is consistent with the expected mean monthly return of 1.5%.

Solution

1. State the hypothesis. Let Î¼ be the mean monthly return on the Jones Fund. The null and alternative hypotheses are stated as:

- H$_0$: Î¼ = Î¼$_0$ (The mean monthly return is 1.5%.)
- H$_a$: Î¼ â‰ Î¼$_0$ (The mean monthly return is not 1.5%.)

Note that this is a two-tailed test.

1. Identify the test statistic. The t-statistic should be used because the population variance is unknown, the sample is small (less than 30) and the population is approximately normally distributed.

2. Specify the level of significance (Î±); Î± is given at 5%.

3. State the decision rule. You need to find the two rejection points for this two-tailed test in a t-table.

A portion of a one-tailed t-table is given below:

- The degree of freedom = n - 1 = 19.
- Since this is a **two-tailed test** and you are given a **one-tailed t-table**, we need to divide the 5% level of significance by 2: 0.05/2 = 0.025.
- In the one-tailed t-table, find the 19 df row and then move to the 0.025 column. The entry of 2.093 is the right-tail rejection point. Since the t-distribution is symmetrical, the left-tail rejection point is -2.093.
- Thus, the decision rule is stated as reject H$_0$ if $t < -2.093$ or $t > 2.093$.

1. Compute the test statistic: $t_{19} = (0.02 - 0.015) / [0.03 / 20^{1/2}] = 0.745$.

2. Make the statistical decision. Because 0.745 does not satisfy either t < -2.093 or t > 2.093, you do not reject the null hypothesis.

Thus, it is reasonable to believe that the actual performance of the Jones Fund is consistent with the expected mean monthly return of 1.5%.

The z-test and the central limit theorem.

When hypothesis testing a population mean, there are generally two options for the test statistic:

- t_{n-1} = (x-bar - Î¼$_0$)/(s/n$^{1/2}$): when the population variance is unknown and must be estimated from the sample.
- z = (x-bar - Î¼$_0$)/(Ïƒ/n$^{1/2}$): when the population variance is known.

The first statistic may be used if either the sample is large (n = 30 or greater) or, if n < 30, it may be used if the sample is at least approximately normally distributed. In most cases, this will be the statistic used, because in most practical problems, the population variance is not known with certainty.

The second statistic is sometimes used with large sample sizes, because the central limit theorem implies that the distribution of a sample mean will be approximately normally distributed as the sample size increases.

Recall that the degrees of freedom in a t-distribution depend on sample size and are generally defined as n-1. This means that as the sample size increases, so the degrees of freedom increase, and a t-graph begins to resemble a z-graph.

In fact, for infinity degrees of freedom (a theoretical concept, because it is not possible to have an infinite sample size), the two graphs are identical, and critical values for a z-distribution can also be found on a t-table in the row that has infinity as its degrees of freedom.

What actually happens is that, as the degrees of freedom increase, the tails of a t-graph flatten out; the graph becomes more peaked in the center and its standard deviation approaches 1 from above. The graph thus begins to resemble a z-graph in all aspects.

As can be seen above, when the degrees of freedom are low, the graph is fairly flat in the center and has long tails and a bigger standard deviation. As the degrees of freedom increase, the tails become narrower and flatter and the graph peaks in the center.

Recall also that the area under any graph is 1, as this area represents a probability. So, as the degrees of freedom increase, the area that is "lost" in the flatter tails is "found" in the center of the graph, and that is why the graph becomes more peaked. However,

- There are differences between the t-test critical values and the z-test critical values (these can be significant but get smaller with large samples).
- The t-test is still the theoretically proper choice unless the population variance is known.

9. Tests Concerning Differences between Means with Independent Samples

```
In practice, analysts often want to know whether the means of two populations are equa
```

If it is reasonable to believe that the samples are from populations at least approximately normally distributed and that the samples are also independent of each other, whether a mean value differs between the two populations can be tested. The test procedure is the same as before. There are just a couple of modifications that need to be made.

As mentioned previously, the null hypothesis involves an equal sign. So, in this situation, the null hypothesis would be that the two unknown population means are equal. The alternative hypothesis would involve one of >,

< or â‰ .

The rest of the testing procedure is the same, but the test statistic is different. It's now time to look at what formula should be used. Be warned, though, that the formulae in this section are horrific.

The test statistic to be used in this section is a t-value, but it varies based on the assumptions. The assumption has been made throughout that the population means are normally distributed.

1. Test statistic for a test of the difference between two population means (normally distributed populations, population variances unknown but assumed equal):

where $s_p^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / (n_1 + n_2 - 2)$ is a pooled estimator of the common variance. The number of degrees of freedom is $n_1 + n_2 - 2$. Normally, the degrees of freedom are given by n - 1, but here there are two samples. Combine the sample sizes and then subtract 1 for each sample, or 2 in total. This gives $n_1 + n_2 - 2$ as the degrees of freedom.

1. Test statistic for a test of the difference between two population means (normally distributed populations, unequal and unknown population variances):

where tables are used to show the t-distribution using "modified" degrees of freedom computed with the formula:

A practical tip is to compute the t-statistic before computing the degrees of freedom. Whether the t-statistic is significant will sometimes be obvious.

Example

From a class of Science students, a sample of 36 is drawn; the mean grade is found to be 62% with a standard deviation of 10. From a class of Arts students, a sample of 49 is drawn; the mean grade is found to be 59% with a standard deviation of 9.6. Assuming that the grades in both classes have a normal distribution, and that the population variances are equal, test at the 5% level for a statistically significant difference in the mean grades of the two classes.

Note that since the sample standard deviations are 10 and 9.6 respectively, the assumption that the population variances are equal seems valid. Had there been a big discrepancy in the sample values, the assumption of equality of population variances would have carried less weight.

"1" will be used to represent the Science students and "2" to represent the Arts students.

Step 1: State the hypotheses.

You are testing for differences in the population means Î¼$_1$ and Î¼$_2$. Since the question does not specify a direction, a two-sided test is appropriate. The hypotheses are therefore:

H_0: Î¼$_1$ - Î¼$_2$ = 0

H_a: Î¼$_1$ - Î¼$_2$ â‰ 0

Step 2: Identify the test statistic and its probability distribution.

The appropriate test statistic is the one that assumes equal variances. It is the t-value (discussed earlier).

Step 3: Specify the significance level.

You are told to test at the 5% level, so Î± = 0.05.

Step 4: State the decision rule.

This is a two-sided test, so you need to split your area equally between both tails. You thus have 2.5% of area in each. The total degrees of freedom are: $n_1 + n_2 - 2 = 36 + 49 - 2 = 83$. Your tables don't have 83 degrees of freedom, so use 80, which is the closest value to 83. From the t-table, the critical values are therefore -1.99 and 1.99.

The value above determines your decision. If your test statistic lies to the left of -1.99 or to the right of 1.99, you will reject H_0; otherwise, you will not reject H_0.

You might notice that your critical values of -1.99 and 1.99 are very close to the corresponding z-values of -1.96 and 1.96. This is because, as explained earlier, as the degrees of freedom increase, the t-values approach the z-values. Since 83 degrees of freedom is a large number, the t-graph here closely resembles a z-graph.

Step 5: Collect the data in the sample and calculate the necessary value(s) using the sample data.

The question gives you the necessary sample values: x-bar$_1$ = 62, x-bar$_2$= 59, $s_1 = 10$, $s_2 = 9.6$, $n_1 = 36$ and $n_2 = 49$.

Recall the formula from earlier: s^2_p = [35 x 100 + 48 x 92.16] / 83 = 95.466. Now you can substitute into the test statistic: = (62 - 59 - 0) / = 1.3988.

The t-value of 1.3988 is now compared with your critical values of -1.99 and 1.99.

Step 6: Make a decision regarding the hypotheses.

Since the value of the test statistic is less extreme (i.e., closer to zero) than the positive critical value, the test statistic falls in the acceptance region. You would thus not reject H_0 at the 5% significance level.

Step 7: Make a decision based on the test results.

You can now conclude that the difference in the average marks of Science and Arts students is not significant when testing at the 5% level.

Note:

- Had you used a p-value approach here, you would have obtained a p-value between 0.1 and 0.2 (closer to 0.2). You can check this for yourself as an exercise. You cannot obtain the p-value exactly from t-tables. Because the value is larger than 0.05, you would not reject H_0, so your results are consistent with those above.
- Had you made Science population 2 and Arts population 1, your test statistic would have worked out to be -1.3988, but this would have made no difference in your conclusion, since the test is two-sided. It is better, however, to be guided by the question in this regard.

10. Tests Concerning Differences between Means with Dependent Samples

```
  In the previous subject, t-tests were examined to discern differences between two popu
```

In this subject, the focus will be on conducting a test based on the means of samples that are related in some way. The data are arranged in paired observations; the test is sometimes known as a **paired comparisons test**. The paired observations are either in or not in the same units.

This test is normally used in two cases:

- a before-and-after situation, where analysts compare data before and after a certain

process/procedure/treatment has taken place.
- when there is a relationship between the values, for example, collecting data from twins.

In both cases, the data in each pair of observations are dependent.

This method involves forming differences by subtracting one value of the pair from the other. The sample is then reduced to a single sample and the test statistic is based on the values of the differences.

In this situation, use the subscript d to indicate the differences being dealt with.

The hypotheses are therefore:

- H_0: Î¼$_d$ = Î¼$_{d0}$
- Ha: Î¼$_d$ â‰ Î¼$_{d0}$

where Î¼$_{d0}$ is some fixed value, commonly zero, about which you are hypothesizing.

The test statistic is then:

where d-bar and $s_{d\text{-bar}}$ are calculated from the sample of differences in the usual way.

Thus, and

Also, because the sample has been reduced to a single sample of size n, the test statistic has n-1 degrees of freedom.

Example

A program that is believed to improve I.Q. levels has been offered to primary school children. I.Q. levels are believed to be normally distributed. A group of 100 children participated in the program. The mean difference in their I.Q. levels (after minus before) was found to be 2, with a standard deviation of 11. Test whether the program is effective at the 5% significance level.

This is clearly a paired comparisons test, as you are given a before-and-after situation. Note also that you are told that the population is normally distributed. However, since the sample size is so large (n =100), this fact is academic, because the central limit theorem states that the sample mean will be normally distributed in any case. (This is just a point for you to take note of.)

Step 1: State the hypotheses.

Although it is perfectly acceptable to work with differences as before-minus-after, you will take the more conventional approach of after-minus-before, because this is how the data are presented.

Note also that you wish to test whether the program is effective, that is, whether I.Q. levels increase after the program. You therefore have a "greater than" test. Under H_0, assume that the program is not effective, so the value of Î¼$_{d0}$ = 0.

The hypotheses are therefore - H_0: Î¼$_d$ = 0 versus Ha: Î¼$_d$ > 0.

Step 2: Identify the test statistic and its probability distribution.

The population is normally distributed and you are dealing with paired comparisons, so a paired comparisons t-test is appropriate here.

Step 3: Specify the significance level.

You are told to test at the 5% level, so Î±= 0.05.

Step 4: State the decision rule.

This is a greater-than test, so the full 5% of the area goes in the right tail. Also, the degrees of freedom are: n-1 = 99. The tables don't have 99 degrees of freedom, so use 100, which is the closest value to 99. From the t-table, the critical value is therefore 1.66.

The value above determines your decision. If your test statistic lies to the right of 1.66, you will reject H_0; otherwise, you will not reject H_0.

You might notice that the critical value of 1.66 is very close to the corresponding z-value of 1.645. This is because, as explained earlier, as the degrees of freedom increase, so the t-values approach the z-values. Since 100 degrees of freedom is a large number, the t-graph here closely resembles a z-graph.

Step 5: Collect the data in the sample and calculate the necessary value(s) using the sample data.

The question gives you the necessary sample values: d-bar = 2, $s_{d\text{-}bar}$ = 11 and n = 100.

The test statistic is: $t = (2 - 0)/(11/100^{1/2}) = 1.818$.

Note that Î¼$_{d0}$ is the value of Î¼d under H_0, and is thus zero.

The t-value of 1.818 is now compared with your critical value of 1.66.

Step 6: Make a decision regarding the hypotheses.

Since the value of the test statistic is more extreme (i.e., further away from zero) than the critical value, you see that the test statistic falls in the rejection region. You would thus reject H_0 at the 5% significance level.

Step 7: Make a decision based on the test results.

You can now conclude that the program is effective in increasing I.Q. levels at the 5% significance level.

Note:

- Had a p-value approach been used here, a p-value between 0.025 and 0.05 would have been obtained. (You can check this for yourself as an exercise.) You cannot obtain the p-value exactly from t-tables. Because this value is smaller than 0.05, you would reject H_0, so your results are consistent with the above.
- Had you worked with before-minus-after, then the value would have been -2, and our test statistic would have been -1.818. Your alternative hypothesis would then have needed to be <, as you are testing whether levels before are smaller than levels after. The critical value would then have been -1.66 and you would still have landed in the rejection region, making no difference to your conclusion. However, doing the test this way seems less logical than the method used above. Either works, though, with the correct modification.

The key issue now is that you understand when to use a test for independence and when to use a paired comparisons test.

To help clarify the issue for you:

- A test of the differences in means (as conducted in subject i) is used when there are two independent samples. Essentially, you have two separate groups, and you wish to compare their population means.
- A test of the mean of the difference (as conducted in subject j) is used when the samples are dependent, either because you have a before-and-after situation or because there is an inherent relation between the pairs.

In the first case, you keep the groups completely separate and combine their sample sizes for the purpose of calculating degrees of freedom.

In the second case, you reduce the two samples to a single sample of differences and treat the entire process from then on as if you were dealing with a single sample.

Another telltale sign is to look at sample sizes. If the sample sizes are different, the test has to be for independent samples, as paired comparisons tests require equal-sized samples. If the sample sizes are the same, either test could be used.

What the two procedures have in common is that they both require normally distributed populations and they both make use of t-tests.

Until now, the examples have been testing means of populations, and have made use of z-tests and t-tests. Now it's time to look at the procedure for testing variances of populations and to introduce two new statistical distributions: the chi-square distribution and the F-distribution.

11. Testing Concerning Tests of Variances (Chi-Square Test)

```
Suppose an analyst is interested in testing whether the variance from a single populat
```

H$_0$: Ïƒ^2 = Ïƒ$_0^2$, versus H$_1$: Ïƒ^2 â‰ Ïƒ$_0^2$

Also note that directional hypotheses could be made instead:

H$_0$: Ïƒ^2 â‰¤ (Ïƒ$_0$)^2, versus H$_1$: Ïƒ^2 > (Ïƒ$_0$)^2, or H$_0$: Ïƒ^2 â‰¥ (Ïƒ$_0$)^2, versus H$_1$: Ïƒ^2 < (Ïƒ$_0$)^2

The test statistic to be used is a chi-square (Ï‡^{2}) statistic with n-1 degrees of freedom.

The formula for the test statistic is:

n = sample size

s^2 = sample variance

Ïƒ$_0^2$ = the hypothesized value for Ïƒ under H$_0$

Unlike t-graphs and z-graphs, a chi-square graph is positively skewed. It is also truncated at zero, and thus is not defined for negative values.

Like the family of t-graphs, the shape of the graph varies; the graph becomes more symmetrical as the degrees of freedom increase.

The graph looks like this:

Critical values for this distribution are found in the chi-square table. The table is titled "Probability in Right Tail," so be aware that this is the region the table gives and modify your calculation accordingly.

It is possible to use both tails in this distribution, despite the fact that it is non-symmetrical.

In contrast to the t-test, the chi-square test is sensitive to violations of its assumptions. If the sample is not actually random or if it does not come from a normally distributed population, inferences based on a chi-square test are likely to be faulty.

Example

You are an institutional investor evaluating a hedge fund that seeks to deliver a return comparable to the domestic broad market equity index while keeping the monthly standard deviation in asset value under 5%. According to data in the prospectus, during the last 30 months the monthly standard deviation in asset value was 4.5%. You wish to test this claim statistically, using a significance level of 0.10. Assume that returns are normally distributed and monthly asset values are independent observations.

The hypotheses are:

H_0: Ïƒ^2 â‰¤ 0.0025, versus H_1: Ïƒ^2 > 0.0025. Note that by squaring the standard deviation, 5% = 0.05, you get 0.0025.

The test statistic will be chi-square with 30 - 1 = 29 degrees of freedom.

The critical value will be found in the table for a chi-square distribution, 29 degrees of freedom, alpha = 0.10. Note that this is a one-tailed test. The critical value is 39.087.

The test statistic will be: chi-square = $[(29) \times 0.045^2] / (0.05)^2 = 23.49$.

The test statistic is not greater than the critical value, so you do not reject the null hypothesis.

Note: If the null hypothesis had a >= symbol in it, then you would reject if the test statistic was less than or equal to the lower alpha point.

The chi-square distribution is asymmetrical. Like the t-distribution, the chi-square distribution is a family of distributions; a different distribution exists for each possible value of degrees of freedom, n - 1 (n is sample size). Unlike the t-distribution, the chi-square distribution is bounded below by 0 (no negative values).

12. Tests Concerning the Equality of Two Variances (F-Test)

```
  Just as in the case of means of populations, now it's time to look at the test for equ
```

The test statistic is , where $(s_1)^2$ and $(s_2)^2$ are the sample variances from two samples, which have n_1 and n_2 observations in each. The samples are random, independent of each other, and generated by normally distributed populations.

The F stands for Fisher, the name of the person who formulated this distribution.

This test statistic will have n_1 - 1 degrees of freedom in the numerator and n_2 - 1 degrees of freedom in the denominator.

The values n_1 - 1 and n_2 - 1 are also the divisors used in calculating $(s_1)^2$ and $(s_2)^2$, respectively.

Another point of interest is that, under H_0, the ratio Ïƒ$_1{}^2$/Ïƒ$_2{}^2$ is 1.

A F-test checks whether or not the ratio of the test statistic, $(s_1)^2/(s_2)^2$, is close to 1. If it is, you will obtain a non-rejection of H_0; if not, you will obtain a rejection of H_0.

For this reason, it is common practice to use the larger of the two ratios $(s_1)^2/(s_2)^2$ or $(s_2)^2/(s_1)^2$, as the actual test statistic. So, whichever of $(s_1)^2$ and $(s_2)^2$ is larger should go in the numerator, with the smaller number in the denominator.

If these values are equal, it makes no difference which value goes on top, as the test statistic will be 1 in either case.

Like the chi-square distribution:

- the F-distribution is bounded below by zero, and therefore values of the F-distribution test statistic cannot be negative.
- the F-distribution is asymmetrical.
- the F-distribution is a family of distributions.
- the F-test is not very robust when assumptions are violated.

Unlike the chi-square distribution, the F-distribution is determined by two parameters: the degrees of freedom in the numerator and the degrees of freedom in the denominator. Note that F(8,6) refers to an F-distribution with 8 numerator and 7 denominator degrees of freedom.

Another area in which F-tests differ from chi-square tests is that, for F-tests, the rejection region is always in the right tail of the graph, never the left tail.

This presents few difficulties. It does mean that if you are conducting a one-sided test, the full value goes into the right tail, whereas if you are conducting a two-sided test, you divide by 2 to obtain the area in the right tail. The remaining half of the area effectively falls away because of the convention of putting the larger sample variance on top in the test statistic. Provided you do this, you will have no problems here.

Critical values for this distribution are found in F-tables. In order to find the required value, you should look up the value that corresponds with the correct numerator and denominator degrees of freedom.

Example

You are an investor skeptical of the value added by Jupiter Fund, an actively managed large-cap mutual fund. An examination of the prospectus reveals that it has tracked its target index fund very closely over the past 10 years, yet it incurs a 1.51% annual expense fee, far in excess of what one might pay to invest in a large-cap index fund. You wonder if Jupiter achieves a smaller standard deviation in asset values to justify its higher costs.

You obtain monthly net asset values for both Jupiter Fund and the market index for the past 10 years. You find that Jupiter fund has a sample variance of 0.0016 and that the market index has a sample variance of 0.0025. At the alpha = 0.05 significance value, is the variance for Jupiter statistically less than that for the market index fund? Assume independence of observations and normally distributed populations.

Solution

Your null hypothesis is:

H_0: (Ïƒ$_1$)^2 â‰¤ (Ïƒ$_2$)^2 versus H_1: (Ïƒ$_1$)^2 > (Ïƒ$_2$)^2 where Jupiter Fund is population 2. This may seem counter-intuitive; the explanation follows below.

The test is a one-tailed F-test, with 119 degrees of freedom in both the numerator and the denominator. Using the convention of placing the higher variance term in the numerator, you get an F-value of 0.0025/0.0016 = 25/16 = 1.5625.

Note that by putting the higher variance term in the numerator, you are, in effect, saying that the market index is population 1. If the statistic is large, then you are finding evidence for the alternative hypothesis, that (Ïƒ$_1$)^2 > (Ïƒ$_2$)^2.

Your critical value will be found with the area in the right tail equal to 0.05. The table in Appendix D contains entries for 60 and 120 degrees of freedom; you can interpolate, but the difference is so small that you can just use the values at 120 degrees of freedom. The critical value is thus 1.35, and your test statistic exceeds this. Thus, you will reject the null hypothesis and conclude that the market index does have significantly larger variance (or alternatively, that Jupiter Fund has significantly smaller variance).

13. Parametric vs. NonParametric Tests

```
    In hypothesis tests, analysts are usually concerned with the values of parameters, suc
```

All hypotheses tests that have been considered in this section are parametric tests.

For example, an F-test relies on two assumptions:

- Populations 1 and 2 are normally distributed.
- Two random samples drawn from these populations are independent.

The F-test is concerned with the difference between the variance of the two populations. Variance is a parameter of a normal distribution. Therefore, the F-test is a parametric test.

There are other types of hypothesis tests, which may not involve a population parameter or much in the way of assumptions about the population distribution underlying a parameter. Such tests are **nonparametric tests**.

Nonparametric tests have different characteristics:

- They are concerned with quantities other than parameters of distributions.
- They can be used when the assumptions of parametric tests do not hold for the particular data under consideration.
- They make minimal assumptions about the population from which the sample comes. A common example is the situation in which an underlying population is not normally distributed. Other tests, such as a median test or the sign test, can be used in place of t-tests for means and paired comparisons, respectively.

Nonparametric tests are normally used in three cases:

- When the distribution of the data to be analyzed indicates or suggests that a parametric test is not appropriate.
- When the data are ordinal or ranked, as parametric tests normally require the data to be interval or ratio. One might be ranking the performance of investment managers; such rankings do not lend themselves to parametric tests because of their scale.
- When a test does not involve a parameter. For instance, in evaluating whether or not an investment manager has had a statistically significant record of consecutive successes, a nonparametric runs test might be employed. Another example: if you want to test whether a sample is randomly selected, a nonparametric test should be used.

In general, parametric tests are preferred where they are applicable. They have stricter assumptions that, when met, allow for stronger conclusions. However, nonparametric tests have broader applicability and, while not as precise, do add to your understanding of phenomena, particularly when no parametric tests can be effectively used.

14. Tests Concerning Correlation

```
    <h4>Pearson's Correlation Coefficient</h4>
```

Pearson's correlation coefficient is a measure of the linear correlation (dependence) between two variables X and Y, giving a value between +1 and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is negative correlation. It tells us about the strength of the linear relationship between x and y points on a regression plot.

The correlation coefficient p measures the strength of a linear relationship in *samples only*. If we obtained a different sample, we would obtain a different correlation and therefore a potentially different conclusion. As always, we want to draw conclusions about populations, not just samples. To do so, we have to conduct a hypothesis test.

The hypothesis test lets us decide whether the value of the population correlation coefficient is "close to 0" or "significantly different from 0" based on the sample correlation coefficient r and the sample size n.

A t-test can be used to test the population correlation coefficient H_0: $p = 0$.

Next, we calculate the value of the test statistic using the following formula:

This is a t-statistic and operates in the same way as other t tests. Calculate the t-value and compare that with the critical value from the t-table at the appropriate degrees of freedom and the level of confidence you wish to maintain.

- If the calculated value is in the tail then cannot accept the null hypothesis that there is no linear relationship between these two independent random variables.
- If the calculated t-value is NOT in the tailed then cannot reject the null hypothesis that there is no linear relationship between the two variables.

Example

A financial analyst estimates the sample correlation between the US dollar and Canadian dollar monthly returns to be 0.4782 over the period of January 2020 to December 2021. At 5% significant level, would you reject the null hypothesis that the population correlation equals to 0?

There are 24 months. The test statistic $t^* = 0.4782\ (24 - 2)^{1/2}/(1\text{-}r^2)^{1/2} = 2.55$. Using the t-distribution table, the critical value is given by $t_{0.025,22} = 2.074$. Since the test statistic is larger than the critical value, we reject the null hypothesis.

A quick shorthand way to test correlations is the relationship between the sample size and the correlation. If |r| â‰¥ 2/âˆšn, then this implies that the correlation between the two variables demonstrates that a linear relationship exists and is statistically significant at approximately the 0.05 level of significance. In the above example, $0.4782 > 2/24^{1/2} = 0.4083$, so the correlation coefficient is significantly larger than 0.

As the formula indicates, there is an inverse relationship between the sample size and the required correlation for significance of a linear relationship. With only 10 observations, the required correlation for significance is 0.6325, for 30 observations the required correlation for significance decreases to 0.3651 and at 100 observations the required level is only 0.2000.

The Spearman Rank Correlation Coefficient

One critical assumption is that the y values for any particular x value are normally distributed about the line. If they are not, we can use Spearman rank correlation coefficient, which is calculated on the ranks of two variables within their respective samples.

Spearman's correlation measures the strength of a monotonic relationship. Basically, this means that if one variable increases (or decreases), the other variable also increases (or decreases). It is a statistical measure of the strength of a monotonic relationship (whether linear or not) between paired data. Its interpretation is similar to that of Pearsons, e.g. the closer the r is to 1, the stronger the monotonic relationship.

15. Test of Independence Using Contingency Table Data

```
<p> </p>When analysis of categorical data is concerned with more than one variable, tw
```

Effects in a contingency table are defined as relationships between the row and column variables; that is, are the levels of the row variable differentially distributed over levels of the column variables? Significance in this hypothesis test means that interpretation of the cell frequencies is warranted. Non-significance means that any differences in cell frequencies could be explained by chance.

Hypothesis tests on contingency tables are based on a statistic called chi-square.

The procedure used to test the significance of contingency tables is similar to all other hypothesis tests. That is, a statistic is computed and then compared to a model of what the world would look like if the experiment was repeated an infinite number of times when there were no effects. In this case the statistic computed is called the chi-square statistic. For the detailed steps please refer to the textbook example.

The chi-square statistic has degrees of freedom of (r - 1) (c - 1), where r is the number of categories for the first variable and c is the number of categories of the second variable.

Introduction to Linear Regression

1. Simple Linear Regression

```
    <p> </p><b>Linear regression</b> is used to quantify the linear relationship between t
```

The variable being studied is called **dependent variable** (or response variable). A variable that influences the dependent variable is called an **independent variable** (or a factor).

For example, you might try to explain small-stock returns (the dependent variable) based on returns to the S&P 500 (the independent variable). Or you might try to explain inflation (the dependent variable) as a function of growth in a country's money supply (the independent variable).

Regression analysis begins with the dependent variable (denoted Y), the variable that you are seeking to explain. The independent variable (denoted X) is the variable you are using to explain changes in the dependent variable. Regression analysis is trying to measure a relationship between these two variables. It tries to measure how much the dependent variable is affected by the independent variable. The regression equation as a whole can be used to determine how related the independent and dependent variables are.

It should be noted that the relationship between the two variables can never be measured with certainty. There always exists other variables that are unknown that may have an effect on the dependent variable.

2. Estimating the Parameters of a Simple Linear Regression

```
    <p> </p>The following regression equation explains the relationship between the depenc
```

$$Y_i = b_0 + b_1X_i + e_i,\ i = 1, 2, ..., n$$

There are two regression coefficients in this equation:

- The Y-intercept is given by b_0; this is the value of Y when X = 0. It is the point at which the line cuts through the Y-axis.
- The slope coefficient, b_1, measures the amount of change in Y for every one unit increase in X.

Also note the error term, denoted by e_i. Linear regression is a technique that finds the best straight-line fit to a set of data. In general, the regression line does not touch all the scatter points, or even many; the error terms represent the differences between the actual value of Y and the regression estimate.

Suppose a regression formula is, $y_i = -23846 + 0.6942 \text{ x } x_i$, where the large constant and the x terms are on a date scale for a spreadsheet program (e.g., Sept 1, 1997 = 35674).

In the regression equation above, the b_1 term, 0.6942, represents the slope of the regression line. The slope, of course, represents the ratio of the vertical rise in the line relative to the horizontal "run." The slope and the intercept (b_0, here -23846) represent parameters calculated by the regression process. The process of linear regression calculates the parameters as those that minimize the squared deviations of the actual data values from the estimated values obtained using the regression equation. The process of determining these parameters involves calculus.

Linear regression involves finding a straight line that fits the scatter plot best. To determine the slope and intercept of the regression line, one must either use a statistical calculator, a software program, or calculate it by hand. Typically, an analyst will not have to calculate regression coefficients by hand.

- For regression equations with only one independent variable, the estimated slope, b_1, will equal the covariance of X and Y divided by the variance of X.

b_1 = Cov(X,Y) / (Ïf_x)^2 = Cov(X,Y)/Var(X)

- The intercept parameter, b_0, can then be determined by using the average values of X and Y in the regression equation and solving for b_0.

The "hat" above the b terms means that the value is an estimate (in this case, from the regression).

The equation that results from the linear regression is called **linear equation**, or the **regression line**.

- Regression line: Y-bar$_i$ = b_0(hat) + b_1(hat) X_i
- Regression equation: $Y_i = b_0 + b_1X_i + e_i$, i = 1, 2, ..., n

The linear equation is derived by applying a mathematical method called the **least squares regression**, also known as the **sum of the squared errors**: the sum of the squared distances of all the points in the observation away from the mean.

Note that analysts never observe the *actual parameter values* b_0 and b_1 in a regression model. Instead they observe only the *estimated values* b_0 and b_1. All of their prediction and testing must be based on the estimated values of the parameters rather than their actual values.

After you calculate the b_0 and b_1, test their significance. Also note that the regression analysis does not prove anything; there could be other unknown variables that affect both X and Y.

3. Assumptions of the Simple Linear Regression Model

```
    Classical normal linear regression assumptions (<b>LINE</b>):<p> </p><ul class="notes"
```

- **L** (Linearity). A linear relation exists between Y_i and X_i. This means the mean value for Y at each level of X falls on the regression line.
- **I** (Independence). The error terms e_i are independent of the values of the independent variable X. That is, thereâ€™s no connection between how far any two points lie from the regression line.
- **N** (Normality). For any value x, the error term has a normal distribution.
- **E** (Homoscedasticity). The variance of the error term (e_i), denoted Ïf^2, is the same for all x. That is, the spread in the Yâ€™s for each level of X is the same.

These assumptions are depicted in the following figure.

How do we check these assumptions? We examine the variability left over after we fit the regression line. We simply graph the residuals and look for any unusual patterns.

If a linear model makes sense, the residuals will:

- have a constant variance;
- be approximately normally distributed (with a mean of zero), and
- be independent of one another.

If the assumptions are met, the residuals will be randomly scattered around the center line of zero, with no obvious pattern. The residuals will look like an unstructured cloud of points, centered at zero.

If there is a non-random pattern, the nature of the pattern can pinpoint potential issues with the model.

For example, if curvature is present in the residuals, then it is likely that there is curvature in the relationship between the response and the predictor that is not explained by our model. A linear model does not adequately describe the relationship between the predictor and the response.

In this example, the linear model systematically over-predicts some values (the residuals are negative), and under-predict others (the residuals are positive).

Example

The yield of wheat per acre for the month of July is thought to be related to the rainfall. A researcher randomly selects acres of wheat and records the rainfall and bushels of wheat per acre.

- Dependent variable: Yield of wheat measured in bushels per acre for July.
- Independent variable: Rainfall measured in inches for July.

Analysis:

- L: The mean yield per acre is linearly related to rainfall.
- I: Field yields are independent; knowing one (X, Y) pair does not provide information about another.
- N: The yields for a given amount of rainfall are normally distributed.
- E: The standard deviation of yields is approximately the same for each rainfall level.

We may encounter problems with the linearity assumption if mean yields increase initially as the amount of rainfall increases after which excess rainfall begins to ruin crop yield. The random selection of fields should assure independence if fields are not close to one another.

Cross Sectional vs. Time-Series Regressions

Cross sectional datasets are those where we collect data on entities only once. For example we collect IQ and GPA information from the students at any one given time (think: camera snap shot).

Time-Series dataset is one where we collect information from the same entity (the same company, asset class, investment fund) over time (think: video).

In cross sectional datasets we do not need to worry about independence assumption. It is â€œassumedâ€ to be met. â€“ We worry about "independence" when we have a time-series dataset.

4. Analysis of Variance

```
    <h4>Sum of Squares</h4>
```

The **sum of squares total**, denoted SST, is the squared differences between the observed dependent variable (Y$_i$) and its mean (Y-bar). You can think of this as the dispersion of the observed variables around the mean - much like the variance in descriptive statistics.

SST = âˆ‘(Y$_i$ - Y-bar)^2

The **sum of squares regression**, or SSR, is the sum of the differences between the predicted value (Y-hat$_i$) and the mean of the dependent variable.

$$\text{SSR = âˆ‘(Y-hat}_i\text{ - Y-bar)}^2$$

It is the explained sum of squares, and tells you how much of the variation in the dependent variable your model explained.

The **sum of squares error**, or SSE, is the difference between the observed value and the predicted value. It measures dispersion of values around regression line, and is the unexplained variability by the regression.

- The smaller the SEE, the more accurate the regression line to predict the value of the dependent variable.

- It is very much like the standard deviation for a single variable, except that it measures the standard deviation of the residual term in the regression.
- If the actual values for the observations are close to the regression line, it means that the relationship between the dependent and independent variables are strong. However, it does not tell us how well the independent variable explains variation in the dependent variable.

Mathematically,

$$\text{SST = SSR + SSE.}$$

The total variability of the data set is equal to the variability explained by the regression line plus the unexplained variability, known as error.

Measures of Goodness of Fit

The **coefficient of determination** is a measure of the proportion of total variance in the dependent variable (Y) that can be explained by variation in the independent variable (X). It is a measure of the goodness of fit of the regression line.

$$R^2 = \text{Explained variation/Total variation} = 1 - \text{Unexplained variation/total variation}$$

The higher the R^2, the better. R^2 values range from 0 to 1.

The easiest way to calculate the coefficient of determination is to square the correlation coefficient, if that is known. Thus, the coefficient of determination is often denoted R^2. This method, however, only works when there is just one independent variable. When there are two or more independent variables in linear regression, analysts need to use the formula R^2 = SSR/SST. R^2 is a ratio of the explained sum of squares to the total sum of squares.

One test statistic that is helpful in identifying significant variables is the F-statistic. The F-statistic can be a blunt tool in multiple regression, in that it tests the null hypothesis that all of the independent variables are equal to zero. That is, the hypotheses are structured as follows: H_0: $b_1 = b_2 = ... = b_k = 0$, and H_a: at least one b_i <> 0. If the null is not rejected, all the slope coefficients are equal to 0, and therefore all slope coefficients are unimportant for predicting Y. You can see that F-test measures how well the regression equation explains the variation in the dependent variable. In simple linear regression, however, it simply tests whether the slope of the one independent variable is zero: H_0: $b_1 = 0$ vs. H_a: b_1 â‰ 0.

To develop the F-test statistic, analysts need several items:

- The total number of observations, n.
- The number of parameters estimated (in simple linear regression, this is two: the intercept and the slope of the regression line).
- The sum of the squared errors (SSE), also known as the residual sum of the squares: SSE = âˆ‘$[Y_i - Y_i\text{(hat)}]^2$.

- The regression sum of the squares (RSS), which is the sum of the squared deviations of the regressed values of Y around the average value of Y: RSS = âˆ‘[Y_i(hat) - Y-bar]^2.

The F-statistic for one independent variable will be:

Note that the F-statistic will have n-1 degree of freedom in the numerator (again, when there is only one independent variable) and n-2 degrees of freedom in the denominator.

A large test statistic (in absolute value) implies rejection of the null hypothesis and therefore a non-zero value for the slope of the regression line. A large F-statistic also implies that the regression model explains much of the variation in the dependent variable. However, for simple linear regression with one independent variable, F-statistics are rarely used, because they contain the same information as the t-statistic. In fact, the F-statistic is simply the square of the t-statistic in such regressions (note, this is not true for multiple regression).

Anova and Standard Error of Estimate in Simple Linear Regression

The **analysis of variance (ANOVA)** method can help in identifying which independent variables are of significant value to the regression model. Compared with the coefficient of determination, an ANOVA procedure provides more details about the sources of the variations.

Typically, the ANOVA procedure is performed using statistical software packages. The output of the ANOVA procedure is an ANOVA table.

Here is an example of an ANOVA table:

The formula for the **standard error of estimate** for a linear regression model with one independent variable is:

{âˆ‘[(Y_i - Î± - Î² X_i)^2/(n - 2)]}$^{1/2}$

n - 2 is called the degrees of freedom: it is the denominator needed to ensure that the estimated standard error of estimate is unbiased.

Note that the errors are actually squared, averaged, and then the square root is taken. Consequently, large outliers will tend to have a greater impact on the value of the standard error than if a simple average was taken. Note too that the "averaging" is done by dividing by (n - 2). This is because the linear regression estimates two parameters, so divide by (n - 2) to make sure the calculated standard error of estimate is unbiased.

In finance world the standard error of estimate is also called **unsystematic variation**.

5. Hypothesis Testing of Linear Regression Coefficients

```
<h4>Hypothesis Tests of the Slope Coefficient</h4>
```

We frequently are interested in testing whether knowledge o an independent variable X is useful in explaining the values of Y. For example, we may want to test whether a linear relationship exists between X and Y. If X and Y are not linearly related, then in the population regression line $E(Y_i) = b_0 + b_1X_i$, we should have $b_1 = 0$. If $b_1 = 0$, the values of X are of no use in predicting Y, and the population regression line will be a horizontal line. If we reject the hypothesis that $b_1 = 0$, then we are saying that the values of X are helpful in predicting Y.

The null hypothesis does not always have the form H_0: $b_1 = 0$, although this is by far the most frequent case.

Suppose an economist claims that, in the U.S., annual income is related to years of education and the slope of the population regression line is approximately $b_1 = \$2000$. That is, the economist claims that an increase of one year in education tends to be associated with an increase of approximately \$2000 in annual income. Suppose we want to test the null hypothesis that the slope of the population regression line is $b_1 = \$2000$. Then we would test the null hypothesis H_0: $b_1 = \$2000$ against the two sided alternative hypothesis H_1: b_1 â‰ \$2000.

Test of hypothesis concerning the value of b_1 are based on the fact that if the basic assumptions of the simple linear regression model hold, then the random variable t = (b-hat$_1$ - B_1)/Sb-hat$_1$ follows the student t distribution with n - 2 degrees of freedom.

The two key measures are the standard error of the parameter and the critical value for the t-distribution associated with the t-test of statistical significance.

- S_{b1} = sqrt [âˆ‘(y_i - y-hat$_i$)^2 / (n - 2)] / sqrt [âˆ‘(x_i - x-bar)^2]
- The t-distribution requires that the number of degrees of freedom be known. For a linear regression with two parameters estimated (the two parameters are the slope and intercept), the number of degrees of freedom is (n - 2), where n is the number of observations. (Generally, the number of degrees of freedom equals the number of observations less the number of parameters estimated.)

The decision rule: reject H_0 in favor of H_1 if t < -t$_{\text{Î±}/2,\ n-2}$ or if t > t$_{\text{Î±}/2,\ n-2}$.

Regarding the calculated t-statistic used to test whether the slope coefficient is equal to zero:

- It is equal to the t-statistic to test whether the pairwise correlation is zero: t = r sqrt (n - 2) / sqrt (1 - r^2)
- It is related to the F-distributed test statistic: $t^2 = F$

Hypothesis Tests of the Intercept

Conducting hypothesis tests and calculating confidence intervals for the intercept parameter b_0 is not done as often as it is for the slope parameter b_1. The reason for this becomes clear upon reviewing the meaning of b_0. The intercept parameter b_0 is the mean of the responses at x = 0.

We can perform the hypothesis testing in a similar manner. One thing to note is the equation for the standard error of the intercept is different. However, there's no need to memorize it.

Hypothesis Tests of Slope When Independent Variable Is an Indicator Variable

A **dummy (indicator) variable** takes on 1 and 0 only. The number 1 and 0 have no numerical (quantitative) meaning. The two numbers are used to represent groups. In short dummy variable is categorical (qualitative).

For instance, we may have a sample (or population) that includes both female and male. Then a dummy variable can be defined as D = 1 for female and D = 0 for male. Such a dummy variable divides the sample into two subsamples (or two sub-populations): one for female and one for male.

Hypothesis testing can be performed in a similar manner when independent variable is a dummy variable.

6. Prediction Using Simple Linear Regression and Prediction Intervals

```
  Suppose we have the following regression equation for the Ivory Tower Mutual Fund.<p>
```

$Y_t = 0.0009 + 0.6154 \text{ x } X_{(1,t)} + 0.3976 \text{ x } X_{(2,t)}$

Where:

$X_{(1,t)}$ = return on the small-cap value index

$X_{(2,t)}$ = return on the small-cap growth index

Suppose further that we know that in a given month, the small-cap value index returned -1.2% and the small-cap growth index returned -1.9%. What is the predicted return for the Ivory Tower Mutual Fund?

The return is simply 0.0009 + 0.6154 (-0.012) + .3976 (-0.019) = -1.4%.

Note that although the prediction of an individual value Y_i and the estimate of the mean value $E(Y_i \mid X_i)$ are the same for a given value X_i, the sampling errors associated with the two predictions are different! The estimate of the mean value of Y_i does not require an estimate of the random error e_i. The confidence interval for the individual Y is always wider than that for mean value.

We often wish to have a confidence interval for the dependent variable for which we are performing the regression. Obtaining a confidence interval for the dependent variable is a bit more complicated because there are multiple sources of variation. First, there is the variation from the standard error of estimate (SEE) in the regression equation. A second source of uncertainty comes from the variation in the estimate of the regression parameters themselves. If these were known with certainty, then the predicted values of the dependent variable would have variance equal to the squared standard error of estimate. However, the parameters themselves are not known with certainty.

The formula for the variance in the prediction error, $(s_f)^2$, of Y, is:

$(s_f)^2 = s^2$ x {1 + 1/n + [(X - X-bar)^2] / [(n - 1) x (Ïƒ$_x$)^2]}

Where:

s^2 = squared standard error of estimate

n = number of observations

X = value of the independent variable used for the specific prediction of Y

X-bar = mean of the independent variable across the observations

(Ïƒ$_x$)^2 = sample variance of the independent variable

Once the variance in the prediction error, $(s_f)^2$, is known, the confidence interval for the dependent variable Y is constructed in a very similar way to the construction of confidence intervals around parameters. The confidence interval will be: Y-hat +or- (t_c) x (s_f)

Where Y-hat is the predicted value of Y, based on a particular value of X, t_c is the critical value for the t-statistic for a significance level of alpha, and s_f is the standard error of the predicted value.

Example

Suppose we are predicting the excess return on GE stock for the next month using the formula: $R_{GE} - R_f$ = Î±$_{GE}$ + Î²$_{GE}$ x $(R_m - R_f)$ + error

You are given the following data regarding 24 monthly observations of the excess return on GE stock and on the S&P 500 over the risk-free rate:

- The mean excess return on the S&P 500 during the observation period was -0.6751%.
- The variance of the excess return on the S&P 500 during the observation period was 0.2453%.
- The standard error of estimate is 0.0622.
- You expect the excess return on the S&P 500 to be 0.1580% next month.
- The Î± for GE stock is 1.68%; the Î² for GE stock is 1.3487.

Find the 95% confidence interval for the excess return on GE stock next month.

Solution

First, we predict the value of the dependent variable for next month: Y-hat = R_{GE} - R_f, the excess return on GE stock. Using the formula and data above we have: 0.0168 + 1.3487(.00158) = 1.893%

Second, compute the variance of the error in this prediction.

$(s_f)^2 = s^2$ x {1 + 1/n + [(X -

X-bar)^2] / [(n-1)x $(s_x)^2$]} =

0.0622^2 x {1 + 1/24 + [(0.00158-(-0.006751)]^2 / [(24

- 1) x (0.002453)]} = 0.42483%

The standard deviation of this forecast error is the square root of this

number, or 6.5179%.

Third, find the critical value for the t-statistic. We will have 22 degrees of freedom, and the value for a 95% confidence interval will be 2.074.

Fourth, calculate the prediction interval. It will be: 1.893% +or- 2.074 x (6.5179%) = [-11.63%, 15.41%]

This is a fairly large interval for just one month of excess return. Clearly, the large standard deviation in forecast error is a big factor in the size of the interval. It is also due in part to having only 24 observations. By collecting a greater number of observations, we could probably lower the standard deviation of forecast error (and, slightly, the critical value for t).

7. Functional Forms for Simple Linear Regression

```
    <p> </p>In the field of investments, the estimated parameters and other relationships
```

If the relationship between the independent variable and the dependent variable is not linear, we can often transform one of both of these variables to convert this relation to a linear form, which then allows the use of simple linear regression.

The **log-lin model**: logarithmic dependent variable - linear independent variable. The model is typically used when the variables may have an exponential growth relationship. For example, if you put some cash in a saving account, you expect to see the effect of compounding interest with an exponential growth of your money! The original model in these types of scenarios isnâ€™t linear in parameters, but a log transformation generates the desired linearity.

The **lin-log model**: linear dependent variable - logarithmic independent variable. This model is typically used when the impact of the independent variable on the dependent variable decreases as the value of the independent variable increases.

The **log-log model**: logarithmic dependent variable - logarithmic independent variable.

The key to selecting the correct form is to examine the goodness of fit measures:

- the coefficient of determination (R^2)
- the F-statistic
- the standard error of the estimate (s_e)

- patterns in the residuals

Economics

Economics (1)

Topics in Demand and Supply Analysis

1. Demand Concepts

```
   The demand function represents buyers' behavior.<p> </p>
```

The quantity demanded of good X depends on its price, consumers' income, and the price of good Y, etc.

Prices influence consumers' purchase decisions. The demand function can be depicted as a negatively sloped demand curve.

- If all other factors are equal, as the price of a good rises, consumer demand falls. This is mainly due to the availability of **substitutes**, which are goods that perform similar functions.
- As the price of a good falls, consumer demand rises.

Therefore, there is an *inverse* relationship between the price of a good and the amount that consumers are willing to buy. The **demand curve** normally slopes *downward.* It tells the analyst the quantity that consumers are willing to buy for each possible price when all other influences on consumers' planned purchases remain the same.

Example 1

Refer to the graph below. What is the quantity of cassettes demanded when their price is $4.00 per week?

Answer: Two cassettes per week. The demand curve tells how much is demanded at each price. To determine the quantity demanded, find $4.00 on the vertical axis and read across until you meet the demand curve. Then read the quantity from the horizontal axis.

When any factor that influences buying plans, other than the price of the good, changes, there is a change in demand for that good. When the quantity of the good that people plan to buy changes at each and every price, there is a new demand curve. These factors include changes in income, number of consumers in the market, changes in the price of a related good, etc.

Example 2

Assume the graph below reflects demand in the automobile market. Which arrow best captures the impact of increased consumer income on the automobile market?

Answer: D. Income is a shift factor of demand. An increase in income increases the number of automobiles demanded at each price. Therefore demand has shifted to the right.

- When demand *increases*, the quantity that people plan to buy *increases* at each and every price, so the demand curve shifts *rightward.*
- When demand *decreases*, the quantity that people plan to buy *decreases* at each and every price, so the demand curve shifts *leftward.*

2. Elasticities of Demand

```
   Elasticity means "responsiveness."  The elasticity of demand measures the responsivene
```

Price elasticity of demand is the percentage change in the quantity of a product demanded divided by the percentage change in the price causing the change in quantity. It indicates the degree of consumer response to variation in price. Specifically, it tells the analyst the percentage change in the quantity demanded for a good caused by a 1% increase in the price of that good.

The change in price is expressed as a percentage of the average price - the average of the initial and new price, and the change in the quantity demanded is expressed as a percentage of the average quantity demanded - the average of the initial and new quantity. Using the average price and average quantity, the same elasticity value is obtained regardless of whether the price rises or falls.

The measure is *units-free* because it is a ratio of two percentage changes and the percentages cancel each other out. Changing the units of measurement of price or quantity leave the elasticity value the same.

Because a change in price causes the quantity demanded to change in the opposite direction, *this ratio is always negative*, although economists always ignore the sign and simply use the absolute value. It is the magnitude, or absolute value, of the measure that reveals how responsive the quantity change has been to a price change.

Example 1

A Pizza Hut store can sell 50 pizzas per day at $7 each or 70 pizzas per day at $6 each. The price elasticity is: [(50 - 70)/60] / [(7 - 6) / 6.5] = -2.17.

Own-Price Elasticity of Demand

Demand can be inelastic, unit elastic, or elastic, and can range from zero to infinity. (Note: the negative sign is ignored.)

- If the elasticity coefficient is greater than 1, demand is **elastic**. A small price change leads to a large change in the quantity demanded. The more elastic the demand, the flatter the demand curve over any specific range. Demands for goods with many substitutes (e.g., juice) are relatively elastic. If the demand curve of a good is completely horizontal, the demand is **perfectly elastic**. Consumers will buy all of that good at the market price.
- When the elasticity coefficient is less than 1, demand is **inelastic**. The more inelastic the demand, the steeper the demand curve. Demands for goods with few substitutes (e.g., cigarettes) are relatively inelastic.
- When the elasticity coefficient is equal to 1, demand is said to be **unitary elastic**.

Because elasticity is a relative concept, the elasticity of a straight-line demand curve will differ at each point along the demand curve. Specifically, a straight-line demand curve is more elastic when price is high. Note that the elasticity is not the slope of the demand curve. Elasticity is used since it is independent of the units of measure.

Example 2

Refer to the graph below. Which of the following is true?

A. Areas C and E are smaller than area A, so demand must be elastic between $10 and $30.

B. Areas C and E are smaller than area A, so demand must be inelastic between $10 and $30.

C. Area F is smaller than areas B and C, so demand must be inelastic between $10 and $30.

Answer: C. Since at $30 the demand is unit elastic, at prices below $30 demand is inelastic. This is because when price rises from $10 to $30, the revenue gained is greater than the revenue lost.

The Factors that Influence the Elasticity of Demand

The elasticity of demand among products varies substantially. The determinants of price and income elasticity of demand are:

- **The closeness of substitutes.** The most important determinant is the availability of substitutes. The closer the substitutes for a good or service, the more elastic the demand for it.

    - Necessities, such as food or housing, generally have inelastic demand.
    - Luxuries, such as exotic vacations, generally have elastic demand.

  When good substitutes for a product are available, a price rise induces many consumers to switch to other products. For example, when the price of apples rises, many consumers simply switch to oranges or other fruits. However, when the price of gasoline rises, most consumers can only slightly cut back their consumption of gasoline, since there is no good substitute for gasoline.

- **The proportion of income spent on the good.** If expenditures on a product are quite small relative to a consumer's budget, the income effect will be small even if there is a substantial increase in the price of the product. This will make the demand less elastic. For example, if the price of matches triples, consumers would not bother to find substitutes, since they only spend a few bucks on matches each year.

- **The time elapsed since a price change.** The more time consumers have to adjust to a price change, or the longer a good can be stored without losing its value, the more elastic the demand for that good. In general, when the price of a product increases, consumers will reduce their consumption by a larger amount in the long run than in the short-run. Therefore, the demand for most products will be more elastic in the long run than in the short run.

The price elasticity of demand tends to increase in the long run.

As changing market conditions raise or lower the price of a product, both consumers and producers will respond. However, their response will not be instantaneous, and it is likely to become larger over time. In general, when the price of a product increases, consumers will reduce their consumption by a larger amount in the long run than in the short run. Thus, the demand for most products will be more elastic in the long run than in the short run. This relationship between the elasticity coefficient and the length of the adjustment period is referred to as **the second law of demand**.

Impact on Total Expenditure

Consumers' total expenditure is the same as total revenues from the suppliers' point of view. One of the most important applications of price elasticity is determining how total consumer expenditure on a product changes when the price changes.

Total Revenues = Total Expenditures = Price x Quantity

According to the law of demand, price and quantity move in opposite directions. When the price changes, total revenue also changes. But a rise in price doesn't always increase total revenue. The change in total expenditures depends on whether the effect of the changes in price or the effect of the changes in quantity is greater.

- When demand is inelastic, a change in price will cause total expenditures to change in the same direction.
- When demand is elastic, a change in price will cause total expenditures to move in the opposite direction.
- When demand elasticity is unitary, total expenditures will remain unchanged as price changes.

Because of the relationship between price and quantity sold, a firm's total revenue can rise, fall or stay the same in response to a change in price. The outcome is determined by the price elasticity of demand. This conclusion is similar to that of total expenditures.

Note that firms attempt to maximize profit (total revenue minus total cost), not revenue.

Income Elasticity of Demand: Normal and Inferior Goods

Definition: The percentage change in the quantity of a product demanded divided by the percentage change in consumer income causing the change in quantity demanded.

Since increases in consumer income will increase the demand for most goods, income elasticity measures the responsiveness of a demand for a good to a change in income. Specifically, it tells the analyst the percentage change in the quantity demanded for a good caused by a 1% increase in consumer income.

Calculation:

The type of product is the primary determinant of income elasticity of demand.

- Most products have positive income elasticity; **normal goods** have positive income elasticity.
    - Necessities have low income elasticities (between 0 and 1); when income rises by 1%, the quantity demanded for necessities will increase by less than 1%.
    - Luxuries have high income elasticities (greater than 1).
- A few commodities (**inferior goods**) have negative income elasticity; as income expands, the demand for them will decline. Examples of inferior goods are margarine, junk food, etc.

Cross-Price Elasticity of Demand: Substitutes and Complements

The cross elasticity of demand is a measure of the responsiveness of demand for a good to a change in the price of a substitute or a complement, other factors remaining the same. The formula for calculating the cross elasticity is:

- The cross elasticity of demand for a substitute is positive.
- The cross elasticity of demand for a complement is negative.

The following figure shows the increase in the quantity of pizza demanded when the price of a burger (a substitute for pizza) rises. The figure also shows the decrease in the quantity of pizza demanded when the price of a soft drink (a complement of pizza) rises.

3. Substitution Effect, Income Effect, Normal and Inferior Goods

```
   There are two different phenomena underlying a consumer's response to a price drop:<p>
```

- As the price of a product declines, the lower opportunity cost will induce consumers to buy more of it since it becomes less expensive - even if they have to give up other products. This is called the **substitution effect**.

To isolate this effect diagrammatically, we move the new budget line inwards and parallel until it is tangent to the old indifference curve. The new slope reflects the new relative prices but the utility is the same as it was originally. The substitution effect is Q2Q6. The substitution effect will always lead to more of the relatively cheaper product being demanded.

- With a fixed amount of money income, a reduction in the price of a product will increase a consumer's real income - the amount of goods and services consumers are able to purchase. Typically, consumers will respond by purchasing more of the cheaper products (as well as other products). This is called the **income effect**. The income effect is identified by shifting the budget line back outwards again. In this case, this leads to an increase in the quantity demanded of Q6 Q4.

The substitution and income effects will generally work in the same direction, causing consumers to purchase more as the price falls and less as the price rises. The indifference curve can be used to separate these two effects.

In the case of a **normal good**, higher real income leads to an increase in quantity demanded; this complements the increase due to the substitution effect. This change is shown in the diagram below.

In the case of an **inferior product**, the income effect leads to a fall in the quantity demanded, which will work against the substitution effect. In the following diagram the substitution effect is Q2 Q5; the income effect is Q5 Q4. However, the substitution effect outweighs the income effect and overall the quantity demanded rises. The overall change in quantity demanded results in an increase of Q2 Q4. This means the demand curve is downward-sloping, because a price fall increases the quantity demanded.

When a good is inferior and the income effect outweighs the substitution effect, it is called a **Giffen good**. This is, however, unlikely, because the substitution effect is almost always stronger than the income effect.

Another exception is the case where an increase in price causes an increase in demand. This results in an upward-sloping demand curve, and the good is called a **Veblen good**.

One possible justification for a Veblen good is that people associate higher prices with status, luxury, and quality, so that a higher price might increase the perceived value of a good.

4. Supply Analysis: Cost, Marginal Return, and Productivity

```
A firm is an institution that hires factors of production and organizes them to produc
```

Productivity: The Relationship between Production and Cost

The total cost of production, TC = w x L + r x K, illustrates that the total cost is the cost of all the firm's inputs. The **cost function**, C = f(Q), is a relationship between the cost of production and the flow of output.

Two things determine the cost of production: the price and productivity. Production costs increase as input prices rise and fall as inputs become more productive. **Input productivity** is a measure of the output per unit of input.

Total, Average, and Marginal Product of Labor

Typical productivity measures for a firm are based on the concepts of total product, average product, and marginal product of labor.

- **Total product of labor** (TP_L) is a short-run concept that is the total quantity that is able to be produced for each level of labor input, holding all other inputs constant.

  The total product curve shows how total product changes with the quantity of variable input employed.

  Total product only provides insight into a firm's production volume relative to the industry; it does not show how efficient a firm is in producing its output.

- **Average product of labor** (AP_L) is the total product of labor divided by number of labor hours.

- **Marginal product of labor** (MP_L), also known as marginal return, is the change in total product divided by the change in labor hours.

As more and more units of a variable resource are combined with a fixed amount of other resources, employment of additional units of the variable resource will eventually increase output only at a decreasing rate. Once diminishing returns are reached, it will take successively larger amounts of the variable factor to expand output by one unit.

The **law of diminishing returns** basically explains the old adage: "too many cooks spoil the broth," or too much of a good thing is bad. It dictates that additional output must fall as more and more labor is added to a fixed amount of capital.

5. Economic Profit vs. Accounting Profit

```
    <b>Accounting profit</b> is the profit used by accountants to determine a firm's net i
```

Accounting profit = Total revenue - Total accounting costs

Economic profit equals a firm's total revenue minus its total economic costs.

Economic profit = Total revenue - Total economic costs

Economic cost takes into account the total opportunity cost of all factors of production. **Opportunity cost** is the next best alternative forgone in making a decision. It is the unearned or nominal profit that the resource-owner did not make from investing in the next best alternative. As a result, you can have a significant accounting profit with little to no economic profit.

Example

Suppose a person uses his own resources, land, capital, and time in the production of goods. The opportunity costs of these resources are shown below:

Accounting Profit = $55,000

Entrepreneur's own forgone salary = $40,000

Foregone interest on capital = $1,000

Foregone rent = $2,000

Economic Profit = 55,000 - 40,000 - 1,000 - 2,000 = $12,000

For publicly traded corporations, economic profit is accounting profit - required return on equity capital.

When economic profit is zero, a firm's accounting profit becomes **normal profit**, which is effectively the total implicit opportunity cost.

Accounting profit = Economic profit + Normal profit

When a firm's total revenues are just equal to its total costs, its economic profit is zero, but it still makes accounting profit. Zero economic profit does not mean that the firm is about to go out of business. Instead, it just indicates that the owners are receiving exactly the market (normal) rate of return on their investment.

6. Marginal Revenue, Marginal Cost and Profit Maximization

```
    Revenue is the income generated from the sale of output in product markets.<p> </p><ul
```

- **Total revenue** (**TR**) is the sum of individual units sold multiplied by their respective prices:
- **Average revenue** (**AR**) =
- **Marginal revenue** (**MR**) is the change in revenue from selling one extra unit of output:

In a perfectly competitive market, each firm is a price taker. Since each unit of output sold by a price taker is sold at the market price, the MR for each unit is also equal to the market price, i.e., P = MR.

Under imperfect competition, a firm's marginal revenue is always less than the price of its good. Why? As the firm reduces price in order to expand output and sales, there will be two conflicting influences on total revenue.

- The increase in sales due to the lower price will, by itself, add to the revenue of the monopoly.
- The price reduction, however, also applies to units that would otherwise have been sold at a higher price. This factor itself will cause a reduction in total revenue.

These two conflicting forces will result in marginal revenue - the change in total revenue - that is less than the sales price of the additional units. Thus, the marginal revenue curve of the firm will always lie below the firm's demand curve, which is also the market's demand curve.

TR is maximized when MR = 0.

Total, Average, Marginal, Fixed, and Variable Costs

To produce more output in the short run, the firm must employ more variable inputs, which means that it must increase its costs. In the short run, a firm's total costs (TC) can be broken down into two categories: fixed costs and variable costs (TC = TFC + TVC). Which costs are fixed and which costs are variable depends on the time horizon being dealt with. For a short time horizon, most costs are fixed. For a long time horizon, all costs are variable.

- **Total Fixed Cost.** The sum of the costs that do not vary with output. They will be incurred as long as a firm continues in business and the assets have alternative uses. Examples of fixed costs include rent, property taxes and insurance premiums.

- **Average Fixed Cost.** Total fixed cost divided by the number of units produced. It always declines as output increases.

- **Total Variable Cost.** The sum of those costs that rise as output increases. Total variable costs are zero if output is zero. Examples are wages paid to workers and payments for raw materials.

- **Average Variable Cost.** The total variable cost divided by the number of units produced.

- **Average Total Cost.** Total cost divided by the number of units produced. It is sometimes called per unit cost.

  ATC is high at low levels of output, decreases as output increases (since fixed costs are spread across more units), and then increases as the firm's maximum capacity is approached (since marginal costs increase).

- **Marginal Cost.** The change in total cost required to produce an additional unit of output.

  The law of diminishing returns implies that the marginal costs of producing each additional unit will increase by increasing amounts. Initially, as output expands, the cost of producing each additional unit of output falls, but then begins to rise as the firm approaches its maximum capacity (e.g., too many workers, congested production lines).

Over the output range with increasing marginal returns, marginal cost falls as output increases. Once a firm confronts diminishing returns, larger and larger additions of the variable factor are required to expand output by one unit. This will cause marginal cost (MC) to rise. As MC continues to increase, eventually it will exceed average total cost. Until that point, MC is below ATC (per unit cost), bringing ATC down. When MC is greater than ATC, the additional units cost more than the average, and ATC must increase. A U-shaped short-run average total cost curve results.

The vertical distance between the AVC and ATC curves measures AFC. The vertical distance between the AVC and ATC curves gets smaller as output increases because AFC decrease as output expands.

Profit Maximization

Maximum economic profit requires that (1) marginal revenue (MR) equals marginal cost (MC) and (2) MC not be falling with output.

- For firms under conditions of perfect competition, price is identical to marginal revenue (MR).
- For firms under conditions of imperfect competition, marginal revenue (MR) is less than price.

7. Breakeven Analysis and Shutdown Decision

```
For a price taker (a firm in a perfectly competitive market):<p> </p>
```

Profit = Total Revenue - Total Cost = (Price - Average Total Cost) x Quantity

However, maximum profit is not always a positive economic profit. In the short run, the firm might break even (making a normal profit), make an economic profit, or incur an economic loss.

1. If the price equals the minimum average total cost, the firm breaks even and makes a **normal profit**. A normal profit is a zero economic profit. In this case total revenue = total cost. The minimum ATC point is often referred to as the **breakeven point**.

2. If the price is higher than the minimum average total cost:

c_1BAP indicates the **economic profit** being made by this firm. The firm is making a positive economic profit since the price per unit exceeds the ATC per unit and the total revenue exceeds the total costs.

1. What would happen to profits if the price fell to below the ATC curve?

The firm therefore will produce q_1 units of output, as shown where MC = MR. At q_1, the firm can only charge P per unit, and yet the ATC per unit is higher, at c_2. This means that the firm is making a total economic loss equal to PBAc_2, or the distance of c_2 to P per unit.

If the firm's current sales revenues can cover its variable cost, and the firm anticipates that the lower market price is temporary, it will continue to operate and will face short-run economic losses. It will produce the quantity at which MC = P. This option is better than "shut down" since the firm is able to cover its variable costs and pay some of its fixed costs. If it were to shut down, the firm would lose the entire amount of its fixed costs.

The **shutdown point** is the output and price at which the firm just covers its total variable cost.

- This point is where average variable cost is at its minimum.
- It is also the point at which the marginal cost curve crosses the average variable cost curve.
- At the shutdown point, the firm is indifferent between producing and shutting down temporarily. It incurs a loss equal to total fixed cost from either action.

If the market price is below the firm's average variable cost, a temporary shutdown is preferable to short-run operation. If the firm continues to operate, operating losses merely add to losses resulting from the firm's fixed costs. Shutdown will reduce losses.

Summary

A firm should stay in the market:

- In the short run, if TR >= TVC (shutdown point), or
- In the long run, if TR >= TC (breakeven point).

Would the decision be different if the firm was operating under perfect competition or if it was a monopoly? The answer is no!

8. Understanding Economies and Diseconomies of Scale

```
<b>Short-Run Cost and Long-Run Cost</b><p> </p>
```

The short-run analysis relates costs to output for a specific size of plant. In the long-run, all resources used by the firm are variable.

For each plant size, there is a set of short-run, U-shaped costs curves for MC, AVC, and ATC. This diagram shows the ATC curves of three (of many) possible plant sizes: small, medium, and large.

Using this information, firms can plan, when in their blueprint stages, the optimal plant size they should be relative to the output they want to produce. For example, if a firm wanted to produce more than Q_1 units of output, it would make sense to build a large firm, since costs per unit would be less than they would be with a small or medium firm.

Long-Run Average Cost Curve

To explain this process, imagine the output level Q_2. Looking at the relevant costs on the vertical axis, the large firm is far cheaper per unit than both the small and medium-sized firms.

Thus, should a firm be planning for output in excess of Q_1, a large firm should be built. For levels of output between Q_0 and Q_1, it would be cheaper per unit if the firm was of a medium size.

- If a firm is planning to produce less than Q_0 units, a small firm would be best.
- For output between Q_0 - Q_1 units, a medium firm is preferable.
- For output in excess of Q_1 units, a large firm is preferable

The long-run average total cost curve is indicated in black.

It shows the minimum average cost of producing each output level when the firm is free to choose among all possible plant sizes. It can best be thought of as a planning curve, because it reflects the expected per-unit cost of producing alternative rates of output while plants are still in the blueprint stage. No single plant size could produce the alternative output rates at the costs indicated by the planning curve.

In reality, there are an infinite number of firm sizes:

The minimum point on the long-run average total cost curve defines the minimum efficient scale for a firm.

Economies and Diseconomies of Scale

Economies of scale are reductions in the firm's per-unit costs that are associated with the use of large plants to produce a large volume of output. They are present over the initial range of outputs when the long-run ATC curve is falling. There are three reasons why economies of scale exist:

- Mass production is more economical.
- Specialization of labor and equipment improves productivity.
- Workers at a larger firm tend to learn more from their experience.
- Bargaining power in input price.

Diseconomies of scale are situations in which the long-run average total costs are greater in larger firms than they are for smaller firms. They are possible: as a firm gets bigger and bigger, bureaucratic inefficiencies may result; principal-agent problems grow; communication breakdowns and bottlenecks can raise input prices. They are present when the long-run ATC curve is rising.

It is also possible to have constant unit costs as the plant size changes. This is known as **constant returns to scale**.

Economies and diseconomies of scale are long-run concepts. They relate to conditions of production when all factors are variable. In contrast, increased and diminishing returns are short-run concepts, applicable only when the firm has a fixed factor of production.

The downward-sloping portion shows economies of scale. The horizontal portion shows constant returns to scale. The upward-sloping portion shows diseconomies of scale.

Minimum efficient scale is the smallest quantity of output at which the long-run average cost reaches its lowest level. If the long-run average cost curve is U-shaped, the minimum point identifies the minimum efficient scale output level.

The Firm and Market Structures

1. Characteristics of Different Market Structures

```
   A financial analyst must understand the characteristics of market structures to better
```

We focus on those characteristics that affect the nature of competition and pricing. They are:

- The number of firms (including the scale and extent of foreign competition).
- The extent of product differentiation (which affects cross-price elasticity of demand).
- The pricing power of seller(s). Can a firm influence the market price?
- Barriers to entry. Exit costs should also be considered.
- Non-price competition such as product differentiation.

The characteristics of each market structure will be discussed in subsequent subjects of this reading.

2. Perfect Competition

```
   An industry with perfect competition displays the following characteristics:<p> </p><u
```

- All the firms in the market are producing an identical product (e.g., wheat of the same grade).
- No barriers limit the entry or exit of firms in the market.
- A large number of firms exist in the market. Established firms have no advantages over new ones.
- Sellers don't have market-pricing power.
- There is no non-price competition.

Perfect competition arises:

- When a firm's minimum efficient scale is small relative to market demand so there is room for many firms in the industry, and
- When each firm is perceived to produce a good or service that has no unique characteristics, so consumers don't care which firm they buy from.

In perfect competition, each firm is a price taker. Price takers are sellers who must take the market price in order to sell their products.

- There is no price decision to make: they will merely attempt to choose the output level that will maximize profit.
- Each price taker's output is small relative to the total market: the output of a firm exerts little or no effect on the market price.

This diagram represents the market demand and supply curve for a certain product - for example, eggs.

As usual, the intersection of the demand and supply curve creates the market price (P) per egg. Now remember that a firm that is a price taker can sell all it wants to at that price, but can sell nothing at a higher price.

Price takers can sell *all* their output at the market price, but they are unable to sell *any* of their output at a price higher than the market price. That is, a price taker faces a *horizontal* demand curve. Each firm's output is a perfect substitute for the output of the other firms, so the demand for each firm's output is *perfectly elastic*.

- They can sell as much as they would like at the going market price.
- There is no need for them to reduce their price in order to sell more.
- Moreover, at any price above the market price there is no demand; their sales would be zero (nobody would buy from that firm because there are so many other firms from which to obtain the product at the market price).
- This reflects the fact that perfectly competitive firms have no control over their price.

When a perfectly competitive market is in long-run equilibrium:

- Quantity supplied and quantity demanded must be equal in the market.
- Firms in the market must earn zero economic profit at the prevailing market price (that is, firms are earning the "normal rate of return"). This occurs when market price = marginal revenue = marginal cost = minimum ATC. Note that accounting profits may still be positive.

Why do firms earn zero economic profit in the long-run equilibrium?

- If firms earn positive economic profit in the long-run equilibrium, these firms will have an incentive to expand their capacity, and new firms will enter the market. This will lead to an increase in supply, forcing the market price down until economic profit is eliminated.
- Conversely, if firms in the market incur economic losses, some firms will leave the market. Accordingly, supply will decline, causing the market price to rise until existing firms can earn the normal rate of return (that is, economic profit is zero).

3. Monopolistic Competition

```
    A <b>monopolistic market</b> is also called a <b>competitive price searcher market</b>
```

Characteristics are:

- **A large number of firms.** This is due to low entry barriers and causes intense competition in these markets. Firms face competition from existing firms and potential entrants to the market.

- **Firms produce differentiated products.** This means that each firm makes a product that is slightly different from the products of competing firms. *This is the most distinctive characteristic of such a market.*

- **Low entry barriers.** Entry into and exit from the market are relatively easy. Sellers in competitive price searcher markets face competition both from firms already producing in the market and from potential new entrants into the market. If profits are present, firms can expect that new rivals will be attracted. Because of the low entry barriers, competitive forces will be strong in monopolistic markets, and firms cannot earn an economic profit in the long run.

- **Competition on quality, price, and marketing.** Demand is not simply given for a monopolistic competitor. The firm has some pricing power and can alter the demand for its products by changing product quality (design, reliability and service), location and by advertising. The firm faces a downward-sloping demand curve. This demand curve is highly elastic because good substitutes for a firm's output are readily available from other suppliers.

Consider two hamburger companies: McDonald's and Burger King.

- Both firms are producing burgers but customers view them as differentiated.
- If McDonald's increases the price of its burger, it will not lose all its customers, as some will continue to pay the higher price, preferring McDonald's.
- Thus, differentiation explains the downward-sloping demand curve. The more firms producing burgers (substitutes), the more elastic McDonald's demand curve will be, since the greater the decrease in quantity demanded as price increases.

The Firm's Short-Run Output and Price Decision

As with price takers, monopolistic competitors maximize profits by expanding output to where MR = MC.

A firm in monopolistic competition operates much like a single-price monopolist.

According to the demand curve, the firm can charge P1 per unit.

- The total revenue earned is the shaded area $0P_1AQ_1$.
- The total cost is the shaded area $0CBQ_1$.
- It earns an economic profit (as in this example) when P > ATC. The total profit is thus the difference between total revenue and total costs, and is given by the shaded area CP_1AB.

A firm might incur an economic loss in the short run when P < ATC.

Long Run: Zero Economic Profit

Whenever firms can freely enter and exit a market, profits and losses play an important role in determining the size of the industry. Economic profits will attract new competitors to the market and economic losses will cause competitors to exit from the market. In the short run, a price searcher may make either economic profits or losses, depending on market conditions. As firms enter the industry, each existing firm loses some of its market share. The demand for its product decreases and the demand curve for its product shifts leftward.

The decrease in demand decreases the quantity at which MR = MC and lowers the maximum price that the firm can charge to sell this quantity. After long-run adjustments have been made, price and quantity fall with firm entry until P = ATC and firms earn zero economic profit.

If firms incur an economic loss, firms exit, to achieve long-run equilibrium.

4. Oligopoly

Literally, <b>oligopoly</b> means "few sellers." This market structure is characterize

- **A small number of rival firms.** The firms are interdependent because each is large relative to the size of the market. The decisions of a firm often influence the demand, price, and profit of rivals, and an oligopolist must consider the potential reaction of rivals.

- **High entry barriers into the market.** Either natural or legal barriers to entry can create oligopoly.

    - Economies of scale are probably the most significant entry barrier here. Achieving minimum per-unit cost is required, and thus a small number of large-scale firms will be able to produce the entire market demand for the product. This is what distinguishes an oligopoly from a monopolistic competitive market.
    - A legal oligopoly might arise even where demand and costs leave room for a larger number of firms.

In short, an oligopoly is competition among the few.

Pricing Strategies

- Like a monopolist, an oligopolist faces a downward-sloping demand curve and seeks to maximize profit, not price.
- Unlike a monopolist, an oligopolist cannot determine the product price that will deliver maximum profit simply by estimating market demand and cost conditions.

A key factor here is the pricing behavior of close rivals, or interdependence between firms. This means that each firm must take into account the likely reactions of other firms in the market when making pricing decisions. Because the reactions of those rivals cannot be determined, the precise price and output that will emerge under an oligopoly cannot be determined. Only a potential range of prices can be indicated.

There are three basic pricing strategies.

1. The assumption of **pricing interdependence** is that firms will match a price reduction and ignore a price increase. The idea is that if a firm raises prices, other firms won't follow, because they won't worry about losing market share to a firm that is raising its prices. However, if the firm lowers its prices, other firms will respond by lowering their prices also, since they don't want to lose market share.

The demand curve that a firm believes it faces has a kink at the current price P and quantity Q.

The kinked demand curve can be thought of as two demand curves.

- Above the price P, an individual firm is afraid of putting up prices. A price increase would, it assumes, not be matched by competitors, hence the demand curve above P is elastic. It will be remembered that if demand is elastic and price rises, revenue falls.
- Similarly, a price fall has the same effect on revenue. This time the firm imagines that dropping its own price leads to others dropping theirs. Overall, quantity demand increases as the demand curve slopes down, but the increase is less than proportionate. That is the demand curve below price P is inelastic.

The kink in the demand curve means that the MR curve is discontinuous at the current quantity - shown by the gap AB in the figure.

Fluctuations in MC that remain within the discontinuous portion of the MR curve leave the profit-maximizing quantity and price unchanged.

For example, if costs increased so that the MC curve shifted upward from MC_0 to MC_1, the profit-maximizing price and quantity would not change.

The beliefs that generate the kinked demand curve are not always correct and firms can figure out this fact. If MC increases enough, all firms raise their prices and the kink vanishes.

1. The assumption of the **Cournot model** is that a firm will embrace another's output decisions in selecting

its profit-maximizing output but that decision is fixed. This means that each firm is naively conjecturing that should either one of them alter their output decisions, the other will not react.

Example

Assume there are 2 firms. The market demand takes the following form: P = 30 - Q, where Q = Q_1 + Q_2 and Q_1 = Q_2 (i.e., industry output constitutes firm 1 and 2's output respectively and both firms share the market).

Also assume that average cost (AC) and marginal cost (MC): AC = MC = 12.

Firm 1's total revenue (TR) is P x Q_1 = (30 - Q) x Q_1 = [30 - (Q_1 + Q_2)] x Q_1 = 30 x Q_1 - $Q1^2$ - Q_1 x Q_2.

Firm 1's MR is thus MR_1 = 30 - 2Q_1 - Q_2

If MC = 12, then Q_1 = 9 - 1/2 x Q_2. This is Firm 1's reaction curve.

For Firm 2, Q_2 = 9 - 1/2 x Q_1.

Solving for Q_1, we find Q_1 = 6, Q_2 = 6 and P = 18.

This equilibrium can be compared with that of perfect competition and monopoly.

- Perfect Competition: Firms set prices equal to MC. So P = 12 and Q = 18.
- Monopoly: TR = PQ = (30 - Q)Q = 30Q - Q^2, MR = 30 - 2Q. As MC = 12 = MR, Q = 9 and P = 21.

We can see that 2 firms operating under Cournot assumptions offer a better welfare outcome than under monopoly.

If the number of firms increases, then the Cournot equilibrium approaches the competitive equilibrium.

1. **Nash equilibrium** is a concept of game theory where a firm does what is best for itself after it takes into account other firms' actions. For example, McDonald's charges $2.99 for a Value Meal based on what Burger King and Wendy's are charging for a similar menu item. McDonald's would reconsider its pricing if its rivals were to change their prices.

An Oligopoly Price-Fixing Game

Collusion is the opposite of competition. It involves cooperative actions by sellers to turn the terms of trade in favor of the group and against buyers.

If there is no collusion and each oligopolist act independently, seeking to maximize profits by offering consumers a better deal than its rivals, the market price would be driven down to its lowest level and firms would be just able to cover their per-unit cost. This is like a pure competitive market.

However, there is a strong incentive for oligopolists to collude, agreeing to raise price and to restrict output. They can form a cartel (such as OPEC) or they can collude without such a formal organization. In this case, the highest price occurs.

Oligopolists have a strong incentive to collude since they can profit by restricting output and raising price. There are six major factors that affect the chances of successful collusion:

- The number and size distribution of sellers
- The similarity of the products
- Cost structure
- Order size and frequency
- The strength and severity of retaliation

- The degree of external competition

Optimal Price and Output

There is no single optimum price and output analysis that fits all oligopoly market situations. Consider a **dominant firm model** where the market consists of a dominant firm and some fringe firms. The dominant firm becomes the price maker. It operates as a monopoly, faces a residual demand curve, and chooses price and output to maximize its profit (MR = MC). Other firms are price takers or followers.

Example

The following figure shows a dominant firm industry. On the left are 10 small firms and on the right is one large firm.

- The demand curve, D, is the market demand curve and the supply curve, S10, is the supply curve of the 10 small firms.
- At a price of $1.50, the 10 small firms produce the quantity demanded. At this price, the large firm would sell nothing.
- But if the price was $1.00, the 10 small firms would supply only half the market, leaving the rest to the large firm.
- The demand curve for the large firm's output is the curve XD on the right. The large firm can set the price and receives marginal revenue that is less than that price along the curve MR.
- The large firm maximizes profit by setting MR = MC. Let's suppose that the marginal cost curve is MC in the figure. The profit-maximizing quantity for the large firm is 10 units. The price charged is $1.00.
- The small firms take this price and supply the rest of the quantity demanded.

5. Monopoly

```
Literally, <b>monopoly</b> means "single seller." It is a market structure characteriz
```

- High entry barriers.
- A single seller of a well-defined product for which there are no good substitutes.

Barriers to entry include legal or natural constraints that protect a firm from potential competitors.

- **Legal barriers** to entry create a legal monopoly, a market in which competition and entry are restricted by the granting of a:
    - Public franchise. The U.S. Postal Service franchises to deliver first-class mail.
    - Government license. Licensing is a requirement that one obtain permission from the government in order to perform certain business activities or work in various occupations. It limits entry and restricts the right to buy and sell goods. Sometimes these licenses cost little and are designed to ensure certain minimum standards. In other cases, they are expensive and designed primarily to limit competition. For example, in many U.S. states a license is required for operating a taxi.
    - Patent and copyright. The entry barrier created by the grant of a patent generally leads to higher consumer prices for products that have already been developed. On the other hand, the absence of patent protection might well lead to a slowdown in the pace of technological innovation.

- **Natural barriers** to entry (for example, economies of scale) create a natural monopoly, which is an industry in which one firm can supply the entire market at a lower price than two or more firms can. In some industries, larger firms will always have lower unit costs. It will be difficult for small firms to enter the market, build a reputation, and compete effectively. Economies of scale tend to eventually result in the market being dominated by one large firm.

- **Other barriers**, such as strong brand loyalty or the increasing returns associated with network effects.

Demand and Supply Analysis

A monopoly faces no competition, and as a result there is no product differentiation. It is a price setter, not a price taker like a firm in perfect competition. Because the monopoly is the only seller in the market, the demand for its product is the market demand curve. It is downward-sloping because demand will decline as price increases.

Marginal Revenue and Price

A monopoly must choose between lower prices with larger quantities sold and higher prices with smaller sales. Although a monopoly can set the price for its products, market forces will determine the quantity sold at alternative prices. To maximize profit, a monopoly must estimate the relationship between price and the quantity of its products demanded.

As the monopoly reduces price in order to expand output and sales, there will be two conflicting influences on total revenue.

- The increase in sales due to the lower price will, by itself, add to the revenue of the monopoly.
- The price reduction, however, also applies to units that would otherwise have been sold at a higher price. This factor will cause a reduction in total revenue.

These two conflicting forces will result in marginal revenue - the change in total revenue - that is less than the sales price of the additional units. *Thus, the marginal revenue curve of the monopoly will always lie below the firm's demand curve, which is also the market's demand curve:*

MR < P

The following example illustrates this concept.

Portico produces beauty soaps.

- If Portico charges $10 for each bar of soap, the demand will be only 1 bar of soap. Both the total revenue and the marginal revenue from the first bar of soap will be $10.
- If the price is $8, the demand will be 2 bars of soap. Now the total revenue will be $16 ($8 x 2). The marginal revenue from the second bar of soap will be $6 ($16 - $10).
- Therefore, Portico's demand curve is line AB and its marginal revenue curve is line AC. Note that line AC lies below line AB.

Marginal Revenue and Elasticity

A single-price monopoly's marginal revenue is related to the elasticity of demand for its good:

- If demand is elastic, a fall in price brings an increase in total revenue. The rise in revenue from the increase in quantity sold outweighs the fall in revenue from the lower price per unit and MR is positive. Total revenue increases.
- If demand is inelastic, a fall in price brings a decrease in total revenue. The rise in revenue from the increase in quantity sold is outweighed by the fall in revenue from the lower price per unit and MR is negative. Total revenue decreases.
- If demand is unit elastic, a fall in price does not change total revenue. The rise in revenue from the increase in quantity sold equals the fall in revenue from the lower price per unit and MR = 0. Total revenue is maximized when MR = 0.

A single-price monopoly never produces an output at which demand is inelastic.

If it did produce such an output, the firm could increase total revenue, decrease total cost, and increase economic profit by decreasing output.

Price and Output Decision

A monopoly faces the same types of technology constraints as a competitive firm but the monopoly faces a different market constraint. The monopoly selects the profit-maximizing level of output in the same manner as a competitive firm, where MR = MC. Therefore, the monopoly will lower price and expand output until marginal revenue is equal to marginal cost.

The ATC curve tells the analyst the average cost. Economic profit is the profit per unit multiplied by the quantity produced. In the above figure, total revenues P0AQ0O exceed the firm's total costs of DBQ0O at the profit-maximizing output level of Q0. Accordingly, the monopoly is making an economic profit of P0ABD.

Unlike a price taker, the monopoly may earn an economic profit, even in the long run, because the barriers to entry protect the firm from market entry by competitor firms.

Monopoly and Competition Compared

Comparing Output and Price

- The market demand curve, D, in perfect competition is the demand curve that the firm faces in a monopoly.
- The market supply curve in perfect competition is the horizontal sum of the individual firm's marginal cost curves, S = MC. This curve is the monopoly's marginal cost curve.
- Equilibrium in perfect competition occurs where the quantity demanded equals the quantity supplied at quantity Q_C and price P_C.
- Equilibrium output for a monopoly, Q_M, occurs where marginal revenue equals marginal cost, MR = MC.
- Equilibrium price for a monopoly, P_M, occurs on the demand curve at the profit-maximizing quantity.
- Because marginal revenue is less than price at each output level, $Q_M < Q_C$ and $P_M > P_C$.

Compared to perfect competition, monopoly restricts output and charges a higher price.

Regulating Natural Monopoly

When demand and cost conditions create a natural monopoly, government agencies regulate the monopoly.

Without regulation, a monopolist would produce Q_0, charge price P_0 and maximize profit. It produces the quantity at which marginal revenue equals marginal cost.

Average cost pricing is when the government instructs the monopolist to produce that output where the demand curve intersects the ATC curve. Here, price decreases to P_1 and output increases to Q_1. Thus, public welfare improves.

This is because the value of the additional production to society (Q_0ABQ_1) exceeds the cost of the additional production Q_0CDQ_1 by the area CABD. The firm is making normal profits (zero economic profits) since the price being charged, P_1, is just sufficient to cover the average cost per unit.

6. Price Discrimination

```
    <b>Price discrimination</b> is a practice whereby a seller charges different consumers
```

- In *first-degree price discrimination* each consumer is charged the maximum he is willing to pay. Consumer surplus is nil, while producer surplus is maximized. The output is the same as in a competitive market.
- In *second-degree price discrimination*, prices vary across units but not people. Consumers self-select into consumption groups and seek the largest surplus.

- In *third-degree price discrimination*, consumers are segregated by demographic or other traits. Prices are determined by the demands of each group.

When sellers can segment their market (at a low cost) into groups with differing price elasticities of demand, price discrimination can increase profits. For each group, the seller will maximize profit by equating marginal cost and marginal revenue. The number of units sold also increases because the discounts provided to price-sensitive groups increase the quantity sold more than the higher prices charged the less price-sensitive groups reduce sales.

Imagine that the MC per unit for a monopoly is constant at $60, producing a horizontal MC curve, as shown below.

The firm produces where MC = MR. It thus produces 100 units and charges $200 per unit. Total revenue (price x quantity) for the firm is thus: $200 x 100 = $20,000. Total costs (cost per unit x quantity) are: $60 x 100 = $6,000. Total profit is thus: $20,000 - $6,000 = $14,000.

Imagine that this firm is an airline and that it now decides to increase its profits using price discrimination. It identifies two groups of people: businessmen, who are fairly price-inelastic, and students, who are fairly price-elastic (responsive to changes in price). By increasing the price of businessmen's tickets and decreasing the price of student's tickets, it can increase its total revenue and thus increase its profit.

The airline starts by doubling the price of businessmen's tickets to $400. By equating the businessmen's MR curve to the MC curve (for simplicity, the MR curve is not shown), the airline finds that the quantity demanded decreases, but by relatively little, given the large increase in price, to 60 tickets.

Next, it equates MC to the students' MR curve and finds that it can decrease the price of students' tickets from $200 to $150, whilst the quantity demanded increases to 150 tickets. (Note: for simplicity, the students' MR curve is not shown).

Therefore, the total revenue is as follows: $400 x 60 + $150 x 150 = $46,500.

Total costs are: $60 x 60 + $60 x 150 = $12,600.

Total profits are thus: $46,500 - $12,600 = $33,900.

This is more than the $14,000 profit the firm made in the absence of price discrimination.

7. Identification of Market Structure

```
Measuring market power is complicated. Ideally, econometric estimates of the elasticit
```

The N-firm concentration ratio is the percentage of market output generated by the N largest firms in the industry. The ratio is used as an indicator of the relative size of firms in relation to the industry as a whole. It may also assist in determining the market form of the industry. The larger the measure of market concentration, the less competition exists in the industry.

The concentration ratio is simple to compute. However, it does not directly quantify market power, meaning it does not take the possibility of entry into account. Another disadvantage is that it ignores mergers among the top market players.

The Herfindahl-Hirschman Index (HHI)

The **Herfindahl-Hirschman index** is the sum of the squared market shares of the top N largest firms in the industry.

$$H = M_1^2 + M_2^2 + M_3^2 + ... + M_N^2$$

where M_i is the market share of an individual firm.

Suppose there are a total of four firms in a specific industry. Three firms have a 20% share each and one has a 40% market share, $H = 0.20^2 + 0.20^2 + 0.20^2 + 0.40^2 = 0.28$.

The advantages of the Herfindahl index are that it reflects more firms in the industry and it gives greater weight to the companies with larger market shares.

Properties of the Herfindahl index:

- It is always smaller than or equal to 1. In a monopoly, the HHI is 1.
- One can classify the competition structure of a market based on this ratio. For example,
    - An H below 0.1 indicates a competitive market.
    - An H of 0.1 to 0.18 indicates moderate competitive.
    - An H above 0.18 indicates uncompetitive.
- If all firms have an equal share, $H = N \times (1/N)^2 = 1/N$. Note that the reciprocal of the index shows the number of firms in the industry.
- When the firms have unequal shares in the industry, the reciprocal of the index indicates the "equivalent" number of firms in the industry. Using the above example, the market structure is equivalent to having 1/0.28 = 3.57 firms of the same size.

Limitations: HHI fails to consider barriers to entry and firm turnover. For example, for some industries, few firms may be currently operating in the market but competition might be fierce, with firms regularly entering and exiting the industry. Even potential entry might be enough to maintain competition.

Aggregate Output, Prices, and Economic Growth

1. Gross Domestic Product

```
<b>Gross Domestic Product (GDP)</b> is the total market value of all domestically proc
```

- Only **final goods and services** count; GDP includes goods and services purchased by final users. **Intermediate goods** purchased for resale or for the production of another good or service are excluded, to avoid double-counting. Their value is embodied in the value of the goods purchased by the end user.

- GDP is a *flow* variable; it measures the market value of production that flows through the economy.

- Financial transactions and income transfers (e.g., social security and welfare payments) are excluded because they represent *exchanges*, not productions, of goods and services. GDP counts transactions that add to current production.

- GDP counts only goods and services produced domestically, whether by citizens or foreigners.

- It includes only goods produced during the current period. Thus, sales of used goods are not counted in GDP. However, sales commissions count toward GDP because they involve services provided during the period.

Government services and household production are estimated and included in the GDP. Activities occurring in the underground economy, although sometimes productive, are not included in GDP.

Nominal and Real GDP

When comparing GDP across time periods, we confront a problem: the nominal value of GDP may increase as the result of either expansion in the quantities of goods produced or higher prices. Since the former will improve our living standards, we have to adjust the nominal values (**nominal GDP**, or **money values**) for the effects of inflation to get real values (**real GDP**).

A **price index** is used for the adjustment. It measures the cost of purchasing a market basket or bundle of goods at a point in time relative to the cost of purchasing the identical market basket during an earlier reference period (e.g., a base year).

Consumer price index (CPI) (not included in the required reading) is an indicator of the general level of prices. It attempts to compare the cost of purchasing the market basket bought by a typical consumer during a specific period with the cost of purchasing the same market basket during an earlier period. The CPI is better at determining *how rising prices affect the money income of* ***consumers***. The CPI is more widely used for price changes over time.

The **GDP deflator** is a price index that reveals the cost during the current period of purchasing the items included in GDP relative to the cost during a base year. Because the base year is assigned a value of 100, as the GDP deflator takes on values greater than 100, it indicates that prices have risen. **It is a broader price index than the CPI** since it is better at *giving an economy-wide measure of inflation.* It is designed to measure the change in the average price of the market basket of goods included in GDP. In addition to consumer goods, *the GDP deflator includes prices for capital goods and other goods and services purchased by businesses and governments.* The GDP deflator also allows the basket of goods to change as the composition of GDP changes, while the CPI is computed using a fixed basket of goods.

We can use the GDP deflator together with nominal GDP to measure the **real GDP** (GDP in dollars of constant purchasing power).

Real GDPi = Nominal GDPi x (GDP Deflator for base year/ GDP Deflator for year i)

Suppose the nominal GDPs in 1992 and 2010 were $6244 and $8509 billion dollars, respectively. This amount has increased by 36.3%. The GDP deflator for 1992 and 2010 was 100 and 112.7, respectively. The real GDP in 2010, therefore, should be:

Nominal GDP in 2010 x GDP deflator in 1992 / GDP deflator in 2010

= 8509 x 100 / 112.7

= $7550 billion dollars

Measured in terms of 1992 dollars, the real GDP in 2010 was only 20.9% higher than that in 1992.

2. The Components of GDP and Related Measures

```
  GDP is a measure of both output and income.  The revenues that firms derive from the s
```

There are two ways of measuring GDP. *GDP derived by these two approaches will be equal.*

The **expenditure approach** totals the expenditures spent on all final goods and services produced during the year. Under this approach, GDP is a measure of *aggregate output*. There are four components of GDP under this approach:

- GDP = C + I + G + (X - M)
- *C (personal consumption expenditures)*: This is the largest component with this approach: durable goods, non-durable goods, and services.
- *I (gross private domestic investment)*: The flow of private sector expenditures on durable assets plus the addition to inventories during a period. It is the production or construction of capital goods that provide a flow of future service. It indicates the economy's future productive capacity.
- *G (government consumption and gross investment)*: Government purchases, not including transfer payments. It includes both (1) expenditures on such items as office supplies, law enforcement, and the

operation of veteran hospitals and (2) the capital purchase of long-lasting capital goods such as missiles, highways, and dams for flood control. Government expenditures, which include transfer payments like social security, are not equal to government consumption.
- *E - M (net exports to foreigners)*: This is exports minus imports. Exports are domestically produced goods and services sold to foreigners. Imports are foreign-made goods and services purchased by domestic consumers, investors and governments. When measuring GDP using the expenditure approach, we must add exports and subtract imports. Net exports may be either positive or negative.

GDP can be measured either from the value of the final output or by summing the value added at each stage of the production and distribution process. The sum of the value added by each stage is equal to the final selling price of the good.

Under the **income approach**, GDP is a measure of *aggregate income*. It is calculated by summing the income payments to resource suppliers and the other costs of producing those goods and services. It includes employee compensation (wages and salaries), self-employment income, rents, profits and interest, etc. Employee compensation is the largest source of income generated by the production of goods and services.

Personal income is the total income received by domestic households and non-corporate businesses. It is available for consumption, saving, and payment of personal taxes. **Personal disposable income** is an individual's available income, after personal taxes are paid, that can be either consumed or saved.

3. Aggregate Demand

```
<b>Aggregate demand</b> (<b>AD</b>) is the quantity of goods and services that househc
```

The IS Curve

GDP = C + I + G + (X - M)

- Consumption (C) is a function of disposable income. Consumption increases when income(Y) increases and/or taxes decrease.

  The **marginal propensity to consume** (**MPC**) is defined as additional current consumption divided by additional current disposable income. The marginal propensity to save (MPS) = 1 - MPC.

- Investment spending depends on the interest rate (i) and output/income level.

- Government purchases are assumed to be unrelated to both interest rates (i) and income (Y). Although tax policy is also considered an exogeneous variable, the actual taxes collected depend on income (Y) and are therefore endogenous. The government's deficit (G - T) increases as income decreases and vice versa.

- Net exports (X - M) depends on relative income and prices between the domestic country and the rest of the world.

We can also derive the following equation, which shows that domestic saving has three uses: investment, government deficits, and trade surplus:

S = I + (G - T) + (X - M), where S is domestic saving.

If we combine these relationships together we can derive the IS curve: the combination of GDP (Y) and the real interest rate (i) such that aggregate income/output equals planned expenditures.

Note that there is an inverse relationship between income and the real interest rate. For example, when interest rates are high, investment falls and therefore Y must fall as well.

Note that changes in Y caused by changes in i are reflected as movements along the IS curve. On the other hand, changes in Y that are brought about by factors other than interest rates will cause Y to change, regardless of the level of interest rates in the economy. For example, changes in government purchases will not change the slope but will change the intercepts; in other words, they will cause the IS curve to shift.

The LM Curve

The IS curve depicts combinations of interest rates and output that clears markets for goods and services. The IS curve by itself does not pin down the interest rate that prevails in the economy. In order to do so, we look at the money market. The LM curve summarizes all the combinations of income and interest rates that equate money demand and money supply.

The quantity theory of money: MV = PY, where V is the velocity of money.

- When interest rates (i) are high, the demand for money is low because money pays no interest; the opportunity cost of holding money rises.
- When Y is high, the demand for money is high; richer people buy more goods and are likely to hold more money.
- When P is hig,h the demand for money is high because we need more money to buy goods.

When the money market is in equilibrium, money supply = money demand. The LM curve summarizes all the combinations of income and interest rates that equate money demand and money supply. It is an upward-sloping relationship between i and Y.

Intuitively, we can explain the upward-sloping LM curve as follows: Let's consider some combination of income and interest rates that equates money demand with the money supply set by the Fed. Now suppose there is an increase in income. The increase in income causes the demand for money to increase. However, money supply is unaffected by the increase in income. The only way that money demand and money supply can be equal again is if interest rates also increase to reduce money demand.

The Aggregate Demand Curve

Combining the IS and LM relationships yields the aggregate demand curve, which depicts the inverse relationship between the price level and real income/output, assuming a constant money supply.

For example, a reduction in the price level will:

- Increase the wealth of people holding the fixed quantity of money.
- Reduce the real rate of interest.
- Make domestically produced goods cheaper than those produced abroad.

All these factors will lead to an increase in the quantity of goods and services demanded at the lower price level (movement along the curve).

4. Aggregate Supply

```
   The aggregate quantity of goods and services supplied depends on three factors: labor
```

$$Y = F(L, K, T)$$

The aggregate supply curve (AS) represents the relationship between the quantity of goods and services supplied and the price level. It is important to distinguish between long-run aggregate supply and short-run aggregate supply.

The *short-run aggregate supply curve* typically slopes upward to the right. In the short run, some prices (e.g., rents, wages) are temporarily fixed as the result of prior commitments. Therefore, firms will expand outputs as the price level increases because higher prices will improve profit margins. Short-run equilibrium occurs when the aggregate quantity of goods and services demanded is equal to the aggregate quantity supplied.

The *long-run aggregate supply curve* is vertical. In the long run, people have sufficient time to alter their behavior to adjust fully to price changes. The sustainable level of output is determined by a nation's resource base, technology, and the efficiency of its institutional factors. The price level has no effect on a nation's long-run aggregate supply. In long-run equilibrium, current output (Y_{full}) will equal the economy's potential GDP, the economy is operating at full employment, and the actual rate of unemployment will equal the natural rate of unemployment.

Aggregated demand and supply determine the level of real GDP and the price level of a nation.

5. Shifts in Aggregate Demand and Supply

```
    <b>Factors that Shift Aggregate Demand</b><p> </p>
```

At each price level, the AD curve shifts to the right due to changes in C, I, G, and X.

- An increase in real wealth: greater wealth increases the demand for all goods.
- Increased optimism about the future: both current consumption and investment increase.
- High capacity utilization: companies have to increase investment spending to expand production.
- Expanding fiscal policy: higher government spending and lower taxes will increase G and C.
- An increase in money supply: higher income and expenditure.
- A lower interest rate - when borrowing is cheaper, investment increases; consumption is cheaper with a lower interest rate.
- A decrease in exchange rate: increases export demand.
- Growth in the global economy: increases export demand.

And vice versa.

Factors that Shift Aggregate Supply

We need to differentiate between the long-run and short-run effects.

Increases in short-run aggregate supply (SRAS) that don't affect long-run aggregate supply are caused by:

- A decrease in resource prices/production costs (e.g., nominal wages, input prices). Unless the lower prices of resources reflect a long-term increase in the supply of resources, they will not alter LRAS.
- A reduction in the expected rate of inflation. If high inflation is expected, suppliers would like to reduce supplies now to sell them at higher prices later but consumers would like to spend more money now.
- Lower business taxes and higher government subsidies.
- Favorable exchange rates for importers of raw materials.

And vice versa.

Long-run supply refers to the economy's long-run production possibilities (maximum rate of sustainable output). Increase in long-run aggregate supply (LRAS) is caused by:

- An increase in the supply of resources. This will expand the economy's sustainable rate of output. Note that an economy's resource base includes physical capital, natural resources and human capital.
- An improvement in technology and productivity. This will increase the average output per unit of resources.

And vice versa.

6. Equilibrium GDP and Prices

```
    <b>Short-run macroeconomic equilibrium</b> occurs when the quantity of real GDP demanc
```

In short-run equilibrium, real GDP can be greater than or less than potential GDP.

Long-run macroeconomic equilibrium occurs when real GDP equals potential GDP - when the economy is on its LAS curve.

Note two things:

- At the price chosen by suppliers (P), the aggregate demand (AD) is exactly equal to the amount suppliers are willing to supply (SAS).
- This equilibrium between SAS and AD coincides with the maximum capacity of the economy, as indicated by the LAS curve. The level of output produced is labeled as Y_f, indicating full employment of resources.

Long-run equilibrium thus occurs where LAS, AD, and SAS coincide.

Economic Growth and Inflation

Economic growth occurs because the quantity of labor grows, capital is accumulated, and technology advances, all of which increase potential GDP and bring a rightward shift of the LAS curve. The following figure illustrates economic growth and inflation.

Inflation occurs because the quantity of money grows faster than potential GDP, which increases aggregate demand by more than long-run aggregate supply.

The AD curve shifts rightward faster than the rightward shift of the LAS curve.

The Business Cycle

The business cycle occurs because aggregate demand and short-run aggregate supply fluctuate.

- A **below full-employment equilibrium** is an equilibrium in which potential GDP exceeds real GDP. The amount by which potential GDP exceeds real GDP is called a **recessionary gap**.
- **Long-run equilibrium** is an equilibrium in which potential GDP equals real GDP.
- An **above full-employment equilibrium** is an equilibrium in which real GDP exceeds potential GDP. The amount by which real GDP exceeds potential GDP is called an **inflationary gap**.

Let's look at the inflation gap.

An economic boom may be the result of an increase in AD. Starting at long-run equilibrium, an increase in aggregate demand shifts the AD curve rightward.

The prices of goods and services increase, which in turn induces suppliers to expand output to a level that is unsustainable in the long run (which is why a boom is followed by an economic contraction). That is, firms increase output and prices - a movement along the SRAS curve.

Since prices are currently high (P_1) and the situation is moving into the long run, people will expect prices to continue to be high. There is an inflationary gap.

Stagflation

In the resource market, a supply shock such as a drought or high oil prices is reflected by a leftward shift of the supply curve of resources. The price of resources, and thus the cost of production, increases. Assuming prices in the goods and services markets are unchanged, the higher costs may be one of the factors that contribute to a recession.

As the SAS curve shifts leftward, real GDP decreases and the price level rises. The combination of recession with inflation is called **stagflation**.

7. Economic Growth and Sustainability

```
    <b>Economic growth</b> is the sustained expansion of production possibilities measurec
```

The standard of living depends on real GDP per person. Real GDP per person is real GDP divided by the population. It grows only if real GDP grows faster than the population grows.

The Production Function and Potential GDP

$$Y = A\,F(L, K)$$

The quantity of real GDP supplied, Y, depends on the quantity of labor, L, the quantity of capital, K, and the state of technology, A (total factor productivity).

This equation shows that output depends on inputs and the level of technology.

- More inputs mean more output. That is, the marginal product of labor (the increase in output generated by increased labor) and the marginal production of capital (the increase in output generated by increased capital) are both positive.
- The higher the level of technology, the more output is produced for a given level of inputs.

The law of diminishing returns: As the quantity of one input increases with the quantities of all other inputs remaining the same, output increases but by ever smaller increments. As capital per hour of labor rises, output rises (the marginal product of capital is positive) but output rises less at high levels of capital than at low levels. This is the key explanation of why the economy reaches a steady state rather than growing endlessly.

Convergence is the process of one economy catching up with another economy. According to the neoclassical growth theory, countries with a low level of capital would have a higher marginal product of capital because of diminishing returns and hence attract more investment and grow faster.

Growth in Y = Growth in technology + W_L (growth in labor) + W_C (growth in capital)

where W_L and $W_C = 1 - W_L$ are the shares of labor and capital in GDP.

Sources of Economic Growth

There are five important sources of growth for an economy:

- Labor supply is the quantity of the work force. It is determined by population growth, the labor force participation rate, and net immigration.
- Human capital measures the quality of the labor force. Human capital acquired through education, on-the-job training, and learning-by-doing is the most fundamental source of economic growth. It is the source of increased labor productivity and technological advance.
- Physical capital results from saving and investment decisions. The accumulation of new capital increases capital per worker and labor productivity.
- Technology. Technological change - the discovery and the application of new technologies and new goods - has contributed immensely to increasing labor productivity. It is the main factor affecting

economic growth in developed countries.
- Natural resources account for some of the differences in growth among countries.

Measures of Sustainable Growth

Labor productivity is the quantity of real GDP produced by an hour of labor. The growth of labor productivity depends on physical capital growth, human capital growth, and technology advances.

Potential GDP = Aggregate hours worked x Labor productivity

Potential growth rate = Long-term growth rate of labor force + Long-term labor productivity growth rate

Understanding Business Cycles

1. Overview of the Business Cycle

```
The <b>business cycle</b> is the fluctuations in the general level of economic activit
```

- **Peak**

  When most businesses are operating at capacity level and real GDP is growing rapidly, a business peak or boom is present.

- **Contraction**

  Aggregate business conditions are slow, real GDP grows at a slower rate or even declines, and the unemployment rate increases. This indicates that the economy begins the contraction, or recessionary, phase of a business cycle.

  - Firms start to cut hours and freeze hiring, followed by outright layoffs.
  - As final demand starts to fall off the downturn in investment spending usually occurs *abruptly*.
  - Inventories accumulate involuntarily, and firms cut production below even reduced sales levels to let their inventories decline. This eventually accelerates the economic downturn.

- **Trough**

  The bottom of the contraction phase is referred to as the recessionary trough. When a contraction is prolonged and characterized by a sharp decline in economic activity, it is called a **depression**.

- **Expansion**

  After the downturn reaches bottom and economic conditions begin to improve, the economy enters the expansion phase of the cycle. Here business sales rise, GDP grows rapidly, and the rate of unemployment declines.

  - Hiring new workers is a costly process. Firms will wait until it's clear that the economy is in this phase. They will then start full-time rehiring as overtime hours rise.
  - As inventories dwindle, businesses ultimately find themselves short of inventory. As a result, they start increasing inventory levels by producing output greater than sales, leading to an economic expansion. This expansion continues as long as the rate of increase in sales holds up and producers continue to increase inventories at the preceding rate.
  - Changes in sales can result in magnified percentage changes in investment expenditures. Suppose a firm is operating at full capacity. When sales of its goods increase, output will have to be increased by increasing plant capacity through further investment. This accelerates the pace of economic expansion, which generates greater income in the economy, leading to further increases in sales. Thus, once the expansion starts, the pace of investment spending accelerates.

> Orders for new equipment are early signals of recovery. Since it usually takes longer to plan and complete large construction projects than for equipment orders, construction projects may be less influenced by business cycles.

The expansion eventually blossoms into another peak.

2. Credit Cycles and Their Relationship to Business Cycles

```
    <p> </p>The <b>credit cycle</b> describes recurring phases of easy and tight borrowing
```

Credit availability is determined by risk and profitability to the lenders. The lower the risk and greater profitability to lenders, the more they are willing to extend loans. During high access to credit in the credit cycle, risk is reduced because investments in real estate and businesses are increasing in value; therefore, the repayment ability of corporate borrowers is sound. Individuals are also more willing to take out loans to spend or invest because funds are cheaper and their incomes are stable or on the rise.

During the contraction period of the credit cycle, interest rates climb and lending rules become more strict, meaning that less credit is available for business loans, home loans, and other personal loans. The contraction period continues until risks are reduced for the lending institutions, at which point the cycle troughs out and then begins again with renewed credit.

The average credit cycle tends to be *longer, deeper and sharper* than the business cycle, because it takes time for a weakening of corporate fundamentals or property values to show up. In other words, there can be an over-extension of credit in terms of amount and period, as spectacularly demonstrated last decade.

3. The Impact of the Business Cycle

```
    <p> </p>Key economic variables change throughout the business cycle.
```

Unemployment increases during business cycle recessions and decreases during business cycle expansions (recoveries). However, **employment levels** follow the cycle *with a delay* as companies initially use overtime before hiring after the onset of recovery and then reduce overtime before reducing employment as the economy passes its peak and enters contraction.

Capital spending is highly sensitive to changes in economic activity, and fluctuates with the business cycle.

The size of **inventory** is small relative to the size of the economy, but inventories can fluctuate dramatically over the business cycle. **Inventory-sales ratio** measures the inventory available for sale to the level of sales. Analysts pay attention to inventories to gauge the position of the economy in the cycle.

Consumer spending, the largest component of output, follows cyclical patterns as workers make decisions based their levels of income, income growth, and employment outlook. For example, an increase in durable spending may be an early indication of economic recovery. Consumer spending is however less cyclical than investment spending though.

The **housing sector** is very sensitive to interest rates. It is also affected by the rate of family formation and expectation of housing price increases.

ImportsExports depend more on cycles of foreign economies. The currency exchange rate plays an important role in this sector.

4. Theories of the Business Cycle

```
    We consider a few fundamentally different theories of the business cycle.<p> </p>
```

Neoclassicial and Austrian Schools - Self-Correcting Economy

The neoclassical economists assumed that the economy would not operate with real GDP (Y) away from the level of natural real GDP (Y^N) for any length of time; if $Y < Y^N$, then firms would be producing below capacity, and would tend to cut nominal wages and prices, which would continue until Y^N was again reached. If $Y > Y^N$, then above-capacity production could support hikes in nominal wages and prices, until real output fell back to Y^N. The consequence was no business cycle in real GDP.

The Austrian school economists argued that business cycles are caused by governments as they try to increase GDP and employment.

Keynesian School - No Self-Correction

It is the changes in output and employment, not price changes, that restores equilibrium in the Keynesian model.

- Aggregate demand fluctuations. Demand is the driving force of the economy. Expectations are the most significant influence on aggregate demand.
- Aggregate supply response. Wages and prices are highly inflexible, particularly in a downward direction. With a sticky price level, the short-run aggregate supply curve is horizontal at a fixed price level.
- Policy response is needed. When aggregate expenditures are deficient, there are no automatic forces capable of assuring full employment. Recessions and depressions result when total spending falls because businesses reduce production. Therefore, government intervention is required to keep the economy at full employment capacity without inflation.

To reduce economic disturbances, fiscal policy must be put into effect at the proper time in the business cycle. Policy changes take time; thus, when they take effect, the recession or inflationary overheating may have passed.

Monetarist School

The economy is self-regulating and it will normally operate at full employment if monetary policy is properly timed and the pace of money growth is kept steady.

The quantity of money is the most significant influence on aggregate demand.

The New Classical Model - Policy Ineffectiveness

Real business cycle theory assumes that real shocks to the economy are the primary cause of business cycles.

- Adverse cost shocks lead to a recession, as individuals should spend less time working because it is not profitable.
- Favorable cost shocks lead to a boom period because it is advantageous to produce as much as possible.

Production fluctuates because of the changing value of output and the changing productivity of the economy. Government intervention is generally not necessary because it may exacerbate this fluctuation or delay the convergence to equilibrium.

The **Neo-Keynesian school** assumes that the prices of most goods don't change daily (sticky price, or menu cost), as the cost of changing prices may outweigh the benefits of changing prices. Therefore, markets do not reach equilibrium quickly.

5. Economic Indicators

```
    <b>Economic indicators</b> are statistics on macroeconomic variables that help in unde
```

Economic indicators can be leading, lagging, or coincident, which indicates the timing of their changes relative to how the economy as a whole changes.

- **Leading**: Leading economic indicators are indicators which change *before* the economy changes. Stock market returns are a leading indicator, as the stock market usually begins to decline before the economy declines and improves before the economy begins to pull out of a recession. Leading economic indicators are the most important type for investors as they help predict what the economy will be like in the future.

- **Lagging**: A lagging economic indicator is one that does not change direction until a few quarters after the economy does. The unemployment rate is a lagging economic indicator as unemployment tends to increase for 2 or 3 quarters after the economy starts to improve.

- **Coincident**: A coincident economic indicator is one that simply moves at the same time the economy does. The Gross Domestic Product is a coincident indicator.

No single indicator is able to forecast accurately the future direction of the economy. In the U.S., economists often refer to the Conference Board's **diffusion index** when judging the moves in the leading index. The diffusion index can measure the breadth of a move in any BCI index, showing how many of an index's components are moving together with the overall index. The index generally turns down prior to a recession and turn up before the beginning of a business expansion.

However, there are two problems with the index.

- There has been significant variability in the lead time of the index. For example, a downturn in the index is not always an accurate indicator of the future.
- The index has often given false alarms. For example, a recession forecasted by a decline in the index does not materialize.

While we cannot predict the future perfectly, economic indicators help us understand where we are and where we are going.

6. Unemployment

```
The U.S. Census Bureau conducts monthly surveys to determine the status of the labor f
```

To be counted as **unemployed**, a person must be actively seeking employment but currently without work.

The **unemployment rate** is the percentage of persons in the labor force who are unemployed. This is a key parameter of conditions in the aggregate labor market.

There are special categories of unemployment, such as:

- Long-term unemployed: people who are unemployed because they do not have the skills required by the openings or reside far from the jobs.
- Frictionally unemployed: people who are not working because they are in between jobs.

Unemployment rate tends to be a *lagging* instead of a leading indicator of the economy, confirming but not foreshadowing long-term market trends. It tends to peak after the trough of the business cycle and bottom after the peak of the business cycle. This is because:

- The employment data is compiled afterwards.
- Employers are reluctant to lay people off when the economy turns bad. For large companies, it can take months to put together a layoff plan. Companies are even more reluctant to hire new workers until they are sure the economy is well into the expansion phase of the business cycle.

Underemployed is a measure of employment and labor utilization in the economy. It looks at how well the labor force is being utilized in terms of skills, experience, and availability to work.

Discouraged workers believe that continuing the job search is fruitless and thus give up looking for a job. They wish to work but because they are not actively searching for work they are excluded from the labor force and are not counted in the unemployment rate. The unemployment rate may fall during recessions as discouraged workers leave the labor force.

Voluntary unemployment refers to the number of persons in an economy without jobs because they choose to be unemployed.

Analysts also use other measures to get a better picture of the employment cycle. These measures include the size of payrolls, hours worked, and the use of temporary workers.

7. Inflation

```
<b>Inflation</b> is a continuing rise in the general level of prices of goods and serv
```

- It is a rise in the price level, not in the price of a particular commodity. Individual prices rise and fall all the time in a market economy, reflecting consumer choices or preferences and changing costs.
- It is an ongoing process, not a one-time jump in the price level.

There are different types of inflation.

- **Deflation** is a decrease in the general price level of goods and services.
- **Disinflation** is a slowing in the rate of increase in the general price level.
- **Hyperinflation** indicates a very high and increasing rate of inflation.

The **annual inflation rate** is simply the percentage change in the price index (PI) from one year to the next:

For example, the CPI is 115 for 2010 and 120 for 2011. The inflation rate during 2011 is: (120 - 115)/115 = 4.35%.

The **Laspeyres index** uses the *same group* of commodities purchased in the base period.

- Advantages: It requires quantity data from only the base period. This allows a more meaningful comparison over time. The changes in the index can be attributed to changes in the price.
- Disadvantages: It does not reflect changes in buying patterns over time. It may also overweight goods when prices increase.

The **Paasche index** uses the *current composition* of the basket. It tends to understate inflation.

The **Fisher index** is the geometric mean of the two indices.

Many countries use their own consumer price indices to track domestic inflation. These indices have different names and baskets.

Inflation is not simply a matter of rising prices. In the long run, inflation occurs if the quantity of money grows faster than potential GDP. In the short run, there are endemic and perhaps diverse reasons for causes at the root of inflation.

Inflation can result from either an increase in aggregate demand (demand-pull inflation) or a decrease in aggregate supply (cost-push inflation).

- **Cost-push inflation** is a result of decreased aggregate supply as well as increased costs of production (itself a result of different factors). It basically means that prices have been "pushed up" by increases in the costs of any of the production factors (money wage rate and money price of raw materials) when companies are already running at full production capacity. Increased costs are passed on to consumers, causing a rise in the general price level (inflation).

    - The **non-accelerating inflation rate of unemployment (NAIRU)**, or the **natural rate of unemployment (NARU)**, is defined as the rate of unemployment when the rate of wage inflation is stable.

    - The **unit labor cost (ULC)** indicator is calculated as total compensation per worker divided by total output per worker. Higher labor costs may pass through to prices.

- **Demand-pull inflation** occurs when total demand for goods and services exceeds total supply. Buyers, in essence, "bid prices up" and cause inflation. This excessive demand usually occurs in an expanding economy.

  The increase in aggregate demand that causes demand-pull inflation can be the result of various economic dynamics. For example, the authorities may allow the money supply to grow faster than the ability of the economy to supply goods and services, so there is "too much money chasing too few goods." Increases in government purchases and depreciations of local exchange rates can also increase aggregate demand and start demand-pull inflation. However, only an ongoing increase in the quantity of money can sustain it.

If an economy identifies what type of inflation is occurring (cost-push or demand-pull), then the economy may be better able to rectify (if necessary) rising prices and loss of purchasing power.

Economics (2)

Monetary and Fiscal Policy

1. What is Money?

```
<b>Fiscal policy</b> refers to the use of government expenditure, tax, and borrowing a
```

The Functions of Money

Money performs three basic functions.

- It serves as a **medium of exchange** to buy and sell goods and services. Money simplifies and reduces the costs of transactions.

  In the absence of money, a **barter** economy would exist. Acquiring a belt, for example, would entail finding a belt maker who happened to want what you had to offer in exchange, making transactions tedious, enormously costly, and inefficient.

  Money permits us to realize the enormous gains from the specialization, division of labor, and mass-production processes that underlie our modern standard living.

- It is used as an **accounting unit** to compare the value and cost of things. As a unit of measurement, like a centimeter, money is used by people to post prices and keep track of revenues and costs.

- It provides a **way of storing value** to allow the movement of purchasing power from one period to another. Although it is not the only way of storing value, it is the most liquid of all assets, due to its function as the medium of exchange. However, many methods of holding money do not yield an interest return and the purchase power of money will decline during a time of inflation.

The Money Creation Process

Reserves are the cash in a bank's vault and deposits at Federal Reserve Banks. Under the fractional reserve banking system, a bank is obligated to hold a minimum amount of reserves to back up its deposits. Reserves held for that purpose, which are expressed as a percentage of a bank's demand deposits, are called required reserves. Therefore, the required reserve ratio is the percentage of a bank's deposits that are required to be held as reserves.

Banks create deposits when they make loans; the new deposits created are new money.

Example

Suppose the required reserve ratio in the U.S. is 20%, and then suppose that you deposit $1,000 cash with Citibank. Citibank keeps $200 of the $1,000 in reserves. The remaining $800 of excess reserves can be loaned out to, say, John. After the loan is made, the money supply increases by $800 (your $1,000 + John's $800). After getting the loan, John deposits the $800 with Bank of America (BOA). BOA keeps $160 of the $800 in reserves and can now loan out $640 to another person. Thus, BOA creates $640 of money supply. The process goes on and on. With each deposit and loan, more money is created. However, the money creation process does not create an infinite amount of money.

The **money multiplier** is the amount by which a change in the monetary base is multiplied to calculate the final change in the money supply. Money Multiplier = 1/b, where b is the required reserve ratio. In our example, b is 0.2, so money multiplier = 1/0.2 = 5.

Definitions of Money

There are different definitions of money. The two most widely used measures of money in the U.S. are:

- The M1 Money Supply: cash, checking accounts and traveler's checks. This is the narrowest definition of the money supply. This definition focuses on money's function as a medium of exchange.
- The M2 Money Supply: M1 + savings + small time deposits + retail money funds. This definition focuses on money's function as a medium of exchange and store of value.

Credit cards are not purchasing power, but instead are a convenient means of arranging a loan. Credit is a liability acquired when one borrows funds, while money is a financial asset that provides the holder with future purchasing power. However, the widespread use of credit cards will tend to reduce the average quantity of money people hold.

Deposits are money, but checks are not - a check is an instruction to a bank to transfer money.

The Quantity Theory of Money

Money Supply (M) x Velocity of Money (V) = Price (P) x Real Output (Y)

- The **velocity of money** is the average number of times a dollar is used to purchase final goods and services during a year. It is computed as V = GDP/Money Supply = PY/M. In essence, it is the turnover rate of money. For example, if the nominal GDP is $200 billion and the money supply is $40 billion: V = 200/40 = 5.
- The equation of exchange reflects two ways of viewing GDP: the left side reflects the monetary flow of expenditures on final products and the right side reflects the sum of the price (P) times the output (Y) of each final product purchased during the period.

If both Y and V are constant, then the equation indicates that an increase in money supply will lead to a proportional increase in price level.

This equation of exchange leads to the **quantity theory of money**, which hypothesizes that a change in the money supply will cause a proportional change in the price level because velocity and real output are unaffected by the quantity of money.

2. The Demand for and Supply of Money

```
   At any given interest rate, the amount of wealth that households and businesses desire
```

People hold (demand) money to conduct transactions, to deal with emergencies (precautionary motive), and for speculative activities.

There is an *inverse* relationship between the demand for money and interest rates when all other influences on the amount of money that people wish to hold remain the same.

A rise in the interest rate brings a decrease in the quantity of money demanded. A fall in the interest rate brings an increase in the quantity of money demanded.

The money supply schedule is vertical since domestic supply of money is determined by the central bank and reserve requirements. The supply of money is not affected by changes in the interest rate.

Money market equilibrium occurs when people are willing to hold all the money supplied by the monetary authorities at the prevailing interest rate; the supply of money equals the demand for money. It occurs at i_e in the diagram.

However, disequilibrium exists at the interest rate i_2.

People are not willing to hold all the money supplied by the monetary authorities as money balances. Instead, they demand high-interest earning assets such as bonds. This will increase the price of bonds, which in turn reduces their interest yield, driving i_2 down towards i_e and eventually restoring equilibrium.

Disequilibrium also exists at the interest rate i_3.

People would like to hold more money balances than the monetary authorities are willing to supply. The resultant low demand for bonds reduces their prices, thus increasing their interest rate (yield), and slowly restores equilibrium at i_e.

The Fisher Effect

$$R_{nom} = R_{real} + R_{inflation} + \text{risk premium}$$

The nominal rate of interest is comprised of three components:

- A real required rate of return.
- A component to compensate lenders for future inflation.
- A risk premium to compensate lenders for uncertainty.

3. The Roles of Central Banks

```
    Central banks all have similar roles:<p> </p><ul class="notes">
```

- Issue currency
- The government's bank, and bank of the banks
- Lender of last resort to the banking sector
- Regulator and supervisor of the payments system
- Set monetary policy
- Regulate banking system

4. The Objectives of Monetary Policy

```
    A nation's monetary policy objectives and the framework for setting and achieving thos
```

The key objective is price stability. It is the source of maximum employment and moderate long-term interest rates.

The Costs of Inflation

Unanticipated inflation is an increase in the general level of prices that was not expected by most decision makers. It is a surprise to most individuals. For example, if someone anticipates an inflation rate of 3% but the actual inflation rate turns out to be 10%, it will catch that person off guard.

Unanticipated inflation redistributes income, creates uncertainty, and can have a potentially destabilizing impact on the economy.

Monetary Policy Tools

Central banks manipulate the money base that creates the change in the money supply. When following an expansionary monetary policy, they increase the growth rate of the money supply. Conversely, when following a restrictive monetary policy, they reduce the growth rate of the money supply.

Central banks have three major means of controlling the money stock.

Open Market Operations. Since it can be undertaken easily and quietly, this is the most common tool used by a central bank to alter the money supply. For example, to increase the money supply, the central bank buys government securities from commercial banks. This increases the money supply.

The Central Bank's Policy Rate. Sometimes banks find themselves in a position where they are not holding enough bank reserves relative to the value of the deposits they hold (i.e., they have extended too many loans!). As a result, they may need to acquire a short-term loan from the central bank to cover this shortage of required reserves. They may apply to the central bank for a loan; the interest charged on the loan is known as **repo rate** (in the U.S. it is called the **discount rate**).

An increase in the policy rate is restrictive on money supply - it tends to discourage banks from shaving their excess reserves to a low level.

In the U.S., borrowing from the Fed amounts to less than one-tenth of 1% of the available loanable funds of banks. A bank can go to the federal funds market to borrow to meet its reserve requirements. The market is where banks with excess reserves extend short-term loans to those banks seeking additional reserves. The interest rate in this market is called the **federal funds rate**. The federal funds rate and the discount rate tend to move together.

Reserve Requirements. The central bank sets the rules. Since banking institutions will want to hold interest-earning assets rather than excess reserves, an increase in the reserve requirements will reduce the supply of money, and vice versa. However, the central banks of developed countries have seldom used their regulatory power over reserve requirements to alter the supply of money, due to its disruptive impact on banking operations - small changes in reserve requirements can sometime lead to large changes in the money supply.

The Transmission Mechanism

When a central bank lowers its official interest rate:

- Other short-term interest rates fall. Short-term rates move closely together and follow the official interest rate.
- The exchange rate falls. The exchange rate responds to changes in the domestic interest rate relative to the interest rates in other countries (the interest rate differential). But other factors are also at work, which make the exchange rate hard to predict.
- The quantity of money and the supply of loanable funds increase. This is because the fall in the interest rate increases reserves and increases the quantity of deposits and bank loans created. The fall in the interest rate also increases the quantity of money demanded.
- The long-term interest rate falls. Long-term rates move in the same direction as the official interest rate. The long-term real interest rate influences expenditure plans.
- Consumption expenditure, investment, and net exports increase. The change in aggregate expenditure plans changes aggregate demand, real GDP, and the price level, which in turn influence the goal of the inflation rate and output gap. Aggregate demand increases.
- Real GDP growth and the inflation rate increase.

When the Fed raises the federal funds rate, the ripple effects go in the opposite direction.

5. Monetary Targeting Rules

```
    <b>Inflation Targeting</b><p> </p>
```

Price stability is the primary goal of inflation-targeting monetary policy strategy. Inflation is usually defined as a range of permissible values (e.g., 1%-3%) rather than as a point value (e.g., 2.4%). The definition of inflation also varies from country to country.

There are three key concepts:

Central bank independence. Central bank independence exists on two dimensions. Goal independence is the freedom that the central bank has to select the objectives of monetary policy, whether they are low inflation, the target rate of unemployment, the level of GDP, etc. Instrument independence is the freedom that the central bank has to pick appropriate policies to produce a certain outcome in the economy. Most inflation-targeting countries only lay out the goals and not the operating procedures; the central bank does have operational independence.

Credibility. Central bankers who are unable to credibly convince the public that they are serious about fighting inflation will be faced with a high inflation rate as a result.

Transparency. It is well known that credibility requires transparency. The benefits of transparency are obvious: it improves the efficiency of monetary policy, allows for a more effective management of expectations, and promotes the discussion and evaluation of monetary policy.

Exchange Rate Targeting

Many countries have viewed pegging their nominal exchange rate to a stable, low-inflation foreign currency as a means of achieving domestic price stability. In a sense, countries that target their exchange rates against an anchor currency attempt to "borrow" the foreign country's monetary policy credibility. However, this monetary policy deprives the central bank of its ability to respond to idiosyncratic domestic shocks. Such countries can become prone to speculation against their currencies.

6. Contractionary and Expansionary Monetary Policies and the Neutral Rate

```
    An <b>expansionary monetary policy</b> decreases the interest rate in order to increas
```

The idea behind the concept of **neutral rate of interest** is that there might be a rate of interest that neither deliberately seeks to stimulate aggregate demand and growth nor deliberately seeks to weaken growth from its current level. In other words, a neutral rate of interest would be one that encourages a rate of growth of demand close to the estimated trend rate of growth of real GDP.

The neutral rate is a useful method of measuring the stance of monetary policy. It has two components:

$$\text{Neutral rate} = \text{Trend growth} + \text{Inflation target}$$

When the neutral rate is reached, the state of equilibrium is attained, implying that the economy is now well-balanced and the price level is stable.

Certainly there can be no such thing as an exact measure of the neutral rate, and it will differ from country to country.

A **demand shock** is a sudden surprise event that increases or decreases demand. If inflation is caused by an unexpected increase in aggregate demand, a contractionary monetary policy might be appropriate, to cause inflation to fall. However, if inflation is caused by a **supply shock** such as a sudden increase in oil price, a contractionary monetary policy might make the situation worse.

Limitations of Monetary Policy

Central banks cannot control the money supply. This is because:

- They cannot control the amount of money that households and corporations put in banks on deposit.
- They cannot control the willingness of banks to create money by expanding credit.

In **quantitative easing** (**QE**), a central bank buys any financial assets to inject money into the economy. It is different from the traditional policy of buying or selling government bonds to keep market interest rates at a specified target value. Risks include the policy being more effective than intended, spurring hyperinflation, or the risk of not being effective enough, if banks opt simply to pocket the additional money in order to increase their capital reserves.

7. Fiscal Policy: Roles, Objectives, and Tools

```
<b>Fiscal policy</b> refers to the use of government expenditure, tax, and borrowing a
```

Government expenditures include transfer payments, the purchase of goods and services (current government spending), and capital expenditure.

Government revenues are generated through **taxes**. There are direct and indirect taxes. Direct taxes are difficult to change without considerable notice. Indirect taxes can be adjusted almost immediately.

The four desirable attributes of a tax policy are simplicity, efficiency, fairness, and revenue sufficiency.

A **budget** is the annual statement of the government's expenditures and tax revenues. A **balanced budget** implies that current government revenue is equal to current government expenditures.

A **budget deficit** exists when total government spending exceeds government revenue. A budget deficit is financed through the issuance of government securities. Such issuance of securities adds to the **national debt**.

A **budget surplus** occurs when revenues exceed spending. Under a budget surplus, excess revenue is applied to the total outstanding debt accumulated during prior periods, therefore reducing it by the amount of the surplus.

There are arguments for and against being concerned with the size of a fiscal deficit.

The arguments against being concerned about national debt:

- The debt is owed internally by fellow citizens.
- Some borrowed money may have been used for capital investment projects or enhancing human capital.
- Large deficits require tax changes which may be desirable.
- Richardian equivalence: the timing of any tax change does not affect consumers' change in spending.
- Debt could improve employment.

The arguments for being concerned about national debt:

- Higher deficits -> higher tax rates -> less incentive to work and invest -> lower long-term growth
- The central bank may have to print money to finance a deficit. This may lead to high inflation.

The **crowding-out effect** is the reduction in private spending as a result of higher interest rates generated by budget deficits that are financed by borrowing in the capital market. It suggests that budget deficits will exert less impact on aggregate demand than the basic Keynesian model implies. Because financing the deficit pushes up interest rates, budget deficits will tend to retard private spending, particularly spending on investment. This reduction will at least partially offset additional spending emanating from the deficit.

Multipliers

As disposable income increases, consumption expenditures increase, but by a smaller fraction than the increase in income. This is reflected in the **marginal propensity to consume** (**MPC**), which is less than one.

Marginal propensity to save (**MPS**) is defined as additional savings divided by additional current disposable income.

$$MPC + MPS = 1$$

If we add tax rate t, then:

$$MPC + MPS = 1 - t$$

An **expenditure multiplier** is the ratio of the change in equilibrium output relative to the independent change in consumption, investment, and government spending or spending on net exports that brings about the change.

The **government purchases multiplier** is the magnification effect of a change in government purchases of goods and services on aggregate demand. It exists because government purchases are a component of aggregate expenditure; an increase in government purchases increases aggregate income, which induces additional consumption expenditure.

The **tax multiplier** is the magnification effect a change in taxes on aggregate demand. An increase in taxes decreases disposable income, which decreases consumption expenditure, aggregate expenditure, and real GDP.

The two multipliers are called the **fiscal multiplier**: $1/[1 - c (1 - t)]$, where c is the MPC and t is the tax rate.

The **balanced budget multiplier** is the magnification effect of a *simultaneous* change in government purchases and taxes on aggregate demand. A $1 increase in government purchases increases aggregate demand initially by $1, but a $1 increase in taxes decreases consumption expenditure by less than $1 initially, so a $1 increase in both purchases and taxes increases aggregate demand.

The **structural surplus or deficit** is the surplus or deficit that would occur if the economy was at full employment and real GDP was equal to potential GDP.

8. Active and Discretionary Fiscal Policy

```
    Fiscal policy actions seek to stabilize the business cycle by changing aggregate demar
```

- *Discretionary.* Discretionary fiscal policy is a policy action that is initiated explicitly by the government.
- *Automatic.* Automatic fiscal policy is a change in fiscal policy triggered by the state of the economy.

To reduce economic disturbances, fiscal policy must be put into effect at the proper time in the business cycle. Policy changes take time; thus, when they take effect, the recession or inflationary overheating may have passed. For example, during an economic downturn, a government uses expansionary fiscal policy to stimulate aggregate demand. Suppose, by the time the expansionary fiscal policy starts to exert its primary impact, the economy's self-corrective mechanism has restored full employment capacity. Therefore, the stimulus injected by expansionary fiscal policy will result in excessive demand and inflation, causing more economic instability.

The use of discretionary fiscal policy is hampered by three time lags:

- *Recognition lag.* There is usually a time lag between when a change in policy is needed and when its need is widely recognized by policymakers. Forecasting a forthcoming recession or boom is a highly imperfect science.
- *Action lag.* There is generally a lag between the time when the need for a fiscal policy change is recognized and the time when it is actually instituted. The time required to change tax laws and government expenditure programs is quite lengthy.
- *Impact lag.* Even after a policy is adopted, it may be 6 to 12 months before its major impact is felt.

Changes in fiscal policy must be timed properly if they are going to exert a stabilizing influence on an economy. However, the use of fiscal policy to calm the business cycle is very difficult; it may accentuate the corrective action of the economy rather than correct the problem for which it was intended. In the real world, a discretionary change in fiscal policy is like a double-edged sword - it has the potential to do harm as well as good. If timed correctly, it will reduce economic instability. If timed incorrectly, however, the fiscal change will increase rather than reduce economic instability.

Automatic Stabilizers

Automatic stabilizers apply stimulus during a recession and restraint during a boom even though no legislative action has been taken. Their major advantage is that they institute counter-cyclical fiscal policy without the delays associated with policy changes that require legislative action. During a recession, they trigger government spending without the authorization of Congress (unemployment compensation and welfare programs). During inflationary overheating, they take spending power out of the economy without the delays caused by legislative actions, thereby minimizing the problem of proper timing.

Income taxes and transfer payments are automatic stabilizers.

When the economy starts to fall into a recession, unemployment increases. Government payments for unemployment compensation will increase while government receipts from the employment tax that finances unemployment benefits will decline. As a result, the unemployment compensation program automatically promotes a budget deficit. Similarly, when the economy expands into an inflationary boom, the program promotes a budget surplus.

When the economy expands into an inflationary boom, personal income will grow sharply. As a result, more people will fall into the "tax due" category, and others will be pushed into higher tax brackets. Therefore, income tax revenues will rise more rapidly than income, reducing the momentum of consumption growth. In addition, higher tax revenues will promote a budget surplus.

9. Interrelationships between Fiscal and Monetary Policy

```
   Governments use fiscal and monetary policies to respond to changes in the business cyc
```

- Central banks can respond with lower interest rates and an expanded money supply. This will lead to an increase in investment and consumer spending. Since investment spending results in a large capital stock, incomes in the future will also be higher. However, inflation may result.
- Keynes argued that the government should boost spending but that if this is financed by higher borrowing, it may result in higher interest rates and lower investment. The net result (by adjusting the increase in G) is the same increase in aggregate demand. However, since investment spending is lower, the capital stock is lower than it would have been and future incomes will be lower.

Monetary policy and fiscal policy are not interchangeable. When the economy is in a recession, monetary policy may be ineffective in increasing spending and income. In this case, fiscal policy might be more effective in stimulating demand. Other economists disagree; they argue that changes in monetary policy can impact consumer and business behaviour quite quickly and strongly.

However, there may be factors which make fiscal policy ineffective aside from the usual crowding-out phenomena. Future-oriented consumption theories based round the concept of rational expectations hold that individuals "undo" government fiscal policy through changes in their own behaviour - for example, if government spending and borrowing rises, people may expect an increase in the tax burden in future years, and therefore increase their current savings in anticipation of this.

Monetary and fiscal policies also differ in the speed with which each takes effect. The time lags are variable and they can conceivably work against one another unless the government and central bank coordinate their objectives.

International Trade and Capital Flows

1. GDP vs. GNP

```
    GDP is the total market value of all domestically produced final goods and services fc
```

- Only **final goods and services** count: GDP includes goods and services purchased by final users. **Intermediate goods** purchased for resale or for the production of another good or service are excluded, to avoid double-counting. Their value is embodied in the value of the goods purchased by the end user.

- GDP is a "*flow*" variable; it measures the market value of production that flows through the economy.

- Financial transactions and income transfers (e.g., social security and welfare payments) are excluded because they represent *exchanges*, not productions, of goods and services. GDP counts transactions that add to current production.

- GDP counts only goods and services produced domestically, whether by citizens or foreigners.

- It includes only goods produced during the current period. Thus, sales of used goods are not counted in GDP. However, sales commissions count toward GDP because they involve services provided during the period.

GNP is the total market value of all final goods and services produced by the citizens of a country. It measures the output that is produced by the "**nationals**" of a country. This figure is the output generated by the labor and capital owned by the citizens of the country, regardless of whether that output is produced domestically or abroad. Consider the case of the United States. GNP is the income earned by Americans, regardless of whether that income is earned in the United States or abroad. It omits the income foreigners earn in the United States, but counts the income that Americans earn abroad. **It is equal to GDP minus the net income of foreigners.**

GNP = GDP + Income received by citizens for factors of production supplied abroad - Income paid to foreigners for the contribution to domestic output

In short, GNP measures the worldwide output of a nation's citizens while GDP measures the domestic output of the nation.

- In general, the bulk of output is produced domestically using resources owned by nationals of the country. Thus, GDP often differs only slightly from GNP.
- These two measures differ substantially only when a country attracts a large number of foreign workers or investments (the country's GDP will exceed its GNP).
- If a relatively large number of a country's citizens work abroad, or its citizens have made substantial investments abroad, the country's GNP will exceed its GDP.

2. International Trade

```
    <b>Benefits and Costs</b><p> </p>
```

Here are the benefits:

- International trade and specialization result in lower prices and higher domestic consumption for imported products, and higher prices and lower domestic consumption for exported products.
- International trade permits the residents of each nation to concentrate on the things they do best (produce at a low cost) while trading for those they do least well.
- Industries experience greater economies of scale.
- Households and firms have greater product variety.
- Resources are allocated more efficiently.

Costs:

- International trade may result in the loss of jobs in developed countries.
- The potential for greater income inequality increases.

Countries have different resource endowments. Some have an abundance of labor whilst others possess fertile lands. Differences in resource endowments result in countries incurring different opportunity costs of production for the same products.

Comparative advantage is the ability to produce a good at a *lower opportunity cost* than others can produce it. Relative costs determine comparative advantage.

If each country has a comparative advantage in producing a specific good, international trade will lead to mutual gain because it allows the residents of each country to:

- Specialize more fully in the production of those things that they do best (i.e., at a lower opportunity cost).
- Import goods when foreigners are willing to supply them at a lower cost than domestic producers.

A nation can have a comparative advantage in producing a good even if it has no **absolute advantage** in producing any good. As long as the relative costs of producing two goods differ in two countries, comparative advantage exists and gains from specialization and trade will be possible. When this is the case, each country will find it cheaper to trade for goods that can be produced only at a high opportunity cost.

We are going to start with some simplifying assumptions:

- Only two countries exist: X and Y.
- Only two products are produced in each country: wine and fish.
- The only resource in each country is labor.

Country X has an absolute advantage in producing both products (i.e., country X's laborers are more efficient than those in country Y).

- Each worker in X is able to produce either 5 fish or 10 bottles of wine per day. The opportunity cost of 1 fish produced in X is 2 bottles of wine.

  A production possibilities frontier (PPF) represents the various maximum combinations of outputs of the two products that a country is able to produce given its current resources.

  Any point on the PPF represents an attainable combination of fish and wine. The slope of the PPF represents the opportunity cost of wine relative to fish. For example, if X was consuming 250 million fish and no wine at point A, the cost of increasing wine consumption to 250 million bottles would be that 125 million fish would have to be sacrificed (i.e., it would move to point B).

- In Y, each worker can produce either 2 fish or 1 bottle of wine per day. One bottle of wine costs 2 fish.

In the absence of trade, a country's consumption possibilities are constrained by the country's production possibilities frontier (i.e., country X could not consume 125 million fish and more than 250 million bottles of wine).

Since there are differences in the opportunity costs of production, trade between the two countries can take place and benefit both.

Let's assume that they set the **terms of trade** at 1 bottle of wine costs 1 fish.

Country X now specializes in producing wine and produces 500 million bottles of wine and no fish. However, it now has the option of consuming all the wine itself or selling it all and obtaining 500 million fish in return. In other words, PPF swivels upwards as shown.

It is now obvious that country X is better off than before, since by selling 250 million bottles of wine in exchange for 250 million fish, point D (consuming 250 million fish and the remaining 250 million bottles of wine), a previously unattainable point, is possible.

How is country Y faring? With the terms of trade set at 1 fish to 1 bottle of wine, Y could specialize in what it does best - producing fish, and produce 500 million fish. It now has the option of selling all of those fish for bottles of wine, so that PPF swivels outwards as shown, to indicate the two options.

Whereas country Y previously could have 250 million bottles of wine and no fish, it can now trade 250 million fish (and be left with 250 million fish) and receive 250 million bottles of wine in exchange. In other words, a previously unattainable point such as F is now attainable.

Note that the nation that has a steeper production possibilities line has a comparative advantage in the good on the *vertical axis* and the other nation has a comparative advantage in the good on the *horizontal axis*.

Comparative advantage results in an expansion of total output.

It also results in mutual gains for each trading partner if each specializes in producing that which it can produce at relatively low cost and uses the proceeds to purchase goods that it could only produce at a higher cost.

For the purposes of trade, it is comparative advantage, not absolute advantage, that matters.

Because trade permits nations to expand their joint output, it also allows each nation to expand its consumption possibilities. Trade between nations will lead to an expansion in total output and mutual gain for each trading partner when each country specializes in the production of goods it can produce at a relatively low cost and uses the proceeds to buy goods it could produce only at a high cost. As long as there is some variation in the relative opportunity cost of goods across countries, each country will always have a comparative advantage in the production of some goods.

Countries gain by consuming output combinations outside their respective production possibilities frontier. Trade does not create winners and losers. It creates only winners.

Models of Comparative Advantage

Adam Smith:

- Principles of division of labor and specialization among countries.
- Each country produces goods that it can produce more of, for the same amount of resources/time.
- Law of absolute advantage.

Ricardian Model:

- Labor is the only production factor.
- Advantageous trade can occur between countries if the countries differ in their technological abilities to produce goods and services. The basis for trade is differences in technology.

Heckscher-Ohlin Model:

- Trade based on resource availability.
- Countries no longed differ by level of technology but by the factors of production with which they are endowed. Goods differ according to the factors of production they require.

3. International Trade Restrictions and Agreements

```
Governments restrict international trade to protect domestic producers from competitic
```

For purposes of international trade policy and analysis:

- A *small country* cannot affect the world price of traded goods. In a small country, trade barriers always generate a net welfare loss arising from distortion of production and consumption decisions and the associated inefficient allocation of resources.

- A *large country*'s production and/or consumption decisions do alter the relative prices of traded goods. Trade barriers can generate a net welfare gain in a large country if it imposes an even larger welfare loss on its trading partners.

Tariffs

A **tariff** is a tax levied on goods imported into a country. It benefits domestic producers and the government at the expense of consumers.

Let's illustrate the impact of a tariff on automobiles. Without a tariff, the world market price (P_W) would prevail in the domestic market. U.S. consumers purchase Q_1 units while U.S. producers supply Q_{d1} units. When the U.S. imposes a tariff (t) on imports of automobiles, U.S. consumers now pay (P_W + t) to purchase automobiles from foreigners. Due to the higher price, U.S. consumers will reduce demand from Q_1 to Q_2, while U.S. producers will increase supply from Q_{d1} to Q_{d2}. Imports from foreigners will reduce from Q_2 to Q_{d2}.

The tariff benefits domestic producers and the government. It protects domestic producers from foreign competition. Consequently, domestic producers can supply goods at a higher price. Domestic producers gain the area S in the form of additional revenue. The government gains the area T in the form of tax revenues collected on imports.

The tariff harms domestic consumers as they have to pay a higher price for fewer goods. They lose the area S + U + T + V. Note that areas U and V are a **deadweight loss** (loss of efficiency) for the economy since they do not benefit either producers or the government.

In effect, a tarrif acts as a subsidy to domestic producers. Potential gains from specialization and trade go unrealized.

Quotas

An **import quota** is a specific limit or maximum quantity (or value) of a good permitted to be imported into a country during a given period. It is designed to restrict foreign goods and protect domestic industries.

Assume that a quota limits imports of automobiles to (Q_2 - Q_{d2}), a quantity below the free trade level of imports (Q_1 - Q_{d1}). Since the quota reduces foreign supply, domestic price will be pushed up to P_2. Due to the higher price, U.S. consumers will reduce demand from Q_1 to Q_2, while U.S. producers will increase supply from Q_{d1} to Q_{d2}. Like a tariff, an import quota benefits domestic producers but harms domestic consumers.

However, different from a tariff, an import quota benefits foreign producers at the expense of the government; with a quota, foreign producers who are granted permission to sell in the domestic market can charge premium prices for the limited supply of foreign goods. The area T represents the gains of those foreign producers. Under a tariff, the government would gain the area T in the form of tariff revenues. This politically granted privilege creates a strong incentive for foreign producers to engage in **rent-seeking activities**. In addition, with a quota foreign producers are prohibited from selling additional units regardless of how much lower their costs are relative to those of domestic producers. Therefore, quotas are more harmful than tariffs in many ways.

A **voluntary export restraint** (**VER**) is an agreement between two governments in which the government of the exporting country agrees to restrain the volume of its own exports. Foreign firms and governments will sometimes agree to limit their exports to a country to avoid the imposition of other types of trade barriers. The economic impact of a VER is similar to that of a quota. The main difference:

- A quota is imposed by the importer.
- A VER is imposed by the exporter.

Export Subsidies

An export subsidy is a government policy to encourage export of goods and discourage sale of goods in the domestic market. The subsidy ends up costing the government instead of generating revenue. Also, unlike an import tariff, an export subsidy increases the amounts traded.

The distortion of production, consumption, and trade decisions generates a welfare loss. This welfare loss is greater for a large country because increased production and export of the subsidized product reduces its global price - that is, it worsens the country's terms of trade.

Example 1

Refer to the graph below. What tariff would the government have to impose on tomatoes imported from Mexico to have the same effect as a quota of Q_1?

A tariff that shifts supply to where it intersects demand at P_3 and Q_1 would be the equivalent of a quota of Q_1. This would happen with a tariff of P_3-P_1.

Example 2

Refer to the graph below. If this graph represents the supply of and demand for an imported product, a tariff of t will result in revenue for the government, shown by area ______.

A tariff of t shifts the supply curve up from S_0 to S_1. Quantity sold is now OE. Tariff revenue is t times quantity OE, shown by the rectangle BDGH.

Example 3

Refer to the graph below. With a quota of 600 tons on lumber imported from Canada, the revenue the government would collect from the import of lumber would be:

Answer: The government does not collect revenue with a quota.

Example 4

Refer to the graph below. Demand and supply are initially D and S1, respectively. Which of the following best describes the effect of a $.50 per pound tariff on Danish hams imported into the United States?

Answer: A tariff shifts the supply curve to the left by the amount of the tax. Equilibrium price is determined where quantity demanded equals quantity supplied, at $2.25 per pound and 60 thousand pounds.

Trading Blocs, Common Markets, and Economic Unions

A **regional trading bloc** is an intergovernmental association that manages and promotes trade activities among countries.

Depending on the level of integration there are different types of regional trading blocs.

- In a **free trade area** (FTA), all trade barriers for the goods and services among members are eliminated, but each member maintains its own policies against non-members. Example: NAFTA
- A **customs union** is an extension of a FTA. The difference is that all members have a common trade policy against non-members.
- A **common market** further extends a customs union by allowing free movement of factors of production among members.

- An **economic union** incorporates all aspects of a common market and, in addition, requires common economic institutions and coordination of economic policies among members.
- In a **monetary union**, all members adopt a common currency.

Regional integration can be viewed as a movement toward freer trade. There are always many benefits of free trade. However, regional integration can lead to changes in the patterns of trade and impose costs on some groups. For example, two static effects can occur as results of a customs union.

- Once a union is created, members agree to eliminate tariffs between themselves. The effect is that, facing lower-priced, zero-tariff imports from members, consumers increase their demand for these goods and new trade will be created - a process called **trade creation**.
- The major loser in this situation is the previous trading partner left outside the bloc; less trade now exists between new members and their old trading partners. The process of efficient producers losing out to inefficient ones is generally referred to as **trade diversion**.

The net welfare effect can be either negative or positive.

There are other challenges in the formation of an RTA, such as cultural differences, historical considerations, national sovereignty, etc.

Capital Restrictions

Trade in assets (capital flows) provides substantial economic benefits by enabling residents of different countries to capitalize on their differences. Capital flows enable countries to borrow in order to improve their ability to produce goods and services in the future. Other benefits include the technology transfer that often accompanies foreign investment and the greater competition in domestic markets that results from permitting foreign firms to invest locally.

The benefits of capital flows do not come without a price, however. Because capital flows can complicate economic policy or even be a source of instability themselves, governments have used capital restrictions to limit their effects.

Capital restrictions are designed to limit or redirect capital account transactions. Countries use capital restrictions to maintain balance of payments, control exchange rate, preserve domestic savings for domestic use, and protect infant industry.

Capital restrictions include prohibitions on investment by foreigners, taxes on the income earned on foreign investments by domestic citizens, quantity restrictions on capital flows, and prohibition of foreign investment in certain domestic industries.

As capital restrictions are often used in conjunction with other policy instruments, their effectiveness often has mixed results.

4. The Balance of Payments

```
   When we buy something from another country, we use the currency of that country to mak
```

A country's **balance of payments accounts** records its international trading, borrowing, and lending.

- It summarizes the transactions of the country's citizens, businesses, and government with foreigners.
- Its accounts reflect all payments and liabilities to foreigners (debits) and all payments and obligations received from foreigners (credits).

Balance-of-payments accounts are recorded using the regular bookkeeping method.

- Any transaction that creates a financial inflow is recorded as a credit. That is, if a country has received money, this is known as a credit. Exports are an example of a credit item.
- Any transaction that creates a financial outflow is recorded as a debit. That is, if a country has paid or

given money, the transaction is counted as a debit. Imports are an example of a debit item.

The main categories of the balance of payments are:

- Current account

  It records payments for imports of goods and services from abroad, receipts from exports of goods and services sold abroad, net interest paid abroad, and net transfers (such as foreign aid payments).

  When combined, goods and services together make up a country's balance of trade (BOT). The BOT is typically the biggest bulk of a country's balance of payments as it makes up total imports and exports.

- Capital account

  This is where is where all international capital transfers are recorded. It also includes net sales of non-produced, non-financial assets. Capital inflow transactions are recorded as credits and capital outflow transactions are recorded as debits.

- Financial account

  This documents all international monetary flows related to investment in financial assets such as bonds and stocks. Also included are government-owned assets such as foreign reserves, gold, and special drawing rights (SDRs) held with the International Monetary Fund.

Analysts often lump financial account and capital account into one category named "capital account," which consists of portfolio investment flows (short-term) and foreign direct investment (long-term).

Example

A U.S. citizen purchases a rug from India for $100. The U.S. debits its current account for $100. Now the Indian rug-maker has two options:

- Deposit the $100 into a U.S. bank. The U.S. asset (a bank deposit) will show up as a credit to the U.S. capital account.
- Convert the $100 to rupees. The Indian bank then has 2 options.
  - Lend the $100 to a customer for the purchase of U.S. goods. This is to credit the U.S. current account.
  - Purchase U.S. government bonds. This is to credit the U.S. capital account.

In each case the balance of payments will balance.

The balance of payments must balance, meaning *the balances of these three components must sum to zero.* A deficit in one area implies an offsetting surplus in other areas. A current-account deficit implies a capital-account surplus (and vice versa).

What do these balances mean in economic terms? A country that runs a current account deficit is spending more than it produces, making up the difference between how much a country saves and how much it invests. A rising current account deficit could imply rising investment or falling saving, or both.

$$CA = S_p + S_g - I$$

To reduce a current account deficit, a country must save more and/or invest less. Higher saving can come from the private sector or from the government through a smaller budget deficit.

Net exports are exports of goods and services, X, minus imports of goods and services, M. Net exports are determined by the government budget and by private saving and investment.

- The government sector surplus or deficit is equal to net taxes, T, minus government purchases of goods

and services, G. This is S_g.
- The private sector surplus or deficit is saving, S_p, minus investment, I.
- Net exports is equal to the sum of the private sector balance and the government sector balance:

$$NX = T - G + S - I = S_p + S_g - I$$

5. Trade Organizations

```
   Created after WWII, the International Monetary Fund, the World Bank, and the World Tra
```

The IMF's mission is to ensure the stability of the international monetary system, the system of exchange rates and international payments which enables countries to buy goods and services from each other. The IMF helps keep country-specific market risk and global systemic risk under control.

The World Bank's mission is to help developing countries fight poverty and enhance environmentally sound economic growth. It helps create the basic economic infrastructure essential for the creation and maintenance of domestic financial markets and a well-functioning financial industry in developing countries.

The WTO provides the legal and institutional foundation of the multinational trading system and is the only international organization that regulates cross-border trade relations among nations on a global scale. Its mission is to foster free trade by providing a major institutional and regulatory framework of global trade rules. Without such global trade rules, today's global transactional corporations would be hard to conceive.

Currency Exchange Rates

1. The Foreign Exchange Market

```
   An <b>exchange rate</b> is the current market price at which one currency can be excha
```

Let's say a:b = S.

- a is the price currency.
- b is the base currency.
- S is the cost of one unit of currency b in terms of currency a.

For example, US$: Â£ = 1.5 indicates that Â£1 is priced at US$1.5.

The exchange rate above is referred to as the **nominal exchange rate**. The **real exchange rate** is the nominal rate adjusted somehow by inflation measures.

For example, if country A has an inflation rate of 10%, country B an inflation rate of 5%, and no changes in the nominal exchange rate took place, then country A now has a currency whose real value is higher than before.

Market Functions and Participants

A foreign exchange market is a place where foreign exchange transactions take place. Measured by average daily turnover, the foreign exchange market is by far the largest financial market in the world. It has important effects, either directly or indirectly, on the pricing and flows in all other financial markets.

There is a wide diversity of global FX market participants that have a wide variety of motives for entering into foreign exchange transactions. Commercial companies undertake FX transactions during cross-border purchases and sales of goods and services. Hedge funds trade FX currencies for hedging or even speculative purposes. Central banks use their FX reserves to stabilize the market and control the money supply. Large dealing banks provide FX price quotes to their clients. With so many different market participants, motives, and strategies, it is very difficult to describe the FX market adequately with simple characterizations.

2. Exchange Rate Quotations

```
   Most countries use a system of <b>direct quotation</b>. A direct exchange rate quote g
```

For example, the price of foreign currency is expressed in yen in Japan and pesos in Mexico. Direct quotation is used in most countries. For an American investor, a quote â‚¬:$ = 1.25 is a direct quote; he is expected to pay $1.25 for a â‚¬. Note that when there are two currencies, the base currency is always mentioned first, the opposite order of the actual ratio (price currency / base currency).

Indirect quotation (FC/DC) is also used in some markets. It is just the opposite of a direct quote; they are reciprocals of each other. For example, a bank in London will quote the value of the pound sterling (GBP) in terms of the foreign currency (i.e., Â£:$ = 1.4410).

Example

For a U.S. resident, Â¥:$ = 0.0085 is the direct quote for Japanese yen and $:Â¥ = 119.46 is the indirect quote for Japanese yen.

In a *direct quote*, an appreciation of the foreign currency (a depreciation of the domestic currency) causes an increase in the direct quote.

- The *domestic currency* moves in the *opposite* direction of the exchange rate.
- The *foreign currency* moves in the *same* direction as the exchange rate.

The opposite is true for an *indirect quote*: the *domestic* (*foreign*) currency moves in the *same* (*opposite*) direction as the exchange rate.

Bid-Ask (Offer) Quotes and Spreads

Dealers (e.g., banks) do not normally charge a commission on their currency transactions but they profit from the spread between the buying and selling rates on both spot and forward transactions. Quotes are always in pairs: the first rate is the buy, or **bid**, price (for a dealer); the second is the sell, or **ask**, **offer** (for a dealer). The ask rate is usually higher than that bid rate, so the dealer can make a profit. The average of the bid and ask price is known as the **midpoint price**: *midpoint price = (Ask + Bid) / 2.*

When direct quotations are converted to indirect quotations, bid and ask quotes are reversed. That is:

- The direct ask price is the reciprocal of the indirect bid price.
- The direct bid price is the reciprocal of the indirect ask price.
- No matter how the quote is made, dealers will always buy low and sell high.

For example, here is a direct quote for the Japanese yen from the U.S. perspective: Â¥:$ = 0.0081-83. That is, the dealer is willing to buy Â¥ at $0.0081 (direct bid price) and sell them at $0.0083 (direct ask price). The indirect bid price is (1/0.0083) $:Â¥ = 120.48 and the indirect ask price is (1/0.0081) = $:Â¥ = 123.45.

The **bid-ask spread** is the spread between bid and ask rates for a currency: Bid-ask spread = ask price - bid price. It is usually stated as a percentage of the ask price:

For example, with GBP quoted at Â£:$ = 1.4419 - 28, the percentage spread is: (1.4428 - 1.4419) x 100 / 1.4428 = 0.062%.

Note that the percentage spread is the same irrespective of whether the exchange rate is expressed in direct or indirect quotations.

The bid-ask spread is based on the breadth and depth of the market for that currency as well as on the currency's volatility.

3. Cross-Rate Calculations

```
    A <b>cross rate</b> is the exchange rate between two countries computed from each cour
```

If we interpret a:b as a "divide" sign, then a:b is actually b/a. Assume we have currencies a, b and c. If a:b and b:c are both known, then a:c = a:b x b:c. For example, if the Mexico peso (MXN) is selling for $0.0923 (MXN:USD = 0.0923) and the buying rate for the EUR is $0.7928 (USD:EUR = 0.7928), then the MXN/EUR cross rate is MXN:EUR = 0.0923 x 0.7928 = 0.0732.

Cross-Rate Calculations with Bid-Ask Spreads

Example

The rate between Japanese Â¥ and the U.S. $ is $:Â¥ = 119.05 - 121.95 and the rate between the euro and the U.S. $ is $:â‚¬ = 0.7920 - 0.7932. The direct quote between the yen and the euro in Japan will be: (Â¥119.05/$)/(â‚¬0.7932/$) = Â¥150.0883/â‚¬, and (Â¥121.95/$)/(â‚¬0.7920/$) = Â¥153.9773/â‚¬.

The lower rate is the bid, and the higher rate is the ask. Therefore, the rate between yen and euro is â‚¬:Â¥ = 150.0883 - 153.9773.

In fact, each cross-currency transaction is the combination of two trades:

- The bid price: a bank will buy U.S. dollars with yen low ($:Â¥ = 119.05), and sell U.S. dollars for euro high ($:â‚¬ = 0.7932). Thus, the bid price is â‚¬:Â¥ = 150.0883.
- The ask price: a bank will sell U.S. dollars for yen high ($:Â¥ = 121.95), and buy U.S. dollars with euro low ($:â‚¬ = 0.7920). Thus, the ask price is â‚¬:Â¥ = 153.9773.

Note that in calculating the cross rates you should always assume that you have to sell a currency at the lower (or bid) rate and buy it at the higher (or ask) rate, giving you the worst possible rate. This method of quotation is how dealers make money in foreign exchange.

Similarly, the direct quote in France or Germany is Â¥:â‚¬ = 0.006494 - 0.006663

4. Forward Calculations

```
    <b>Spot and Forward Exchange Rates</b><p> </p>
```

In the **spot market**, currencies are traded for immediate delivery. In the **forward market**, contracts are made to buy or sell currencies for future delivery.

In a typical forward transaction, a U.S. company buys textiles from England with payment of Â£1 million due in 90 days. The importer is thus **short Â£** - that is, it owes Â£ for future delivery. Suppose the present price of Â£ is $1.71. Over the next 90 days, however, Â£ might rise against the U.S. dollar, raising the U.S. dollar cost of the textiles. The importer can guard against this exchange risk by immediately negotiating a 90-day forward contract with a bank at a price, say, Â£:$ = 1.72. In 90 days the bank will give the importer Â£1 million and the importer will give the bank 1.72 million U.S. dollars. By **going long** in the forward market, the importer is able to convert a short underlying position in Â£ to a zero net exposed position.

Three points are worth noting:

- The gain or loss on the forward contract is unrelated to the current spot rate of Â£:$ = 1.71.
- The forward contract gain or loss exactly offsets the change in the U.S. dollar cost of the textile order that is associated with movements in the GBP's value.
- The forward contract is not an option contract. Both parties must perform the agreed-on behavior.

Forward exchange rates are often quoted as a premium, or discount, to the spot exchange rate. A base currency is at a **forward discount** if the forward rate is below the spot rate, whereas a **forward premium** exists if the forward rate is above the spot rate.

For example, if the one-month forward exchange rate is $:â‚¬ = 0.8020 and the spot rate is $:â‚¬ = 0.8000, the $ quotes with a premium of 0.0020 â‚¬/$. In the language of currency traders, the $ is "strong" relative to the â‚¬.

Consequently, when a trader announces that a currency quotes at a premium (discount), the premium (discount) should be added to (subtracted from) the spot exchange rate to obtain the value of the forward exchange rate.

Occasionally, forward rates are presented in terms of percentages relative to the spot rate:

Interest Rate Parity

According the **interest rate parity (IRP)** theory, the currency of the country with a lower interest rate should be at a forward premium in terms of the currency of the country with the higher rate. In an efficient market with no transaction costs, the interest differential should be (approximately) equal to the forward differential.

The exact relationship between the forward rate and the spot rate of two currencies is as follows:

- The exchange rate is d:f = S for the spot rate and F for the forward rate.
- Both i_d and i_f are periodic interest rates, which should be computed as i = annual interest rate x number of days till the forward contract expires / 360.
- It is assumed that there are no transaction costs.

Example

Suppose that the annual interest rate in the U.S. is 5%. The spot exchange rate Â£:$ = 1.50 and the 180-day forward rate is Â£:$ = 1.45. The U.S. periodic interest rate (180) is: 0.05 x 180 / 360 = 0.025. If interest rate parity holds:

$(1.45 - 1.5)/1.5 = (0.025 - i_{UK})/(1+i_{UK}) \Rightarrow i_{UK} = 6\%$

Therefore, the annual UK interest rate is approximately 12%.

Similarly, you can calculate the forward rate based on the two interest rates and the spot rate.

Interest parity ensures that the return on a hedged (or "covered") foreign investment will just equal the domestic interest rate in investments of identical risk, thereby eliminating the possibility of having a money machine. When this condition holds, the **covered interest differential** - the difference between the domestic interest rate and the hedged foreign rate - is zero.

If the difference is not zero, **covered interest arbitrage** will generate profits without any risk or investment.

For example, suppose the interest rate on GBP (Â£) is 12% in London and the interest rate on a comparable U.S. dollar investment in New York is 7%. The pound spot rate is Â£:$ = 1.75 and the one-year forward rate is Â£:$ = 1.68. These rates imply a forward discount on sterling of 4% [(1.68 - 1.75)/1.75] and a covered yield on sterling approximately equal to 8% (12% - 4%). Suppose the borrowing and lending rates are identical and the bid-ask spread in the spot and forward markets is zero. An arbitrageur will:

- Borrow $1,000,000 in New York at 7%;
- Convert the $1,000,000 to Â£571,428.57 at Â£1 = $1.75;
- Invest the Â£571,428.57 in London at 12% for one year, and sell Â£640,000 forward at a rate of Â£1 = $1.68 for delivery in one year.
- At the end of the year, collect Â£640,000 from his investment in London, deliver it to the bank's foreign exchange department in return for $1,075,200, and use $1,070,000 to repay the loan in New York. The arbitrageur will earn $5,200 on this set of transactions with no investment at all.

5. Exchange Rate Regimes

```
   The <b>exchange rate regime</b> is the way a country manages its currency in relation
```

An ideal currency regime would have three properties:

- The exchange rate between any two currencies would be credibly fixed.
- All currencies would be fully convertible.
- Each country would be able to undertake fully independent monetary policy in pursuit of domestic objectives, such as growth and inflation targets.

However, these conditions are not consistent. A country cannot have a fixed exchange rate and fully convertible currency without giving up its ability to implement independent monetary policy.

In a **flexible exchange rate regime**, the exchange rate is determined by the market forces of supply and demand, and therefore fluctuates freely in the market. The central bank intervenes in the foreign exchange market only to smooth temporary imbalances. The advantages are that the exchange rate reflects economic fundamentals at a given point in time and governments are free to adopt independent monetary and fiscal policies. However, exchange rates can be extremely volatile in this regime.

A **fixed exchange rate** is an exchange rate that is set at a determined amount by government policy. The distinguishing characteristic of a fixed rate, unified currency regime is the presence of only one central bank with the power to expand and contract the supply of money. Those linking their currency at a fixed rate to the U.S. dollar or the euro are no longer in a position to conduct monetary policy. They essentially accept the monetary policy of the nation to which their currency is tied. They also accept the exchange-rate fluctuations of that currency relative to other currencies outside of the unified zone.

In practice, most regimes fall between these extremes. The type of exchange rate regime used varies widely among countries and over time.

No Separate Legal Tender

In this regime a country does not have its own legal tender. There are two sub-types:

- **Dollarization**. The country uses another country's currency as its domestic currency. The benefit is the elimination of exchange rate fluctuations. However, this leads to the loss of monetary policy autonomy.
- **Monetary union**. In this case a group of countries share a common currency, e.g., the European Union and the euro.

Currency Board System

The monetary authority is required to maintain a fixed exchange rate with a foreign currency. Its foreign currency reserves must be sufficient to ensure that all holders of its own currency can convert them into the reserve currency. That is, the monetary authority will only issue one unit of local currency for each unit of foreign currency it has in its vault.

The major benefit is currency stability and the main drawback is the loss of ability for the country to set its own monetary policy.

Fixed Parity

The country tries to keep the value of its currency constant against another country but it has no legal obligation to do so. This is also known as the pegged exchange rate system. There can be a very small percentage allowable deviation (band) on both sides of the rate.

Target Zone

This is a fixed parity with a somewhat wider band.

Crawling Peg

In this case, the exchange rate is fixed and then adjusted periodically to keep pace with the inflation rate.

Crawling Band

This is initially a fixed parity, followed by widening band around the central parity. It is used to gradually exit from the fixed parity.

Managed Float

A country's exchange rate is adjusted based on the country's internal or external targets.

Independently Float

In this case, the market determines the exchange rate.

6. Exchange Rates, International Trade, and Capital Flows

Countries that attract a net inflow of foreign capital tend to run current account def

$$X - M = (S - I) + (T - G)$$

This relationship shows that a trade surplus is equal to the sum of public and private savings. A country saves more than enough to fund its investment (I) in plants and equipment. If a country runs a trade deficit, it has to rely on foreign capital to finance its investment (a capital surplus).

Now we analyze the impact of the exchange rate on trade and capital flows.

The Elasticities Approach

This approach emphasizes price changes as a determinant of a country's balance of payments and exchange rate.

The exchange rate is an important price in an economy. When a country's currency depreciates, domestic goods become relatively cheaper and foreign goods relatively more expensive in the global market. Hence, we would expect exports to rise and imports to decline. The elasticities approach considers the responsiveness of imports and exports to a change in the value of a country's currency.

For example, if import demand is highly elastic, a depreciation of the domestic currency will cause a disproportionate decline in the country's imports.

The Marshall-Lerner condition states that a depreciation of domestic currency can improve a country's balance of payments only when the sum of the demand elasticity of exports and the demand elasticity of imports exceeds unity.

The J-Curve is an observed phenomenon.

What is observed is that, following a depreciation or devaluation, a country's balance of payments worsens before it improves. This is because, in the short-run, exports and imports volume does not change that much, so that the price effect dominates, leading to a worsening of the current account.

Absorption Approach

This approach assumes that prices remain constant and emphasizes changes in real domestic income. Hence, the absorption approach is a real-income theory of the balance of payments.

Absorption refers to the total goods and services taken off the market domestically. In other words, absorption equals the sum of consumption plus investment.

Whether a currency depreciation can improve the current account (then the balance of payments) depends on its effect on national income and on domestic expenditure (absorption).

- On the supply side, effective depreciation requires idle resources in the economy.
- On the demand side, effective depreciation requires the Marshall-Lerner condition to be met.

Financial Statement Analysis

Financial Reporting and Analysis (1)

Introduction to Financial Statement Analysis

1. The Roles of Financial Reporting and Financial Statement Analysis

```
    The role of <b>financial reporting</b> is to provide information about a company's fir
```

The role of **financial statement analysis**, on the other hand, is to take these financial statements and other information to evaluate the company's past, current, and prospective financial position and performance for the purpose of making rational investment, credit, and similar decisions.

The primary users of financial statements are equity investors and creditors.

- *Equity investors* are primarily interested in the company's long-term earning power, growth, and ability to pay dividends.
- *Short-term creditors* (e.g., banks and trade creditors) are more interested in the company's immediate liquidity, because they seek an early payback of their investment.
- *Long-term creditors* (e.g., corporate bond owners such as insurance companies and pension funds) are primarily concerned with the company's long-term asset position and earning power.

2. Major Financial Statements

```
    Financial statements are the most important outcome of the accounting system. They con
```

The four principal financial statements are:

- Income statement (statement of earnings)
- Balance sheet (statement of financial position)
- Cash flow statement
- Statement of changes in owners' or stockholders' equity

These four financial statements, augmented by footnotes and supplementary data, are interrelated. In addition, there are other sources of financial information, such as management discussion and analysis, auditor's reports, etc.

Income Statement

The income statement summarizes revenues earned and expenses incurred, and thus measures the success of business operations*for a given period of time.* It explains some but not all of the changes in the assets, liabilities, and equity of the company between two consecutive balance sheet dates.

The income statement lists income and expenses as they are directly related to the company's recurring income. The format of the income statement is not specified by U.S. GAAP and actual format varies across companies. The following is a generic sample:

The goal of income statement analysis is to derive an effective measure of future earnings and cash flows. Analysts need data with predictive ability, hence income from continuing (recurring) operations is considered to be the best indicator of future earnings. As operating expenses do not include financing costs such as interest expenses, operating income (EBIT) is independent of the company's capital structure.

In the typical income statement this means segregating the results of normal, recurring operations from the effects of nonrecurring or extraordinary items to improve the forecasting of future earnings and cash flows. The idea here is that recurring income is persistent. If an item in the unusual or infrequent component of income from continuing operations is deemed not to be persistent, then recurring (pre-tax) income from continuing operations should be adjusted.

The net income figure is used to prepare the statement of retained earnings.

Balance Sheet

A balance sheet provides a "snapshot" of a company's financial condition. Think of the balance sheet as a photo of the business at a specific point in time. It reports major classes and amounts of assets, liabilities, stockholders' equity, and their interrelationships as of a specific date.

Assets = Liabilities + Stockholders' Equity

- Assets are the economic resources controlled by the company.
- Liabilities are the financial obligations that the company must fulfill in the future. Liabilities are typically fulfilled by payment of cash. They represent the source of financing provided to the company by the creditors.
- Equity ownership is the owner's investments and the total earnings retained from the commencement of the company. Equity represents the source of financing provided to the company by the owners.

Cash Flow Statement

The primary purpose of the cash flow statement is to provide information about a company's cash receipts and cash payments during a period. It reports the cash receipts and cash outflows classified according to operating, investment, and financing activities.

The cash flow statement is useful because it provides answers to the following simple yet important questions:

- Where did the cash come from during the period?
- What was the cash used for during the period?
- What was the change in the cash balance during the period?

The statement's value is that it helps users evaluate liquidity, solvency, and financial flexibility.

- **Liquidity** refers to the "nearness to cash" of assets and liabilities, or having enough cash available to pay debts when they are due.
- **Solvency** refers to the company's ability to pay its debts as they mature. Cash flows reflect the company's liquidity and long-term solvency.
- **Financial flexibility** refers to a company's ability to respond and adapt to financial adversity and unexpected needs and opportunities. For example, cash flow information can be used to evaluate the effects of major investment and financing decisions.

The details of income statements, balance sheets and cash flow statements will be covered in Study Session 6.

Statement of Changes in Owners' Equity

This statement reports the amounts and sources of changes in equity from capital transactions with owners. It reports ownership interests in order of preference upon liquidation and dividends. For example, the first item listed gets paid off first after creditors in the event of liquidation.

3. Other Financial Information Sources

```
<b>Financial Notes and Supplementary Schedules</b><p> </p>
```

Financial footnotes are an integral part of financial statements. They provide information about the accounting methods, assumptions and estimates used by management to develop the data reported in the financial statements. They provide additional disclosure in such areas as fixed assets, inventory methods, income taxes, pensions, debt, contingencies such as lawsuits, sales to related parties, etc. They are designed to allow users to improve assessments of the amounts, timing, and uncertainty of the estimates reported in the financial statements.

Supplementary Schedules: In some cases additional information about the assets and liabilities of a company is provided as supplementary data outside the financial statements. Examples include oil and gas reserves reported by oil and gas companies, the impact of changing prices, sales revenue, operating income, and other information for major business segments. Some of the supplementary data is unaudited.

Management Discussion and Analysis (MD&A)

This requires management to discuss specific issues on the financial statements, and to assess the company's current financial condition, liquidity, and its planned capital expenditure for the next year. An analyst should look for specific concise disclosure as well as consistency with footnote disclosure.

Note that the MD&A section is not audited and is for public companies only.

Auditor's Reports

See next subject for details.

Other Sources of Information

- **Interim reports.** Publicly held companies must file form 10-Q (interim report) on a quarterly basis. It is far less detailed than annual financial statements, as it contains unaudited basic financial statements, unaudited footnotes to financial statements, and management discussion and analysis.

- **Proxy statements.** An analyst should look for litigation, executive compensation, and related-party transactions, known as proxy statements. Proxy statements should be considered an integral part of the financial report, and they may contain special compensation "perks" for officers and directors, as well as lawsuits and other contingent obligations facing the company.

- **Companies' websites, press releases, and conference calls.**

4. Auditor's Reports

```
The auditor (an independent certified public accountant) is responsible for seeing tha
```

Though hired by the management, the auditor is supposed to be independent and to serve the stockholders and the other users of the financial statements.

An **auditor's report** (also called the **auditor's opinion**) is issued as part of a company's audited financial report. It tells the end-user the following:

- Whether the financial statements are presented in accordance with generally accepted accounting principles.
- It identifies those circumstances in which such principles have not been consistently observed in the current period in relation to the preceding period.
- Informative disclosures in the financial statements are to be regarded as reasonably adequate unless otherwise stated in the report.

An auditor's report is considered an essential tool when reporting financial information to end-users, particularly in business. Since many third-party users prefer or even require financial information to be certified by an independent external auditor, many auditees rely on auditor reports to certify their information in order to attract investors, obtain loans, and improve public appearance. Some have even stated that financial information without an auditor's report is "essentially worthless" for investing purposes.

The Types of Audit Reports

There are four common types of auditor's reports, each one representing a different situation encountered during the auditor's work. The four reports are as follows:

- An **unqualified opinion report** is issued by an auditor when the financial statements presented are free of material misstatements and are in accordance with GAAP, which, in other words, means that the company's financial condition, position, and operations are fairly presented in the financial statements. It is the best type of report an auditee may receive from an external auditor. It is regarded by many as the equivalent of a "clean bill of health" to a patient, which has led many to call it the "clean opinion."

- A **qualified opinion report** is issued when the auditor encountered one or two situations that did not comply with generally accepted accounting principles; however, the rest of the financial statements are fairly presented. This type of opinion is very similar to an unqualified or "clean opinion," but the report states that the financial statements are fairly presented with a certain exception which is otherwise misstated.

- An **adverse opinion** is issued when the auditor determines that the financial statements of an auditee are materially misstated and generally do not comply with GAAP. It is considered the opposite of an unqualified or clean opinion, essentially stating that the information contained to assess the auditee's financial position and results of operations is materially incorrect, unreliable, and inaccurate.

- A **disclaimer of opinion**, commonly referred to simply as a disclaimer, is issued when the auditor could not form, and consequently refuses to present, an opinion on the financial statements. This type of report is issued when the auditor tried to audit a company but could not complete the work due to various reasons and does not issue an opinion.

Auditor's Report on Internal Controls

Following the enactment of the Sarbanes-Oxley Act of 2002, the Public Company Accounting Oversight Board (PCAOB) was established in order to monitor, regulate, inspect, and discipline audit and public accounting firms of public companies. The PCAOB Auditing Standards No. 2 now requires auditors of public companies to include an additional disclosure in the opinion report regarding the auditee's internal controls, and to opine about the company's and auditor's assessment of the company's internal controls over financial reporting. These new requirements are commonly referred to as the COSO Opinion.

5. Financial Statement Analysis Framework

```
  The financial statement analysis framework provides steps that can be followed in any
```

- Articulate the purpose and context of the analysis.

What is the purpose of the analysis? Evaluating an equity or debt investment? Or issuing a credit rating?

The context needs to be defined clearly too: Who is the intended audience? What is the nature and content of the final report? What is the time frame? What is the budget?

- Collect input data.

Gather a company's financial data from financial statements and other sources described in Subject c (other financial information sources). Also gather information on the economy and industry to understand the environment in which the company operates.

- Process data.

Compute ratios or growth rates, prepare common-size financial statements, create charts, perform statistical analyses, make adjustments to financial statements, etc.

- Analyze / interpret the processed data.

Interpret the output to support a conclusion (e.g., a buy decision).

- Develop and communicate conclusions and recommendations.

Communicate the conclusion or recommendation in an appropriate format.

- Follow up.

Periodic review is required to determine if the original conclusions and recommendations are still valid.

Financial Reporting Standards

1. The Objective of Financial Reporting

```
An awareness of the reporting framework underlying financial reports can assist in sec
```

The objective of financial reporting:

- The objective of financial statements is to provide information about a company's financial position, performance, and any changes in financial position; this information should be useful to a wide range of end-users for the purpose of making economic decisions.
- Financial reporting requires policy choices and estimates. These choices and estimates require judgment, which can vary from one preparer to the next. Accordingly, standards are needed to attempt to ensure some type of consistency in these judgments.

2. Financial Reporting Standard-Setting Bodies and Regulatory Authorities

```
Private sector standard-setting bodies and regulatory authorities play significant but
```

International Accounting Standards Board (IASB)

This is essentially the international equivalent of the Financial Accounting Standards Board (FASB).

- It was preceded by the International Accounting Standards Committee (IASC), which was established in 1973.
- It is comprised of 14 members (12 full-time, 2 part-time); seven members are liaisons with a national board.
- It works toward harmonization of international accounting standards.
- The standard development process is open.
- Standards are principles-based.
- Since the establishment of the IASB, the focus is on global standard-setting rather than harmonization per se.

International Organization of Securities Commissions (IOSCO)

This is essentially the international equivalent of the U.S. Securities and Exchange Commission (SEC).

- It works to achieve improved market regulation internationally.
- It works to facilitate cross-border listings.
- It advocates for the development and adoption of a single set of high-quality accounting standards.

Financial Accounting Standards Board (FASB)

The FASB is a non-governmental body that sets accounting standards for all companies issuing audited financial statements. All FASB pronouncements are considered authoritative; new FASB statements immediately become part of GAAP.

U.S. Securities and Exchange Commission (SEC)

In the U.S., the form and content of the financial statements of companies whose securities are publicly traded are governed by the SEC through its regulation S-X. Although the SEC has delegated much of this responsibility to the FASB, it frequently adds its own requirements. The SEC functions as a highly effective enforcement mechanism for standards promulgated in the private sector.

Audited financial statements, related footnotes, and supplementary data are presented in both annual reports sent to stockholders and those filed with the SEC. These filings often contain other valuable information not presented in stockholder reports.

Convergence of Global Financial Reporting Standards

As capital markets become more international in scope, the need for global accounting standards and the demand for multiple listings has grown. The IASB and FASB, along with other standard-setters, are working to achieve convergence of financial reporting standards.

Pros:

- Expedite the integration of global capital markets and make the cross-listing of securities easier.
- Facilitate international mergers and acquisitions.
- Reduce investor uncertainty and the cost of capital.
- Reduce financial reporting costs.
- Allow for easy adoption of high-quality standards by developing countries.

Cons:

- Significant differences in standards currently exist.
- The political cost of eliminating differences.
- Overcoming nationalism and traditions.
- Will cause "standards overload" for some firms.
- Diverse standards for diverse places are acceptable.

3. The International Financial Reporting Standards Framework

```
   The IFRS Framework sets forth the concepts that underlie the preparation and presentat
```

Objectives of Financial Statements

The Framework identifies the central objective of financial statements as providing information about a company that is useful in making economic decisions. Financial statements prepared for this purpose will meet the needs of most end-users. Users generally want information about a company's financial performance, financial position, cash flows, and ability to adapt to changes in the economic environment in which it operates.

The Framework identifies end-users as investors and potential investors, employees, lenders, suppliers, creditors, customers, governments, and the public at large.

Qualitative Characteristics of Financial Statements

The Framework prescribes a number of qualitative characteristics of financial statements. The key characteristics are relevance and reliability. Preparers can face a dilemma in satisfying both criteria at once. For example, information about the outcome of a lawsuit may be relevant, but the financial impact cannot be measured reliably.

Financial information is **relevant** if it has the capacity to influence an end-user's economic decisions. Relevant information will help users evaluate the past, present, and most importantly, future events in a company.

To be **reliable**, financial information must represent faithfully the effects of the transactions and events that it reflects. The true impact of transactions and events can be compromised by the difficulty of measuring transactions reliably.

- Financial information faithfully represents transactions and events when accounted for in accordance with their substance and economic reality and not merely their legal form. Commonly, a legal agreement will purport that a company has "sold" assets to a third party. However, an analysis of the substance of the arrangement indicates that the company retains control over the future economic benefits and risks embodied in the asset, and should continue to recognize it on its own balance sheet.
- Financial information is reliable if it is free from material error and is complete. Information is material if its omission or misstatement could influence decisions that end-users make on the basis of the financial statements. Information is reliable when it is neutral or free from bias and prudence. A degree of prudence when preparing financial information enhances its reliability. However, a company should not use prudence as the basis for the recognition of, for example, excessive provisions.

Financial information must be easily understandable in addition to being relevant and reliable. Preparers should assume that end-users have a reasonable knowledge of business and economic activities, and an ability to comprehend complex financial matters.

End-users must be able to compare a company's financial statements through time in order to identify trends in financial performance (comparability). Hence, policies on recognition, measurement, and disclosure must be applied consistently over time. Where a company changes its accounting for the recognition or measurement of transactions, it should disclose the change in the Basis of Accounting section of its financial statements and follow the guidance set out in IFRS.

The application of qualitative characteristics and accounting standards usually results in financial statements that show a true and fair view, or fairly present a company's financial position and performance.

The Elements of Financial Statements

The Framework outlines definition and recognition criteria for assets, liabilities, equity, revenues and expenses as the elements of financial statements.

Valuation of elements is based on:

- Historical costs
  - Assets are recorded at the amount of cash or cash equivalents paid or the fair value of the consideration given to acquire them at the time of their acquisition.
  - Liabilities are recorded at the amount of proceeds received in exchange for the obligation.
- Current costs
  - Assets are carried at the amount of cash or cash equivalents that would have to be paid if the same asset was acquired currently.
  - Liabilities are carried at the undiscounted amount of cash or cash equivalents that would be required to settle the obligation currently.
- Realizable (settlement) value
  - Assets are carried at the amount of cash or cash equivalents that could currently be obtained by selling the asset in an orderly disposal.
  - Liabilities are carried at their settlement values; that is, the undiscounted amounts of cash or cash

equivalents expected to be paid to satisfy the liabilities in the normal course of business.

- Present value
    - Assets are carried at the present discounted value of the future cash inflows that the item is expected to generate in the normal course of business.
    - Liabilities are carried at the present discounted value of the future net cash outflows that are expected to be required to settle the liabilities in the normal course of business.
- Fair value

  This is the amount at which an asset could be exchanged, or a liability settled, between knowledgeable, willing parties in an arm's length transaction, which may involve either market measures or present-value measures.

Constraints and Assumptions

There are three inherent constraints.

- Timeliness. To be relevant, information must be timely; however, it takes time to get reliable information. How to balance between relevance and reliability?
- Benefit versus cost. Does the cost of providing financial information exceed the benefits derived from the information?
- Balance between qualitative characteristics. For example, should financial statements omit non-quantifiable information, such as work force creativity which can result in superior financial performance?

Underlying assumptions:

- Accrual basis. Financial statements shall be prepared on the accrual basis of accounting. No effects of payments are to be considered. All activities are reported to the financial statements of the periods to which they relate.
- Going concern principle. Financial statements are normally prepared on the assumption that a company is a going concern and will continue in operation for the foreseeable future. There is no intention or need to liquidate the company in the future. If those needs or intentions exist, the financial statements shall be prepared on another basis.

Example

A company sells goods and gets the negotiated selling price. In terms of special circumstances, the company isn't able to determine the costs for the mentioned proceeds.

Revenue and costs from the same activity have to be recognized simultaneously. Therefore, the revenue from a sale shall be recognized when the costs incurred or to be incurred with respect to the transaction can be measured reliably. Revenues won't be able to be shown in this case.

4. General Requirements for Financial Statements

```
   The objective of IAS No. 1 is to prescribe the basis for the presentation of general-p
```

Components of Financial Statements

A complete set of financial statements comprises:

- a balance sheet
- an income statement

- a statement of changes in equity showing either:
    - all changes in equity, or
    - changes in equity other than those arising from transactions with equity-holders acting in their capacity as equity-holders
- a cash flow statement
- notes, comprising a summary of significant accounting policies and other explanatory notes

Fundamental Principles Underlying the Preparation of Financial Statements

A company whose financial statements comply with IFRS shall make an explicit and unreserved statement of such compliance in the notes. Financial statements shall not be described as complying with IFRS unless they comply with all the requirements of IFRS.

Underlying principles:

- *Fair presentation.* Financial statements shall present fairly the financial position, financial performance, and cash flows of a company. In virtually all circumstances, a fair presentation is achieved by compliance with applicable IFRS.
- *Going concern.* A business is presumed to be a going concern. If management has significant concerns about the company's ability to continue as a going concern, the uncertainties must be disclosed.
- *Accrual basis.* IAS No. 1 requires that a company prepare its financial statements, except for cash flow information, using the accrual basis of accounting.
- *Consistency.* The presentation and classification of items in the financial statements shall be retained from one period to the next unless a change is justified either by a change in circumstances or requirements of new IFRS.
- *Materiality and Aggregation.* Each material class of similar items must be presented separately in the financial statements. Dissimilar items may be aggregated only if they are individually immaterial.

Presentation requirements:

- *No offsetting.* Assets and liabilities, and income and expenses, may not be offset unless required or permitted by IFRS.
- *Classified balance sheet.* A business must normally present a classified balance sheet, separating current and non-current assets and liabilities. Only if a presentation based on liquidity provides information that is reliable and more relevant may the current/non-current split be omitted.
- *Minimum information on the face of the financial statements.* IAS No. 1 specifies the minimum line item disclosures on the face of, or in the notes to, the balance sheet, the income statement, and the statement of changes in equity.
- *Minimum information in the notes.* IAS No. 1 specifies disclosures about information to be presented in the financial statements.
- *Comparative information.* Comparative information shall be disclosed in respect of the previous period for all amounts reported in the financial statements, both on the face of financial statements and in notes.

5. Comparison of IFRS with Alternative Reporting Systems

```
   A significant number of the world's listed companies report under either IFRS or U.S.
```

6. Effective Financial Reporting

```
    Effective financial reporting frameworks share three characteristics:<p> </p><ul class
```

- *Transparency.* Transparent financial statements facilitate investing in the same way that maps facilitate sailing - by clarifying the environment and therefore reducing the risk of the unknown. They permit investors to see clearly what's under the surface and what risks they might face.

- *Comprehensiveness.* Financial statements should encompass the full spectrum of transactions that have financial consequences.

- *Consistency.* Information about a particular company is more useful if an investor can compare it with similar information about other companies and with similar information about the same company for some other time period. The purpose of comparison is to detect and explain both similarities and differences. High-quality accounting requires accounting for similar transactions and circumstances similarly and accounting for different transactions and circumstances differently.

Barriers to a Single Coherent Framework

Effective standards can have conflicting approaches on valuation, the bases for standard setting, and resolution of conflicts between balance sheet and income statement focuses.

- *Valuation.* Some valuation approaches (non-historical-cost approaches) may require considerable judgment.

- *Standard-setting approach.*

    - Simply stated, **principles-based** accounting provides a conceptual basis for accountants to follow instead of a list of detailed rules. One starts with laying out the key objectives of good reporting in the subject area and then provides guidance explaining the objective and relating it to some common examples. While rules are sometimes unavoidable, the intent is not to try to provide specific guidance or rules for every possible situation. Rather, if in doubt, the reader is directed back to the principles.
    - **Rules-based** approaches are characterized by a list of specific rules, numerical tests for classifying certain transactions, exceptions, and alternative treatments.
    - IASB attempts to follow a principles-based approach to standard-setting, as such accounting standards are grounded in the IASB framework. The FASB is now adopting an objectives-oriented approach to U.S. standard-setting.

- *Measurement.* Financial reporting standards can be established taking an asset/liability approach or a revenue/expense approach.

    - The asset/liability approach requires a definition of what constitutes an asset and what constitutes a liability. For example, it links profit to changes in assets and liabilities.
    - The revenue/expense approach focuses more on the income statement. It relies on concepts such as the matching principle to determine profit.

7. Monitoring Developments in Financial Reporting Standards

```
    Reporting standards evolve rapidly. Analysts should monitor ongoing developments in fi
```

There are three areas that require special attention:

1. New products or types of transactions

There may be no explicit guidance in the financial reporting standards to report a new type of transaction or new product. For example, the creation of new financial products has outpaced the establishment of relevant accounting standards. An analyst should identify such items and gain an understanding of their business purposes. The analyst can then evaluate the potential effect of such items on financial statements (e.g., cash flow implications).

1. Evolving standards and the role of CFA Institute

Analysts can improve their investment decision-making by keeping current on financial reporting standards; various web-based sources provide the means to do so. In addition, analysts can contribute to improving financial reporting by sharing their end-users' perspectives with standard-setting bodies, which typically invite comments concerning proposed changes.

1. Company disclosures

Companies typically provide information in the footnotes to the financial statements regarding:

- Critical and significant accounting policies, such as revenue recognition and timing of expense reporting, and
- The impact of recently issued accounting changes. The conclusions include:
    - The standard does not apply.
    - The standard will have no material impact.
    - Management is still evaluating the impact.
    - The impact of adoption is discussed.

Financial Reporting and Analysis (2)

Understanding Income Statements

1. Components and Format of the Income Statement

```
  The income statement presents information on the financial results of a company's acti
```

Here are common components:

- **Sales or revenue**: amount charged for the delivery of goods or services.
    - Follows the revenue recognition rule: Revenue is recognized even though cash may not be collected until the following accounting period.
    - **Net sales** = gross sales - sales returns and allowances - discounts.
    - Amount of sales and trends in net sales over time are used to analyze a company's progress.

- **Cost of goods sold** is the amount paid for merchandise sold, or the cost to manufacture products that were sold, during an accounting period.

- **Gross margin** = net sales - costs of goods sold. Also called *gross profit*.

  Management is interested in both:

    - The amount of gross margin; and
    - The percentage of gross margin (gross margin/net sales).

  Both are useful in planning business operations.

- **Operating expenses** are expenses other than the cost of goods sold that are incurred in running a business.
    - These expenses are grouped into categories: selling expenses, general and administrative expenses, and other revenues and expenses.
    - Careful planning and control of operating expenses can improve a company's profitability.

- **Income from operations** (also called **operating income**) is the difference between gross margin and operating expenses. It represents the income from a company's normal, or main, business. It is used to compare the profitability of companies or divisions within a company.

- **Other revenues and expenses** are not part of a company's operating activities. These include:

  - Revenues or expenses from investments (e.g., dividends and interest).
  - Interest and other expenses from borrowing.
  - Any other revenue or expense not related to the company's normal business operations.

  They are also called non-operating revenues and expenses.

- **Income before income taxes** is the amount a company has earned from all activities - operating and non-operating - before taking into account the amount of income taxes the company incurred.

  This is used to compare the profitability of two or more companies or divisions within a company. Comparisons are made before income taxes are deducted because companies may be subject to different income tax rates.

- **Income taxes** (also called *provision for income taxes*) represent the expense for federal, state, and local taxes on corporate income.

  The income taxes account is shown as a separate item on the income statement. Tax rates are substantial (usually 15-38%) and have a significant effect on business decisions. Most other types of taxes are shown among operating expenses.

- **Net income** is what remains of the gross margin after operating expenses are deducted, other revenues and expenses are added or deducted, and income taxes are deducted. It is the final figure, or "bottom line," of the income statement.

  Net income = Income before income taxes - income taxes

  Net income is an important performance measure.

  - It represents the amount of business earnings that accrue to stockholders.
  - It is the amount transferred to retained earnings from all income generating activities during the year.
  - It is often used to determine whether a business has been operating successfully.

The following is a sample income statement for company XYZ for fiscal years ending 2006 and 2007 (expenses are in parentheses).

Income Statement For Company XYZ FY 2006 and 2007		
(Figures USD)	***2006***	***2007***
Net Sales	1,500,000	2,000,000
Cost of Sales	(350,000)	(375,000)
Gross Income	1,150,000	1,625,000
Operating Expenses (SG&A)	(235,000)	(260,000)
Operating Income	915,000	1,365,000
Other Income (Expense)	40,000	60,000
Extraordinary Gain (Loss)	-	(15,000)
Interest Expense	(50,000)	(50,000)
Net Profit Before Taxes (Pretax Income)	905,000	1,360,000
Taxes	(300,000)	(475,000)

Net Income	605,000	885,000

2. Revenue Recognition

There are two revenue and expense recognition issues when accrual accounting is used t

- **Timing**: when should revenue and expense be recognized?
- **Measurement**: how much revenue and expense should be recognized?

Revenue is generally recognized when it is (1) realized or realizable, and (2) earned.

The general rule for revenue recognition includes the "concept of realization." Two conditions must be met for revenue recognition to take place:

1. Completion of the earnings process

The company must have provided all or virtually all the goods or services for which it is to be paid, and it must be possible to measure the total expected cost of providing the goods or services. No remaining significant contingent obligation should exist. This condition is not met if the company has the obligation to provide future services (such as warranty protection) but cannot estimate the associated expenses.

2. Assurance of payment

The quantification of cash or assets expected to be received for the goods or services provided must be reliable.

These conditions are typically met at the time of sale, but there are many exceptions, which will be discussed next.

The Converged Revenue Recognition Standard

In May 2014, the IASB and FASB each issued a converged standard for revenue recognition. Key aspects of the converged accounting standards:

The core principle of the new standard is for companies to recognize revenue to depict the transfer of goods or services to customers in amounts that reflect the consideration (that is, payment) to which the company expects to be entitled in exchange for those goods or services.

Companies under contract to provide goods or services to a customer will be required to follow a five-step process to recognize revenue:

1. Identify contract(s) with a customer
2. Identify the separate performance obligations in the contract
3. Determine the transaction price
4. Allocate the transaction price to the separate performance obligations
5. Recognize revenue when the entity satisfies each performance obligation

There is new guidance on whether revenue should be recognized at a point in time or over time. The standard provides detailed guidance on various issues such as identifying distinct performance obligations, accounting for contract modifications, and accounting for the time value of money. Detailed implementation guidance is included on topics such as sales with a right of return, customer options for additional goods or services, etc. The standard also introduces new guidance on costs of fulfilling and obtaining a contract and specifying the circumstances in which such costs should be capitalized. Costs that do not meet the criteria must be expensed when incurred.

The standard introduces new, increased requirements for disclosure of revenue in a reporter's financial statements.

3. Expense Recognition

```
   The <b>matching principle</b> states that operating performance can be measured only i
```

Expenses incurred to generate revenues must be matched against those revenues in the time periods when the revenues are recognized.

- If the revenues are recognized in the current period, the associated expenses should be recognized in the current period and appear in the income statement.
- If revenues are expected to be recognized in future periods, the associated expenses are capitalized (appearing on the balance sheet of the current period as an asset). When the revenues are recognized in future periods, the asset is converted to expenses in those periods.

The problem of expense recognition is as complex as that of revenue recognition. For costs that are not directly related to revenues, accountants must develop a "rational and systematic" allocation policy that will approximate the matching principle. However, matching permits certain costs to be deferred and treated as assets on the balance sheet when in fact these costs may not have future benefits. If abused, this principle permits the balance sheet to become a "dumping ground" for unmatched costs.

The Matching of Inventory Costs with Revenues

Please refer to Reading 21 [Inventories] for details.

Some issues in expense recognition:

Doubtful Accounts

Account receivables arise from sales to customers who do not immediately pay cash. There are always some customers who cannot or will not pay their debts. The accounts owed by these customers are called *uncollected accounts*. Therefore, accounts receivables are valued and reported at net realizable value - the net amount expected to be received in cash, which is not necessarily the amount legally receivable. The chief problem in recording uncollectible accounts receivable is establishing the time at which to record the loss.

Under the **direct write-off method**, uncollectible accounts are charged to expense in the period that they are determined to be worthless. No entry is made until a specific account has definitely been established as uncollectible. This method is easy and convenient to apply. However, it usually does not match costs with revenues of the period, nor does it result in receivables being stated at estimated realizable value on the balance sheet.

Advocates of the **allowance method** believe that bad debt expense should be recorded in the same period as the sale to obtain a proper matching of expenses and revenues and to achieve a proper carrying value for accounts receivable. In practice, the estimate of bad debt is made either on the percentage-of-sales basis (income statement approach) or outstanding-receivables basis (balance sheet approach).

Warranties

Warranty costs are a classic example of a loss contingency. Although the future cost amount, due date, and customer are not known for certain, a liability is probable and should be recognized if it can be reasonably estimated.

Depreciation and Amortization

Please refer to Reading 22 [Long-Lived Assets] for details.

Financial Analysis Implications

In expense recognition, choice of method (i.e., the depreciation method and the inventory method) as well as estimates (i.e., uncollectible accounts, warranty expenses, assets' useful life, and salvage value) affect a company's reported income. An analyst should identify differences in companies' expense-recognition methods and adjust reported financial statements where possible to facilitate comparability. Where the available information does not permit adjustment, an analyst can characterize the policies and estimates as more or less conservative and thus qualitatively assess how differences in policies might affect financial ratios and judgments about companies' performances.

4. Non-Recurring Items and Non-Operating Items

```
   The goal of analyzing an income statement is to derive an effective indicator to predi
```

Segregating the results of recurring operations from those of non-recurring items facilitates the forecasting of future earnings and cash flows. Generally, analysts should exclude items that are non-recurring in nature when predicting a company's future earnings and cash flows. However, this does not mean that every non-recurring item in the income statement should be ignored. Management tends to label many items in the income statement as "non-recurring," especially those that reduce reported income. For the purpose of analysis, an important issue is to assess whether non-recurring items are really "non-recurring," regardless of their accounting labels.

There are four types of non-recurring items in an income statement.

1. Discontinued operations

Discontinued operations are not a component of persistent or recurring net income from continuing operations. To qualify, the assets, results of operations, and investing and financing activities of a business segment must be separable from those of the company. The separation must be possible physically and operationally, and for financial reporting purposes. Any gains or disposal will not contribute to future income and cash flows, and therefore can be reported only after disposal, that is - when realized.

- Subsidiaries and investees also qualify as separate components.
- Disposal of a portion of a business component does not qualify as discontinued operations. Instead, this is recorded as an unusual or infrequent item.

2. Extraordinary items

Extraordinary items are BOTH unusual in nature AND infrequent in occurrence, and material in amount. They must be reported separately (below the line) net of income tax.

Common examples are:

- Expropriations by foreign governments.
- Uninsured losses from earthquakes, eruptions, and tornadoes.

But ...

Starting 2016, businesses and other organizations should report transactions that are either unusual or rare on the income statement or disclose them in their financial statement footnotes. **There is no such concept of "extraordinary items".**

The International Financial Reporting Standards (IFRS) do not include extraordinary items in their accounting practices.

3. Unusual or infrequent items

These are either unusual in nature OR infrequent in occurrence but not both. They may be disclosed separately (as a single-line item) as a component of income from continuing operations. They are reported pre-tax in the income statement and appear "above the line," while the other three categories are reported on an after-tax basis and "below the line" and excluded from net income from continuing operations.

Common examples are:

- Gains or losses from disposal of a portion of a business segment.
- Gains or losses from sales of assets or investments in affiliates or subsidiaries.
- Provisions for environmental remediation.
- Impairment, write-offs, write-downs, and restructuring costs (such as those costs related to the integration of acquired companies).

4. Changes in accounting principles

Changes in accounting principles, such as from LIFO to another inventory method or from the percentage-of-completion method to the completed-contract method, can be either voluntary changes or changes mandated by new accounting standards. They are reported in the same manner as extraordinary items and discontinued operations. The cumulative impact on prior period earnings should be reported as a separate line item on the income statement on an after-tax basis. They are typically reported through retrospective application, which means that the financial statements for all fiscal years shown in a company's financial report are presented as if the newly adopted accounting principle had been used throughout the entire period.

Changes in accounting estimates, such as changes in asset lives or salvage value when recording depreciation expenses, are not considered changes in accounting principles. The impact of such a change is only prospective, and no retroactive or cumulative effects are recognized.

A change from an incorrect to an acceptable accounting method is treated as an error and its impact is reported as a prior period adjustment.

Non-operating Items: Investing and Financing Activities

Non-operating items are reported separately from operating items. For example, if a non-financial service company invests in equity or debt securities issued by another company, any interest, dividends, or profits from sales of these securities will be shown as non-operating income.

Summary

Non-recurring items should be scrutinized to assess whether they are truly "non-recurring." For example, gains or losses from the sale of fixed assets are classified as unusual or infrequent items. However, for a car rental company that retires part of its fleet of cars annually, such gains or losses are rather recurring in nature. Some non-recurring charges are, in fact, prior period expenses taken too late or future expenses taken too early. For example, asset write-downs may indicate that prior period depreciation or amortization changes were insufficient. Therefore, completely ignoring such non-recurring items in financial analysis may result in an overestimation of a company's earning trend.

5. Earnings per Share

```
    <b>Earnings per share</b> (<b>EPS</b>) is a measure that is widely used to evaluate th
```

A company's capital structure is **simple** if it consists of only common stock or includes no potential common stock that upon conversion or exercise could dilute earnings per common share. Companies with simple capital structures only need to report basic EPS.

A **complex capital structure** contains securities that could have a dilutive effect on earnings per common share. **Dilutive securities** are securities that, upon conversion or exercise, could dilute earnings per share. These securities include options, warrants, convertible bonds, and preferred stocks.

Companies with complex capital structures must report both basic EPS and diluted EPS. Calculation of diluted EPS under a complex capital structure allows investors to see the adverse impact on EPS if all diluted securities are converted into common stock.

Basic EPS

To calculate EPS in a simple capital structure:

The current year's preferred dividends are subtracted from net income because EPS refers to earnings available to the common shareholder. Common stock dividends are not subtracted from net income.

Since the number of common shares outstanding may change over the year, the weighted average is used to compute EPS. The weighted average number of common shares is the number of shares outstanding during the year weighted by the portion of the year they were outstanding. Analysts need to find the equivalent number of whole shares outstanding for the year.

Three steps are used to compute the weighted average number of common shares outstanding:

- Identify the beginning balance of common shares and changes in the common shares during the year.
- For each change in the common shares:
    - Compute the number of shares outstanding after each change in the common shares. Issuance of new shares increases the number of shares outstanding. Repurchase of shares reduces the number of shares outstanding.
    - Weight the shares outstanding by the portion of the year between this change and next change: weight = days outstanding / 365 = months outstanding / 12
- Sum up to compute the weighted average number of common shares outstanding.

Stock Dividends and Splits

In computing the weighted average number of shares, stock dividends and stock splits are only changes in the units of measurement, not changes in the ownership of earnings. A stock dividend or split does not change the shareholders' total investment (i.e., it means more pieces of paper for shareholders).

When a stock dividend or split occurs, computation of the weighted average number of shares requires restatement of the shares outstanding before the stock dividend or split. It is not weighted by the portion of the year after the stock dividend or split occurred.

Specifically, before starting the three steps of computing the weighted average, the following numbers should be restated to reflect the effects of the stock dividend/split:

- The beginning balance of shares outstanding;
- All share issuance or purchase prior to the stock dividend or split.
- No restatement should be made for shares issued or purchased after the date of the stock dividend or split.

If a stock dividend or split occurs after the end of the year but before the financial statements are issued, the weighted average number of shares outstanding for the year (and any other years presented in comparative form) must be restated.

Example

1. 01/01/15 - 100,000 shares issued and outstanding at the beginning of the year
2. 07/01/15 - 10% stock dividend
3. 09/01/15 - 3 for 1 stock split
4. 10/01/15 - 50,000 shared issued

The weighted average number of shares is:

100,000 x 1.1 x 3 x 9/12 + (100,000 x 1.1 x 3 + 50,000) x 3 / 12 = 342,500

Diluted EPS

If a company has a complex capital structure, it must report two EPS figures: basic EPS and diluted EPS.

The securities could be either dilutive or antidilutive. **Antidilutive securities** are those that, upon conversion or exercise, increase earnings per share or reduce loss per share. The likelihood of conversion or exercise of antidilutive securities is considered remote.

Diluted EPS shows the maximum potential adverse effect on EPS if dilutive securities are converted to common stock. The purpose is to show the "worst case" scenario. Therefore, the computation of diluted EPS does not consider antidilutive securities, which increase EPS. In computing diluted EPS, analysts need to check each potentially dilutive security individually to see whether it is dilutive or antidilutive. All antidilutive securities should be excluded and cannot be used to offset dilutive securities.

Only income from continuing operations (excluding discontinued operations, extraordinary items, and accounting changes) is considered in determining diluted EPS.

To compute diluted EPS, start from basic EPS and remove the adverse effect of all dilutive securities outstanding during the period. In computing diluted EPS, the adverse effects of dilutive securities are removed by adjusting the numerator and the denominator of the basic EPS formula.

- Identify all potentially dilutive securities: convertible bonds, options, convertible preferred stock, warrants, etc.
- Compute the basic EPS. The effect of potentially dilutive securities is not included in the computation.
- Determine the effect of each potentially dilutive security on EPS to see whether it is dilutive or antidilutive. How? Compute the adjusted EPS assuming the conversion occurs. If adjusted EPS < (>) basic EPS, the security is dilutive (antidilutive).
- Exclude all antidilutive securities from the computation of diluted EPS.
- Use basic and dilutive securities to compute diluted EPS.

Diluted EPS =

If-Converted Method - Convertible Securities

If a company has a complex capital structure containing convertible bonds and preferred stocks, then diluted EPS will treat these securities as if they were converted to common stocks from the first of the year (or when issued, if issued during the current year).

The effect of a convertible bond on EPS:

- Upon conversion, the numerator (net income) of the basic EPS formula will be increased by the amount of interest expense, net of tax associated with those potential common shares. Why? If converted, there would be no interest for the bond, so income available to common shares will increase accordingly. After-tax interest is used because bond interest is tax deductible while net income is computed on an after-tax basis.

- Upon conversion, the denominator (weighted average number of shares outstanding) of the basic EPS formula will be increased by the number of shares created from the conversion, weighted by the time that these shares would be outstanding: number of shares due to conversion = par value of the convertible bond / conversion price

  The time outstanding would be the entire year if the bond was issued in a previous year, or a fraction of the year if the bond is issued in the current year.

The effect of a convertible preferred stock on EPS:

- Upon conversion, the numerator of the basic EPS formula would increase by the amount of the preferred dividends. If converted, there would be no dividends for the convertible preferred stock so income available to common shares would increase accordingly. Unlike bond interests, preferred dividends are not tax-deductible.

- Upon conversion, the denominator of the basic EPS formula would increase by the number of shares created from the conversion weighted by the time that these shares would be outstanding: number of shares due to conversion = number of convertible preferred shares outstanding x conversion rate.

  The time outstanding would be the entire year, if the preferred stock was issued in a previous year, or a fraction of the year, if the preferred stock is issued in the current year.

Treasury Stock Method - Options and Warrants

This method assumes that options and warrants are exercised at the beginning of the year (or date of issue if later) and the proceeds from the exercise of options and warrants are used to purchase common stock for the treasury. There is no adjustment to net income in the numerator.

- Upon exercise of the options or warrants, the company receives the following amount of proceeds: exercise price of the option x number of shares issued to holders of the options or warrants.

- The company will then use the proceeds from the exercise of options and warrants to buy back common shares at the average market price for the year.

- The net change in the number of shares outstanding is the number of shares issued to holders of the options or warrants less the number of shares acquired from the market.

If the exercise price of the option or warrants is lower than the market price of the stock, dilution occurs. If it is higher, the number of common shares is reduced and an antidilutive effect occurs. In the latter case, exercise is not assumed.

Like the if-converted method, the treasury stock method makes the following assumptions:

- If an option or warrant was issued in a previous year, it is assumed to be exercised at the beginning of the current year. Thus, the net change in common shares is outstanding for the entire year.
- If the option or warrant is issued during the current year, its exercise occurs at the date of issuance. Thus, the net change in common shares is outstanding for a fraction of the current year, starting from the date of issuance.

Note: If there are restrictions on the proceeds received when warrants are exercised, dilutive EPS calculations must reflect the results of those agreements.

6. Analysis of the Income Statement

```
<b>Common-Size Analysis of the Income Statement</b><p> </p>
```

This topic will be discussed in detail in Reading 20 [Financial Analysis Techniques].

Income Statement Ratios

The following operating profitability ratios measure the rates of profit on sales (profit margins).

- **Net Profit Margin** shows how much profit is generated on every dollar of sales.

  Net income is earnings after tax but before dividends (EBIT - interest - taxes). It should be based on earnings from the company's continuing operation because the analysis is to forecast the company's future performance. Thus analysts should not consider earnings from discontinued operations, gains or losses from the sale of discontinued operations, and non-recurring income or expenses.

- **Gross Profit Margin** equals percent of sales available after deducting cost of goods sold.

This percentage is available to cover selling, general and administrative costs, and also earn a profit. It indicates the basic cost structure of a company and shows the company's cost-price position. Comparing this ratio with the industry average over time shows the company's relative profitability within the industry.

- A declining gross profit may indicate increasing costs of production or declining prices.
- The ratio can be affected by changes in the company's product mix: a change toward items with higher (lower) margin raises (reduces) the gross profit margin.
- A small change in gross profit can result in a much larger change in profit margin if the company has high fixed costs.

7. Comprehensive Income

```
   <b>Comprehensive income</b> includes both net income and other revenue and expense ite
```

Comprehensive income is the sum of net income and other items that must bypass the income statement because they have not been realized, including items like an unrealized holding gain or loss from available-for-sale securities and foreign currency translation gains or losses. These items are not part of net income, yet are important enough to be included in comprehensive income, giving the user a bigger, more comprehensive picture of the organization as a whole.

The following table is from the Statement of Stockholders' Equity section of the 3M's 2001 annual report.

This section describes the composition of comprehensive income. It begins with net income and then includes those items affecting stockholders' equity that do not flow through the income statement. For 3M, these items include:

- Cumulative translation adjustment.
- Minimum pension liability adjustment.
- Unrealized gains (losses) on available-for-sale investments.
- Unrealized gains (losses) on derivative investments.

FASB has taken the position that income for a period should be all-inclusive comprehensive income. Comprehensive income may be reported on an income statement or separate statement, but is usually reported on a statement of stockholders' equity.

Understanding Balance Sheets

1. Components and Format of the Balance Sheet

```
   The starting place for analyzing a company is typically the balance sheet. Think of th
```

- Assets are the economic resources controlled by the company.
- Liabilities are the financial obligations that the company must fulfill in the future. Liabilities are typically fulfilled by payment of cash. They represent the source of financing provided to the company by the creditors.
- Equity ownership is the owner's investments and the total earnings retained from the commencement of the company. Equity represents the source of financing provided to the company by the owners.

The balance sheet provides users, such as creditors and investors, with information regarding the sources of finance available for projects and infrastructure. At the same time, it normally provides information about the future earnings capacity of a company's assets as well as an indication of cash flow implicit in the receivables and inventories.

The balance sheet has many limitations, especially relating to the measurement of assets and liabilities. The lack of timely recognition of liabilities and, sometimes, assets, coupled with historical costs as opposed to fair value accounting for all items on the balance sheet, implies that the financial analyst must make numerous adjustments to determine the economic net worth of the company.

The analyst must understand the components, structure, and format of the balance sheet in order to evaluate the liquidity, solvency, and overall financial position of a company.

Balance Sheet Format

Balance sheet accounts are *classified* so that similar items are grouped together to arrive at significant subtotals. Furthermore, the material is arranged so that important relationships are shown.

The table below indicates the general format of balance sheet presentation:

This format is referred to as the **account format**, which follows the pattern of the traditional general ledger accounts, with assets at the left and liabilities and equity at the right of a central dividing line. A **report format** balance sheet lists assets, liabilities, and equity in a single column.

Balance Sheet Components

Current Assets

These are cash and other assets expected to be converted into cash, sold, or consumed either in one year or in the operating cycle, whichever is longer. The **operating cycle** is the average time between the acquisition of materials and supplies and the realization of cash through sales of the product for which the materials and supplies were acquired. The cycle operates from cash through inventory, production, and receivables back to cash. Where there are several operating cycles within one year, the one-year period is used. If the operating cycle is more than one year, the longer period is used.

Long-Term Investments

Often referred to simply as investments, these are to be held for many years and are not acquired with the intention of disposing of them in the near future.

- Investments in securities such as bonds, common stock, or long-term notes that management does not intend to sell within one year.
- Investments in tangible fixed assets not currently used in operations, such as land held for speculation.
- Investments set aside in special funds, such as a sinking fund, pension fund, or plant expansion fund. The cash surrender value of life insurance is included here.
- Investments in non-consolidated subsidiaries or affiliated companies.

Property, Plant, and Equipment

These are properties of a durable nature used in the regular operations of the business. With the exception of land, most assets are either depreciable (such as a building) or consumable.

Intangible Assets

These lack physical substance and usually have a high degree of uncertainty concerning their future benefits. They include patents, copyrights, franchises, goodwill, trademarks, trade names, secret processes, and organization costs.

Other Assets

These vary widely in practice. Examples include deferred charges (long-term pre-paid expenses), non-current receivables, intangible assets, assets in special funds, and advances to subsidiaries.

Current Liabilities

These are obligations that are reasonably expected to be liquidated either through the use of current assets or the creation of other current liabilities within one year or within the operating cycle, whichever is longer.

The excess of total current assets over total current liabilities is referred to as **working capital**. It represents the net amount of a company's relatively liquid resources; that is, it is the liquid buffer, or margin of safety, available to meet the financial demands of the operating cycle.

Long-Term Liabilities

These are obligations that are not reasonably expected to be liquidated within the normal operating cycle but instead at some date beyond that time. Bonds payable, notes payable, deferred income taxes, lease obligations, and pension obligations are the most common long-term liabilities. Generally they are of three types:

- Obligations arising from specific financing situations, such as issuance of bonds, long-term lease obligations, and long-term notes payable.
- Obligations arising from the ordinary operations of the enterprise, such as pension obligations and deferred income tax liabilities.
- Obligations that are dependent upon the occurrence or non-occurrence of one or more future events to confirm the amount payable, or the payee, or the date payable, such as services or product warranties and other contingencies.

Owner's Equity

The complexity of capital stock agreements and the various restrictions on residual equity imposed by state corporation laws, liability agreements, and boards of directors make the owner's equity section one of the most difficult sections to prepare and understand. This section is usually divided into three parts:

- Capital stock: the par or stated value of the shares issued.
- Additional paid-in capital: the excess of amounts paid in over the par or stated value.
- Retained earnings: the corporation's undistributed earnings.

2. Measurement Bases of Assets and Liabilities

```
  Asset and liability values reported on a balance sheet may be measured on the basis of
```

Current Assets

Current assets are presented in the balance sheet in order of liquidity. The five major items found in the current assets section are:

- Cash. Valued at its stated value. Cash restricted for purpose other than payment of current obligations or for use in current operations should be excluded from the current asset section.
- Marketable securities. Valued at cost or lower of cost and market value.
- Accounts receivables. Amounts owed to the company by its customers for goods and services delivered. Valued at the estimated amount collectible.
- Inventories. Products that will be sold in the normal course of business. They should be measured at the lower of cost or net realizable value. Refer to Reading 21 [Inventories] for details.
- Pre-paid expenses. These are expenditures already made for benefits (usually services) to be received within one year or the operating cycle, whichever is longer. Typical examples are pre-paid rent, advertising, taxes, insurance policies, and office or operating supplies. They are reported at the amount of un-expired or unconsumed cost.

Current Liabilities

Current liabilities are typically paid from current assets or by incurring new short-term liabilities. They are not reported in any consistent order. A typical order is: accounts payable, notes payable, accrued items (e.g., accrued warranty costs, compensation and benefits), income taxes payable, current maturities of long-term debt, unearned revenue, etc.

Tangible Assets

These are carried at their historical cost less any accumulated depreciation or accumulated depletion. See Reading 22 [Long-Lived Assets] for details.

Intangible Assets

Intangible assets are long-term assets that have no physical substance but have a value based on rights or privileges that belong to their owner. Generally, **identifiable intangible assets** are recorded only when purchased (at acquisition costs). The cost of internally developed identifiable intangible assets is typically expensed when incurred. For example, R&D costs are not in themselves intangible assets. They should be treated as revenue expenditures and charged to expense in the period in which they are incurred. One exception is that IFRS allows costs in the development stage to be capitalized if certain criteria (including technological feasibility) are met.

A company should assess whether the useful life of an intangible asset is finite or infinite and, if finite, the length of its life. The straight-line method is typically used for amortization.

Goodwill is an example of an **unidentifiable intangible asset** which cannot be acquired singly and typically possesses an indefinite benefit period. It stems from such factors as a good reputation, loyal customers, and superior management. Any business that earns significantly more than a normal rate of return actually has goodwill.

Goodwill is recorded in the accounts only if it is purchased by acquiring another business at a price higher than the fair market value of its net identifiable assets. It is not valued directly but inferred from the values of the acquired assets compared with the purchase price. It is the premium paid for the target company's reputation, brand names, customers or suppliers, technical knowledge, key personnel, and so forth.

Goodwill only has value insofar as it represents a sustainable competitive advantage that will result in abnormally high earnings. Analysts need to be aware of the possibility, however, that the goodwill recognized by accountants may, in fact, represent overpayment for the acquired company. Since goodwill is inferred rather than computed directly, it will increase as the payment price increases. It is only after the passage of time that analysts will be able to evaluate the extent to which the purchase price was justified.

Under U.S. GAAP SFAS No. 142, goodwill is no longer amortized, but is tested annually for impairment. It is not amortized. Impairment of goodwill is a non-cash expense which is charged against income in the current period.

3. Financial Instruments: Financial Assets and Financial Liabilities

```
   Financial instruments are contracts that give rise to both a financial asset of one cc
```

Measured at fair market value:

Financial assets:

- Financial assets held for trading.
- Available-for-sale financial assets.
- Derivatives (whether stand-alone or embedded in non-derivative instruments).
- Non-derivative instruments with fair value exposures hedged by derivatives.

Financial liabilities:

- Derivatives.
- Financial liabilities held for trading.
- Non-derivative instruments with fair value exposures hedged by derivatives.

Measured at cost or amortized cost:

Financial assets:

- Unlisted instruments (there is no reliable valuation measure).
- Held-to-maturity investments (bonds).
- Loans and receivables.

Financial liabilities:

- All other liabilities (such as bonds payable or notes payable).

Accounting for Gains and Losses on Marketable Securities

- **Held-to-maturity securities.** Debt securities that management intends to hold to their maturity dates. At year-end, they are reported at cost adjusted for the effect of interest (debit the securities account and credit the interest income account) and unrealized holding gains and losses are not recognized.

- **Trading securities.** Debt and equity securities bought and held mainly for sale in the short term to generate income on price changes. At year-end, they are reported at their fair market value. Any unrealized holding gains or losses are recognized on the company's income statement as part of net income. When they are sold, the realized gains or losses will also appear on the income statement. Realized gains and losses are not affected by any unrealized gains or losses recognized before.

- **Available-for-sale securities.** Debt and equity securities not classified as held-to-maturity or trading securities. Unrealized gains and losses are reported as part of other comprehensive income (in contrast, the unrealized gains or losses of trading securities are reported in the income statement as part of net income). Other than that, they are accounted for in the same way as trading securities.

4. Equity

```
 Equity is a residual value of assets which the owner has claim to after satisfying oth
```

- *Contributed capital.* The amount of money which has been invested in the business by the owners. This includes preferred stocks and common stocks. Common stock is recorded at par value with the remaining amount invested contained in additional paid-in capital.

- *Minority interest.*

- *Retained earnings.* These are the total earnings of the company since its inception less all dividends paid out.

- *Treasury stock.* This is a company's own stock that has

    - Already been fully issued and was outstanding;
    - Been reacquired by the company; and
    - Not been retired.

It decreases stockholder's equity and total shares outstanding.

- *Accumulated comprehensive income.* This includes items such as the minimum liability recognized for under-funded pension plans, market value changes in non-current investments, and the cumulative effect of foreign exchange rate changes. Refer to Reading 17 [Understanding the Income Statement] for details.

Statement of Changes in Shareholders' Equity

This statement reflects information about increases or decreases to a company's net assets or wealth. It reveals much more about the year's stockholders' equity transactions than the statement of retained earnings.

- The statement of shareholders' equity is a financial statement that summarizes changes that occurred during the accounting period in components of the stockholders' equity section of the balance sheet. For

example, it includes capital transactions with owners (e.g., issuing shares) and distributions to owners (i.e., dividends).
- The shareholders' equity section of the balance sheet lists the items in contributed capital and retained earnings on the balance sheet date.

5. Uses and Analysis of the Balance Sheet

```
<b>Common-Size Analysis of Balance Sheets</b><p> </p>
```

This topic will be discussed in detail in Reading 20 [Financial Analysis Techniques].

Balance Sheet Ratios

Liquidity ratios measure the ability of a company to meet future short-term financial obligations from current assets and, more importantly, cash flows. Each of the following ratios takes a slightly different view of cash or near-cash items.

- **Current Ratio** is a measure of the number of dollars of current assets available to meet current obligations. It is the best-known liquidity measure. A current ratio of less than 1 indicates the company has negative working capital.

- **Quick Ratio** (**Acid-Test Ratio**) eliminates less liquid assets, such as inventory and pre-paid expenses, from the current ratio. If inventory is not moving, the quick ratio is a better indicator of cash and near-cash items that will be available to meet current obligations.

- **Cash Ratio** is the most conservative liquidity ratio, determined by eliminating receivables from the quick ratio. As with the elimination of inventory in the quick ratio, there is no guarantee that the receivables will be collected.

Solvency ratios measure a company's ability to meet long-term and other obligations.

- **Long-Term Debt-Equity Ratio** is an indicator of the degree of protection available to the creditors in the event of insolvency of a company. Higher debt-equity ratio indicates higher financial risk.

- **Debt-Equity Ratio** includes short-term debt in the numerator.

  The total debt includes all liabilities, including non-interest-bearing debt such as accounts payables, accrued expenses, and deferred taxes. This ratio is especially useful in analyzing a company with substantial financing from short-term borrowing.

- **Total Debt Ratio** =

- **Financial Leverage Ratio** =

Financial statement analysis aims to investigate a company's financial condition and operating performance. Using financial ratios helps to examine relationships among individual data items from financial statements. Although ratios by themselves cannot answer questions, they can help analysts ask the right questions in financial statement analysis. As analytical tools, ratios are attractive because they are simple and convenient. However, ratios are only as good as the data upon which they are based and the information with which they are compared.

From the earlier discussion it is obvious that there are a significant number of estimates and subjective information that go into financial statements and therefore it is imperative that the end user understands the numbers before calculating and relying on ratio analyses based on these numbers.

Understanding Cash Flow Statements

1. Classification of Cash Flows and Non-Cash Activities

```
   The cash flow statement provides important information about a company's cash receipts
```

Cash receipts and cash payments during a period are classified in the statement of cash flows into three different activities:

Operating Activities

These involve the cash effects of transactions that enter into the determination of net income and changes in the working capital accounts (accounts receivable, inventory, and accounts payable). Cash flows from operating activities (CFOs) reflect the company's ability to generate sufficient cash from its continuing operations. CFOs are derived by converting the income statement from an accrual basis to a cash basis. For most companies, positive operating cash flows are essential for long-run survival.

The major operating cash flows are (1) cash received from customers, (2) cash paid to suppliers and employees, (3) interest and dividends received, (4) interest paid, and (5) income taxes paid.

Special items to note:

- Interest and dividend revenue, and interest expenses, are considered operating activities, but dividends paid are considered financing activities. Note that interest expense is reported on the income statement while dividends flow through the retained earnings statement.

  Remember that an interest/dividend item is an operating activity if it appears on the income statement. For example, payments of dividends do not appear on the income statement, and thus are not classified as operating activities.

- All income taxes are considered operating activities, even if some arise from financing or investing.

- Indirect borrowing using accounts payable is not considered a financing activity - such borrowing would be classified as an operating activity.

Investing Activities

These include making and collecting loans and acquiring and disposing of investments (both debt and equity) and property, plants, and equipment. In general, these items relate to the long-term asset items on the balance sheet. Investing cash flows reflect how a company plans its expansions.

Examples are:

- Sale or purchase of property, plant and equipment.
- Investments in joint ventures and affiliates and long-term investments in securities.
- Loans to other entities or collection of loans from other entities.

Financing Activities

These involve liability and owner's equity items, and include:

- Obtaining capital from owners and providing them with a return on (and a return of) their investments.
- Borrowing money from creditors and repaying the amounts borrowed.

In general, the items in this section relate to the debt and the equity items on the balance sheet. Financing cash flows reflect how the company plans to finance its expansion and reward its owners.

Examples:

- Dividends paid to stockholders (not interest paid to creditors!). Note that the cash outflow caused by dividends is determined by dividends paid, not dividends declared. Dividends paid are not reflected in the retained earnings account. The amount is provided in the supplementary information.
- Issue or repurchase of the company's stocks.
- Issue or retirement of long-term debt (including the current portion of long-term debt).

Purchase of debt and equity securities from other entities (sale of debt or equity securities of other entities) and loans to other entities (collection of loans to other entities) are considered investing activities. However, issuance of debt (bonds and notes) and equity securities is a financing cash inflow, and payment of dividend, redemption of debt, and reacquisition of capital stock are financing cash outflows.

Non-cash Activities

Some investing and financing activities do not flow through the statement of cash flows because they don't require the use of cash:

- Retiring debt securities by issuing equity securities to the lender.
- Converting preferred stock to common stock.
- Acquiring assets through a capital lease.
- Obtaining long-term assets by issuing notes payable to the seller.
- Exchanging one non-cash asset for another non-cash asset.
- The purchase of non-cash assets by issuing equity or debt securities.

For example, if a company purchases $200,000 of land by issuing a long-term bond, this transaction is a non-cash one, as it does not involve direct outlays of cash. Therefore, it is excluded from the statement of cash flows. These types of transactions should be disclosed in a separate schedule as part of the statement of cash flows or in the footnotes to the financial statements.

Differences between IFRS and U.S. GAAP

The above discussions are based on the U.S. GAAP. Under IFRS there is some flexibility in reporting some items of cash flow, particularly interest and dividends.

- Interest and dividends received:
    - Under U.S. GAAP, interest income and dividends received from investment in other companies are classified as CFO.
    - Under IFRS, interest and dividends received may be classified as either CFO or CFI.
- Interest paid:
    - Under U.S. GAAP, interest paid is classified as CFO.
    - Under IFRS, interest paid may be classified as either CFO or CFF.
- Dividends paid:
    - Under U.S. GAAP, dividends paid are classified as CFF.
    - Under IFRS, dividends paid may be classified as either CFO or CFF.

2. Preparing the Cash Flow Statement

```
   The beginning and ending cash balances on the statement of cash flows tie directly to
```

Net income differs from net operating cash flows for several reasons.

- One reason is non-cash expenses, such as depreciation and the amortization of intangible assets. These

expenses, which require no cash outlays, reduce net income but do not affect net cash flows.
- Another reason is the many timing differences existing between the recognition of revenue and expense and the occurrence of the underlying cash flows.
- Finally, non-operating gains and losses enter into the determination of net income, but the related cash flows are classified as investing or financing activities, not operating activities.

There are two methods of converting the income statement from an accrual basis to a cash basis. Companies can use either the direct or the indirect method for reporting their operating cash flow.

- The **direct method** discloses operating cash inflows by source (e.g., cash received from customers, cash received from investment income) and operating cash outflows by use (e.g., cash paid to suppliers, cash paid for interest) in the operating activities section of the cash flow statement.

  - It adjusts each item in the income statement to its cash equivalent.
  - It shows operating cash receipts and payments. More cash flow information can be obtained and it is more easily understood by the average reader.

- The **indirect method** reconciles net income to net cash flow from operating activities by adjusting net income for all non-cash items and the net changes in the operating working capital accounts.

  - It shows why net income and operating cash flows differ.
  - It is used by most companies.

- The direct and indirect methods are *alternative formats* for reporting net cash flows from operating activities. Both methods produce the same net figure (*dollar amount* of operating cash flow).

- Under IFRS and U.S. GAAP, both the direct and indirect methods are acceptable for financial reporting purposes. However, the direct method discloses more information about a company. Partly because companies want to limit information disclosed, the indirect method is more commonly used.

- The reporting of investing and financing activities is the same for both direct and indirect methods. Only the reporting of CFO is different.

Direct Method

Under the direct method, the statement of cash flows reports net cash flows from operations as major classes of operating cash receipts and cash disbursements. This method converts each item on the income statement to its cash equivalent. The net cash flows from operations are determined by the difference between cash receipts and cash disbursements.

Assume that Bismark Company has the following balance sheet and income statement information:

Additional information:

- Receivables relate to sales and accounts payable relates to cost of goods sold.
- Depreciation of $5,000 and pre-paid expense both relate to selling and administrative expenses.

Direct Method:

- Cash sales: sales on the accrual basis are $242,000. Since the receivables have decreased by $8,000, the cash collections are higher than accrual-basis sales.

  Sales: $242,000
  Add decrease in receivables: $8,000
  Cash sales: $250,000

- Cash purchases: since inventory decreased by $2,000, goods purchased in prior years were used as the cost of goods sold. Since accounts payable decreased by $12,000, more cash was paid in 2000 for goods than is reported under accrual accounting.

  Costs of goods sold: $105,000
  Deduct decrease in inventories: $2,000
  Add decrease in accounts payable: $12,000
  Cash purchases: $115,000

- Cash selling and administrative expenses: the selling and administrative expenses include a non-cash charge of $5,000 related to depreciation. In addition, pre-paid expenses (assets) increased by $1,000 and should be added to the selling and administrative expenses.

  Selling and administrative expenses: $58,000
  Deduct depreciation expense: $5,000
  Add increase in pre-paid expense: $1,000
  Cash selling and administrative expenses: $54,000

- Cash income taxes: income tax on the accrual basis is $30,000. Tax payable, however, has increased by $5,000. This means a portion of the taxes has not been paid. As a result:

  Income tax expense: $30,000
  Deduct increase in taxes payable: $5,000
  Income tax paid: $25,000

The presentation of the direct method for reporting net cash flow from operating activities:

Indirect Method

The indirect method uses net income (as reported in the income statement) as the starting point in the computation of net cash flows from operating activities. Adjustments to net income necessary to arrive at net cash flows from operating activities fall into three categories: non-cash expenses, timing differences, and non-operating gains and losses. Adjustments reconcile net income (accrual basis) to net cash flows from operating activities. In other words, the indirect method adjusts net income for items that affected reported net income but did not affect cash.

The four-step process:

1. Start with net income.

2. Add back non-cash charges such as depreciation and amortization of intangibles. Cash payments for long-lived assets such as plants and intangibles occur when they are purchased. Purchase of these assets is reflected as an investing activity at that time. When depreciation expense is recognized in the current period, it simply indicates the paper allocation of original purchase cost to this period. As a result, expenses increase without a corresponding cash outlay. Since depreciation does not affect cash flow, it should be added back to net income to compute net CFO.

3. Add back losses and subtract gains from investing or financing activities. Examples include gains/losses from sale of property, plants and equipment (investing activity) or gains/losses from early retirement of debt (financing activity). Why? Disposal of fixed assets will be used to illustrate this. The gains and losses from the disposal of fixed assets appear on the income statement. However, disposal of fixed assets is an investing activity, so the entire cash receipt is shown as an investing cash inflow. Therefore, the gains or losses should be removed from net income so as to prevent double-counting cash flows. Note that it is the proceeds from disposal, not the gain or loss, that constitute the cash flow.

4. Adjust for changes in operating related accounts (current assets and current liabilities other than cash, short-term borrowings, and short-term investments). For example, an increase in current assets ties up cash, thereby

reducing operating cash flow. An increase in current liabilities postpones cash payments, thereby freeing up cash and increasing operating cash flows in the current period. An increase in assets reduces cash and should be deducted from net income. An increase in liabilities increases cash and should be added to net income.

Note that short-term investments are considered an investing activity and short-term borrowing is considered a financing activity.

Example

Selton Co.'s balance sheet and income statement are presented below.

Additional information:
(a) Operating expenses include a depreciation expense of $34,000 and amortization of pre-paid expenses of $2,000.
(b) Land was sold at its book value for cash.
(c) A cash dividend of $48,000 was paid in 2000.
(d) An interest expense of $8,000 was paid in cash.
(e) Equipment with a cost of $36,000 was purchased for cash. Equipment with a cost of $24,000 and a book value of $18,000 was sold for $16,000 cash.
(f) Bonds were redeemed at their book value for cash.
(g) Common stock ($1 par value) was issued for cash.

Explanations of the adjustments to the net income of $57,000 are as follows:

a. Accounts receivable: The decrease of $2,000 should be added to net income to convert from the accrual basis to the cash basis.
b. Inventories: The increase of $60,000 represents an operating use of cash for which an expense was not incurred. This amount is therefore deducted from net income to arrive at cash flow from operations.
c. Pre-paid expense: The decrease of $2,000 represents a change to the income statement for which there was no cash outflow in the current period. The decrease should be added back to net income.
d. Accounts payable: When it increases, the cost of goods sold and the expense on a cash basis are lower than they are on an accrual basis. The increase of $3,000 should be added back to net income.
e. Depreciation expense: The depreciation expense for the building is $20,000. Due to the sale of equipment, the depreciation for equipment is (24,000 - 18,000) + 20,000 - 12,000 = $14,000. This amount plus $20,000 should be added back to net income to determine net cash flow from operating activities.
f. Loss on sale of equipment: The loss of $2,000 on sale of equipment should be added back to net income since the loss did not reduce cash (but it did reduce net income).

Cash flows from investing and financing activities:

a. Land: The sale of land for $20,000 is an investing cash inflow.
b. Equipment: The purchase of equipment for $36,000 is an investing cash outflow and the sale for $16,000 is an investing cash inflow.
c. Bonds payable: This financing activity used $40,000 cash.
d. Common stock: Common stock of $80,000 was issued as a financing cash inflow.
e. Retained earnings: The increase of $9,000 is the result of net income of $57,000 from operations and the financing activity of paying cash dividends of $48,000.

The statement of cash flows is prepared as follows:

Conversion of Cash Flows from the Indirect to the Direct Method

Although the indirect method is most commonly used by companies, the analyst can generally convert it to the direct format by following a simple three-step process.

- Aggregate all revenue and all expenses.
- Remove all non-cash items from aggregated revenues and expenses and break out remaining items into relevant cash flow items.
- Convert accrual amounts to cash flow amounts by adjusting for working capital changes.

3. Cash Flow Statement Analysis

```
<b>Evaluation of the Sources and Uses of Cash</b><p> </p>
```

Analysts should assess the sources and uses of cash between the three main categories and investigate what factors drive the change of cash flow within each category. For example, if operating cash flow is growing, does that indicate success as the result of increasing sales or expense reductions? Are working capital investments increasing or decreasing? Is the company dependent on external financing? Answers to questions like these are critical for analysts and can help form a foundation for evaluating the financial health of an industry or company.

Please refer to the textbook for specific examples.

Common-Size Analysis of the Statement of Cash Flows

This topic will be discussed in detail in Reading 20 [Financial Analysis Techniques].

Free Cash Flow to the Firm and Free Cash Flow to Equity

From an analyst's point of view, cash flows from operation activities have two major drawbacks:

- CFO does not include charges for the use of long-lived assets. Recall that depreciation is added back to net income in arriving at CFO.
- CFO does not include cash outlays for replacing old equipment.

Free Cash Flow (**FCF**) is intended to measure the cash available to a company for discretionary uses after making all required cash outlays. It accounts for capital expenditures and dividend payments, which are essential to the ongoing nature of the business.

The basic definition is cash from operations less the amount of capital expenditures required to maintain the company's present productive capacity.

Free cash flow = CFO - capital expenditure

Free Cash Flow to the Firm (**FCFF**): Cash available to shareholders and bondholders after taxes, capital investment, and WC investment.

FCFF = NI + NCC + Int (1 - Tax rate) - FCInv - WCInv

- NI: Net income available to common shareholders. It is the company's earnings after interest, taxes and preferred dividends.
- NCC: Net non-cash charges. These represent depreciation and other non-cash charges minus non-cash gains. The add-back of net non-cash expenses is usually positive, because depreciation is a major part of total expenses for most companies.
- Int (1 - Tax rate): After-tax interest expense. Add this back to net income because:
    - FCFF is the cash flow available for distribution among all suppliers of capital, including debt-holders, and
    - Interest expense net of the related tax savings was deducted in arriving at net income.

  The add-back is after-tax, because the discount rate in the FCFF model (WACC) is also calculated on an after-tax basis.
- FCInv: Investment in fixed capital. It equals capital expenditures for PP&E minus sales of fixed assets.

- WCInv: Investment in working capital. It equals the increase in short-term operating assets net of operating liabilities.

Example

Quinton is evaluating Proust Company for 2014. Quinton has gathered the following information (in millions):

- Net income: $250
- Interest expense: $50
- Depreciation: $130
- Investment in working capital: $20
- Investment in fixed capital: $100
- Tax rate: 30%
- Net borrowing: $180
- Proust has launched a new product in the market. It has capitalized $200 as an intangible asset out of a product launch expense of $240.
- During the year, Proust has written down restructuring non-cash charges amounting to $30.
- The tax treatment of all non-cash items is the same as that of other items in the books. There are no differed taxes incurred.

Calculate the FCFF for Proust for the year.

Solution

NCC = Depreciation + non-cash restructuring charges - Cash expenses during the year in which they are capitalized = 130 + 30 - 200 = -$40 million

FCFF = NI + NCC + Int (1 - Tax rate) - FCInv - WCInv = 250 + (-40) + 50 (1 - 0.3) - 20 - 100 = $125 million

FCFF can also be computed from cash flow from operating activities (CFO).

FCFF = CFO + Int (1 - Tax rate) - FCInv

The convenience of this approach to calculation of FCFF is that CFO is already adjusted for non-cash charges and changes in working capital accounts.

Example

Uwe is doing a valuation of TechnoSchaft for fiscal year 2004, using the following information (in millions).

- CFO: $250
- Depreciation: $80
- Interest expense: $50
- Tax rate: 30%
- Investment in working capital: $60
- Investment in fixed capital: $240
- Net borrowing: $180

Calculate the FCFF for the company for the year.

Solution

FCFF = CFO + Int (1 - tax rate) - Investment in fixed capital = 250 + 50 (1 - 0.3) - 240 = $45 million

As CFO is given, information on WCInv and non-cash charges is not required.

Free Cash Flow to Equity (**FCFE**): Cash available to stockholders after payments to and inflows from bondholders. This is the cash flow from operations net of capital expenditures and debt payments (including both interest and repayment of principal).

FCFE = FCFF + Net borrowing - Int (1- Tax rate)

FCFE can be calculated from net income. Recall that FCFF = NI + NCC + Int (1 - Tax rate) - FCInv - WCInv. Then:

FCFE = NI + NCC + Net borrowing - FCInv - WCInv

FCFE can be calculated from CFO.

FCFE = CFO + Net borrowing - FCInv

This is different from the formula given in the textbook since net debt repayment should be included in net borrowing!

Cash Flow Ratios

The cash flow statement may also be used in financial ratios measuring a company's profitability, performance, and financial strength.

Performance Ratios

- Cash flow to revenue = CFO / Net revenue: cash generated per dollar of revenue.
- Cash return on assets = CFO / Average total assets: cash generated from all resources.
- Cash return on equity = CFO / average shareholders' equity: cash generated from owner resources.
- Cash to income = CFO / Operating income: cash-generating ability of operations.
- Cash flow per share = (CFO - Preferred dividends) / number of common shares outstanding: operating cash flow on a per-share basis.

Coverage Ratios

- Debt coverage = CFO / Total debt: financial risk and financial leverage.
- Interest coverage = (CFO + Interest Paid + Taxes paid) / Interest paid: ability to meet interest obligations.
- Reinvestment = CFO / Cash paid for long-term assets: ability to acquire assets with operating cash flows.
- Debt payment = CFO / Cash paid for long-term debt repayment: ability to pay debt with operating cash flows.
- Dividend payment: CFO / Dividends paid: ability to pay dividends with operating cash flows.
- Investing and financing: CFO / Cash outflows for investing and financing activities: ability to acquire assets, pay debts, and make distributions to owners.

Financial Analysis Techniques

1. Analysis Tools and Techniques

```
   Financial analysis techniques are useful in summarizing financial reporting data and e
```

Ratios

Ratios express one quantity in relation to another. As analytical tools, ratios are attractive because they are simple and convenient. They can provide a profile of a company, its economic characteristics and competitive strategies, and its unique operating, financial, and investment characteristics.

Ratio analysis is essential to comprehensive financial analysis. However, analysts should understand the following aspects when dealing with ratios:

- A ratio is not "the answer." A ratio is an indicator of some aspect of a company's performance *in the*

past. It does not reveal **why** things are as they are. Also, a single ratio by itself is not likely to be very useful. For example, a current ratio of 2:1 may be viewed as satisfactory. If, however, the industry average is 3:1, such a conclusion may be questionable.
- Differences in accounting policies can distort ratios (e.g., inventory valuation, depreciation methods).
- Not all ratios are necessarily relevant to a particular analysis. Analysts should know the questions for which they want to find answers and know the questions that particular ratios can help answer.
- Ratio analysis does not stop with computation; interpretation of the result is also important.

Limitations:

There are a significant number of estimates and subjective information that go into financial statements and therefore it is imperative that the analyst understands the numbers before calculating and relying on ratio analyses based on these numbers. An analyst needs to ask questions like:

- How homogeneous is the company? Are the ratios comparable between divisions within a company? It is critical to derive comparable industry ratios. However, many companies have multiple lines of business, making it difficult to identify the appropriate industry to use in comparing companies. Companies are required to provide segmented information that allows the user to see the impact of various segments on the overall company.

- Are the results of the ratio analysis consistent? An analyst needs to look at several ratios in conjunction in order to form a sensible conclusion. The total portfolio of the company should be used instead of only one set of ratios. A company must be viewed along all these lines since the company may have strengths and/or weaknesses in different areas. For example, a highly profitable company may have very poor short-term liquidity.

- Is the ratio within a reasonable range for the industry? Analysts must look at a range of values for a particular ratio because a ratio can be too high or too low.

- Are alternative companies' accounting treatments comparable? In comparsons companies, even within the same industry, companies may be using different accounting treatments and/or different estimates to capture the same event. Companies can use different estimates to calculate depreciation or bad debt expenses. Companies can use different inventory methods and may have operating versus capital leases in the financial statements. All of these accounting choices and estimates affect financial statements. Alternative treatments can cause a difference in results for the same events, especially when dealing with non-U.S. companies.

Common-size Analysis

Raw numbers hide relevant information that percentages frequently unveil. Common-size statements normalize balance sheet, income statement, and cash flow statement items to allow easier comparison of different-sized companies. They reduce all the dollar amounts to a percentage of a common amount.

- A **common-size balance sheet** expresses all balance sheet accounts as a percentage of total assets.

- A **common-size income statement** expresses all income statement items as a percentage of sales. Common-size income statement ratios are especially useful in studying trends in costs and profit margins.

- Analysts can use common-size statement analysis for the **cash flow statement**. Two prescribed approaches are the total cash inflows/total cash outflows method and the percentage of net revenue method.

These ratios can be used to compare financial statements of different-size companies or of the same company over different periods.

- **Cross-sectional analysis** compares a company to the industry or other comparable companies for a particular ratio. Comparable companies should be from the same industry, employ the same technology, appeal to similar clienteles, pursue similar marketing strategies, and be similar in size (as measured by sales or total assets). The division of the industry into subsets according to size and characteristics may be valuable in making meaningful comparisons. For example, the "computer industry" covers a huge range of companies and it may not be meaningful to compare IBM to the ratios exhibited by small companies. Multi-industry companies can be handled by appropriate cross-sectional analysis with companies operating across the same range of industries, or alternatively by constructing composite industry average ratios to match a company's structure. However, comparability may be impaired due to the fact that no two firms are exactly the same, and companies may use different depreciation and inventory valuation methods.

- **Time series (trend) analysis** compares a company's performance to itself over time to examine the trend of a particular ratio. It aims to detect changes in the company's operations over time. However, several problems are inherent:
  - Ratios may be affected by changes in a company's internal factors, such as production, marketing, and financial policies.
  - Ratios are based on accounting data provided on a historical cost basis; thus, inflation may cause spurious trends in ratios or render them non-comparable.
  - Comparability may also be impaired due to changes in depreciation or inventory valuation methods across time.
  - The benchmark comparison ratios may suffer from the same problems, leaving the company's problems or strengths unnoticed.

Common-size statement ratios are useful to:

- Compare companies of different sizes.
- Identify trends over time within an individual company.
- Examine the relative size of items or the relative change in items in a company's financial statement.

Graphs

Graphs facilitate comparison of performance and financial structure over time, highlighting changes in significant aspects of business operations. In addition, graphs provide analysts with a visual overview of risk trends in a business. Graphs may also be used effectively to communicate analysts' conclusions regarding aspects of financial conditions and risk management.

2. Common Ratios

```
   The ratios presented in the textbook are neither exclusive nor uniquely "correct." The
```

Activity Ratios

A company's operating activities require investments in both short-term (inventory and accounts receivable) and long-term (property, plant, and equipment) assets. Activity ratios describe the relationship between the company's level of operations (usually defined as sales) and the assets needed to sustain operating activities. They measure how well a company manages its various assets.

- **Receivables turnover** measures the liquidity of the receivables - that is, how quickly receivables are collected or turn over. The lower the turnover ratio, the more time it takes for a company to collect on a sale and the longer before a sale becomes cash.

  This ratio provides a better level of detail than the current or quick ratio. A company could have a favorable current or quick ratio, but if the receivables turn over very slowly, these ratios would not be a good measure of liquidity. The same applies for the inventory turnover below.

  This ratio also implies an average collection period (the number of days it takes for the company's customers to pay their bills):

Remember, as with all ratios, these ratios are industry specific. The nature of the industry dictates a higher or lower receivables or inventory turnover. For receivables turnover, analysts don't want to derive too much from the norm, since a low number indicates slow-paying customers that cause capital to be tied up in receivables and bad debt and a high number indicates overly stringent credit terms that hurt sales. If a company's credit policy is 30 days and the days of sales outstanding is 45 days, then the credit policy needs to be reviewed.

- **Inventory turnover** measures how fast the company moves its inventory through the system. The lower the turnover ratio, the longer the time between when the good is produced or purchased and when it is sold.

  An abnormally high inventory turnover and a short processing time could mean inadequate inventory, which could lead to outages, backorders, and slow delivery to customers, adversely affecting sales. An extremely low inventory turnover value implies capital is being tied up in inventory and could signal obsolete inventory.

- **Payable turnover** measures the length of time a company has to pay its current liabilities to suppliers. This ratio examines the use of trade credit. The longer the time, the better it is for the company, since it is an interest-free loan and offsets the lack of cash from receivables and inventory turnovers.

  The following measures the number of days it takes for the company to pay its bills.

- **Working capital turnover** measures how efficiently a company generates revenue with its working capital. Working capital is defined as current assets minus current liabilities.

- **Total asset turnover** is a measure of how many dollars of sales are generated by a dollar of assets. This number is smaller in capital-intensive industries since they have a higher investment in property, plants and equipment. It is also affected by the amount of leasing that is done by a company. Therefore, the ratio is very industry specific.

  Accounting choices do affect ratios. Analysts must always be aware of accounting choices and how they affect the calculation of ratios. If a company leases most of its assets and they are classified as operating leases, which means they are not recorded on the balance sheet, then the company will have fewer assets and therefore a higher turnover ratio.

  If a company writes down assets due to impairment, in the years following the write-down, assets will be lower and the turnover ratio higher.

  The range of the asset turnover values should be consistent with the industry.

  - An exceedingly high ratio might imply too few assets for the potential business (sales) or the use of outdated, fully depreciated assets.
  - A low ratio might imply capital tied up in an excess of assets relative to the needs of the company.

  The utilization of specific assets (e.g., receivables, inventories) should be examined to identify the cause of the change in total asset turnover.

- **Fixed asset turnover** is a measure of a company's utilization of fixed assets.

net assets = gross fixed assets - depreciation on fixed assets.

This ratio should be compared to the industry norm. The impact of leased assets should be considered, particularly for industries such as retail and airlines that lease most of their fixed assets.

An abnormally high turnover ratio indicates that a company does not have enough capacity to meet potential demand or that the company may have obsolete equipment. Therefore, age of assets also affects this ratio. A higher ratio, caused by old assets, might look positive on the surface, but it would also imply the need for capital expenditures in the near future. Analysts must consider other factors (like age of assets) when considering which company is in a better position for the future. Age of assets is viewed as a factor by analysts because older assets imply the need for future cash outflows to replace assets. The ratio might look better for a company with older assets but the company is actually in a less favorable position than a comparable company that has replaced its assets and is using more efficient assets.

An abnormally low turnover ratio indicates that too much capital is tied up in excess assets.

Liquidity Ratios

Liquidity ratios measure the ability of a company to meet short-term obligations. Major liquidity ratios such as current ratio, quick ratio, and cash ratio are discussed in Reading 18 [Understanding Balance Sheets].

- The **defensive interval ratio** measures how long a company can pay its daily cash expenditures using only its existing liquid assets, without additional cash flow coming in. It provides an intuitive "feel" for a company's liquidity, albeit a most conservative one. It compares currently available "quick" sources of cash with the estimated outflows needed to operate the company: projected expenditures.

  This ratio represents a "worst case" scenario, indicating the number of days a company could maintain its current level of operations with its present cash resources but without considering any additional revenues.

- The **cash conversion cycle** is the time period that exists from when the company pays out money for the purchase of raw materials to when it gets the money back from the purchasers of the company's finished goods. In short, it measures the number of days the company's cash is tied up in the business. When the company buys raw materials, it commits capital to inventory. At the same time, suppliers provide interest-free loans to the company by carrying its payables, thus offsetting the company's capital commitment. After the products are sold, the capital is then tied up in receivables for the collection period. The cash conversion cycle is a measure of how fast a dollar spent returns to the company in payment for a sale. A very high cash conversion cycle indicates that too much capital is tied up in the sales process.

  Cash Conversion Cycle = DOH + DSO - Number of Days of Payables

Note: Analysts should be aware of the impact of accounting choices and accounting transactions on liquidity ratios. For example, payment of an accounts receivable has no effect since one current asset is increasing and another is decreasing. Capitalizing a lease decreases the current ratio, since capitalizing a lease puts a liability on the balance sheet and the portion due in the next year is classified as a current liability. An increase in the turnover ratio decreases the number of days for collection of a receivable or sale of inventory and hence shortens the cash conversion cycle. Use of LIFO versus FIFO in periods of rising prices results in a lower inventory balance and hence a lower current ratio.

Solvency Ratios

Solvency ratios measure the ability of a company to meet long-term obligations. Major solvency ratios such as **debt ratios** are discussed in Reading 18 [Understanding Balance Sheets].

Coverage Ratios

- **Interest coverage ratio** measures how many times over a company could pay its interest out of earnings. The higher the ratio, the less likely that the company cannot meet its interest payments, and consequently, the lower the financial risk.

  This ratio shows how far earnings could decline before it would be impossible to pay interest charges from current earnings.

- **Fixed charge coverage** measures a company's ability to meet all fixed payment obligations, including interest payment on debt, entire lease payments (not limited to the interest component), and dividends on preferred stock. Lease payments are fixed payments just like debt and interest payments. Preferred dividends are grossed-up (on the same basis as the other items in the formula), because preferred dividends are paid from after-tax dollars.

Analysts may be required to adjust these ratios for accounting choices as well. If interest is capitalized, it reduces the interest expense in the current year and the interest is added to the cost of the asset. As a result, interest coverage will look more positive, since some of the interest for the current year will not be included in the interest expense and, hence, the ratio will be higher.

Profitability Ratios

Profitability ratios measure the ability of a company to generate profits from revenue and assets.

Return on Sales

Net profit margin and gross profit margin are discussed in Reading 17 [Understanding Income Statements].

- **Operating profit margin** relates a company's operating income to its net sales.

  Operating profit, also known as earnings before Interest and Taxes (EBIT), is gross profit minus sales, general and administrative expenses (SG&A). It is profit before interest and taxes. The variability of this ratio over time is a prime indicator of the business risk of a company.

- **Pre-tax margin**:

Since profit margin is valuable as a predictor of future earnings, an analyst needs to decide whether to back out other items, such as restructuring charges, in determining what represents income from "continuing" operations. These non-recurring items are considered part of continuing operations but may not be the best predictor of future earnings. Given the current problems in financial reporting, analysts should consider whether certain income statement items should be added/deleted from net income to obtain a better indicator of future earnings. In addition, any "quality of earnings" items should be considered, as should their potential effect on these operating performance ratios. For example, if a company decreased its bad debt expense calculation, it would improve the current year's net income but this might result in a larger expense being recorded in subsequent years. Analysts should be prepared to answer numerous ratio questions based on quality of earnings issues and their effects on ratios.

Return on Investment

The following ratios measure the percentage returns on capital employed.

- **Operating ROA** =

- **ROA** measures the return earned by a company on its assets.

  The problem is that net income is the return to equity-holders, whereas assets are financed by both equity-holders and creditors.

- **Return on total capital** indicates a company's return on all the capital employed (debt, preferred stock, and common stock). It measures return on all sources of funding.

  where total capital = debt + equity.

  This ratio should match the perceived risk of the company. A return well below the industry norm shows the company is losing its competitiveness within the industry.

- **ROE (Return on equity)** measures return on total equity capital only. This includes both preferred and common equity owners.

- **Return on common equity** is also a useful indicator. This ratio is

  This ratio reflects the financial risk assumed by the common stockholder.

Integrated Financial Ratio Analysis

Ratios can also be combined and evaluated as a group to better understand how they fit together and how efficiency and leverage are tied to profitability. The information from one ratio category can be helpful in answering questions raised by another category. The most accurate overall picture comes from integrating information from all sources. Please refer to the textbook for examples.

3. The DuPont System

```
The breakdown of ROE into component ratios to assess the impact of those ratios is ger
```

Traditional DuPont equation:

- ROE = net income / common equity
- ROE = (net income / net sales) x (net sales / common equity). Therefore, ROE = (net profit margin) x (equity turnover).
- ROE = (net income / net sales) x (net sales / total assets) x (total assets / common equity)

Each of these components impacts the overall return to shareholders. An increase in profit margin, asset turnover, or leverage can all increase the return. There is a downside as well. If a company loses money in any year, the asset turnover or financial leverage multiplies this loss effect.

This implies that to improve its return on equity, a company should become more:

- Profitable (increase net profit margin, e.g., pricing and expense control).
- Efficient (increase total asset turnover, e.g., efficiency of asset use).
- Leveraged (increase its financial leverage ratio).

A company's over- or underperformance on ROA is due to one or both of these causes, or "drivers."

The **extended DuPont model** takes the above three factors and incorporates the effect of taxes and interest based on the level of financial leverage. It takes the profit margin and backs up to see the effect of interest and taxes on the overall return to shareholders. Therefore the extended model starts with EBIT (Earnings Before Interest and Taxes) rather than net income.

- EAT = EBT (1 - t), where t is the company's average tax rate. Substituting EBT(1 - t) for EAT in the expanded ROE equation gives us ROE = (EBT / sales)(sales / assets)(assets / equity)(1 - t).

- EBT = EBIT - I, where I equals the company's total interest expense. Substituting (EBIT - I) into the ROE equation for EBT gives us ROE = [(EBIT / sales)(sales / assets) - (interest expense / assets)] (assets / equity) (1 - t).
- Restated in accounting terms:

ROE = [(operating profit margin) x (total asset turnover) - (interest expense rate)] x (financial leverage multiplier) x (tax retention rate)

High financial leverage does not always increase ROE; higher financial leverage will lead to a higher interest expense rate, which may offset the benefits of higher leverage.

This breakdown will help an analyst understand what happened to a company's ROE and why it happened.

4. Ratios Used in Equity Analysis, Credit Analysis, and Segment Analysis

```
<b>Equity Analysis</b><p> </p>
```

Analysts need to evaluate a company's performance in order to value a security. One of their valuation methods is the use of valuation ratios.

Some common valuations ratios are:

- P/E: Price per share / Earnings per share.

  P/E is widely recognized and used by investors. Earning power is a chief driver of investment value, and EPS is perhaps the chief focus of security analysts' attention.

- P/CF: Price per share / Cash flow per share.

  Cash flow is less subject to manipulation by management than earnings. Thus, P/CF ratios can be used to compare companies with different degrees of accounting aggressiveness. Moreover, cash flow is generally more stable than earnings, so P/CF ratios are more stable than P/Es. When EPS is abnormally high, low, or volatile, P/CF ratios are more reliable than P/Es.

- P/S: Price per share / Sales per share.

  - Sales growth is the driving force for the growth of earnings and cash flows.
  - Sales are positive even when EPS is negative.
  - Sales are generally less subject to distortion or manipulation than are other fundamentals.

- P/BV: Price per share / Book value per share.

  Similarly, book value is generally positive even when EPS is negative. Since book value per share is more stable than EPS, P/B may be more meaningful than P/E when EPS is abnormally high or low, or is highly variable.

These price multiples and dividend-related quantities are discussed in more detail in Study Session 12 (Equity Analysis and Valuation).

Credit Analysis

Credit analysis is the evaluation of credit risk.

How does an analyst who has calculated a ratio know whether it represents good, bad, or indifferent credit quality? Somehow, the analyst must relate the ratio to the likelihood that a borrower will satisfy all scheduled interest and principal payments in full and on time. In practice, this is accomplished by testing financial ratios as predictors of the borrower's propensity not to pay (to default). For example, a company with high financial leverage is statistically more likely to default than one with low leverage, all other factors being equal.

Similarly, high fixed-charge coverage implies less default risk than low coverage. After identifying the factors that create high default risk, the analyst can use ratios to rank all borrowers on a relative scale of propensity to default.

Many credit analysts conduct their ratio analyses within ranking frameworks established by their employers. In the securities field, bond ratings provide a structure for analysis. Credit rating agencies such as Moody's and Standard & Poor's use financial ratios when assigning a credit rating to a company's debt issues. For example, credit ratios used by Standard & Poor's include EBIT interest coverage, funds from operations to total debt, total debt to EBITDA, and total debt to total debt plus equity.

Much research has been performed on the ability of ratios to assess the credit risk of a company (including the risk of bankruptcy) and predict bond ratings and bond yields.

Segment Analysis

A company may be involved in many different businesses, may do business in many different geographic areas, or may have significant number of customers. It is difficult to analyze a company with multiple business lines because of the inherent differences in financial structures, risk characteristics, etc,. amount the different lines. Aggregation of financial results for all the lines tends to obscure the true picture.

A company must disclose information related to various subdivisions of its business.

- Under IAS 14 (Segment Reporting), disclosures are required for reportable segments.
- U.S. GAAP requirements are similar to IFRS but less detailed.

Based on the limited segment information that companies are required to present, a variety of useful ratios can be computed for business segments to evaluate how units within a business are doing. These ratios include segment margin, segment turnover, segment ROA, and segment debt ratio.

Analysts should pay attention to a segment's relative performance and see how this may fit into the overall business strategy. For example, in January 2000 Motorola's stock price was $139, on its way to a March peak of $180. But Meyer, an analyst for an investment research company, didn't buy the Street's analysis. "The market was valuing each segment at best-in-class multiples," he said. "But Motorola was not best in class in semiconductors, or wireless, or satellites." The margins of each of Motorola's segments compared poorly with top performers (such as Intel), and Meyer issued warnings that accurately foreshadowed the stock's subsequent plunge. "It was the segment information that helped me to really get a grip on the company," he said.

5. Model Building and Forecasting

```
   The results of financial analysis provide valuable inputs into forecasts of future ear
```

- **Sensitivity Analysis**. This is the study of how the variation in the output of a model can be apportioned to different sources of variation. (e.g., what will be the net income if more debt is issued?)

- **Scenario Analysis**. This considers both the sensitivity of financial outcome to changes in key financial variables and the likely range of variable values. The least "reasonable" set of circumstances (low unit sales, high construction costs, etc.) and the most "reasonable" set are specified first. The financial outcomes under the bad and good conditions are then calculated and compared to the expected, or base-case, outcome. Even though there are an infinite number of possibilities, scenario analysis only considers a few discrete outcomes.

- **Monte Carlo Simulation**. This is a risk analysis technique in which a computer is used to simulate probable future events and thus estimate the profitability and risk of a project. Random values of input variables are generated on a computer. The mean of the target variable is computed to measure the expected value. Standard deviation (or coefficient of variation) is computed to measure risks.

Financial Reporting and Analysis (3)

Inventories

1. Cost of Inventories

There are two basic issues involved in inventory accounting:<p> </p>

1. Determine the cost of goods available for sale: Beginning Inventory + Purchases.
2. Allocate the cost of total inventory costs (cost of goods available for sale) between two components: COGS on the income statement and the ending inventory on the balance sheet. Note that COGS = (Beginning Inventory + Purchases) - Ending Inventory. The cost flow assumption to be adopted includes specific identification, average cost, FIFO, LIFO, etc. This issue will be discussed in subsequent subjects.

Determination of Inventory Cost

IFRS and SFAS No. 151 provide similar treatment of the determination of inventory costs.

The cost of inventories, **capitalized inventory costs**, includes all costs incurred in bringing the inventories to their present location and condition.

- It includes production costs, invoice price (net of discount), transportation costs, taxes, part of fixed production overhead, etc.
- It does not include all abnormal costs incurred due to waste of materials, abnormal waste incurred for labor and overhead conversion costs from the production process, any storage costs, or any administrative overhead and selling costs. These costs are typically expensed in the accounting period instead of being considered inventory costs.

2. Inventory Valuation Methods

In some cases, it's possible to specifically identify which inventory items have been

The remaining three methods are referred to as cost flow assumptions under GAAP and cost formulas under IFRS. They should be applied only to an inventory of homogeneous items. The cost flow assumption may or may not reflect the physical flow of inventory.

Weighted Average Cost

Using the weighted average cost method, the average cost of all units in the inventory is computed and used in recording the cost of goods sold. This is the only method in which all units are assigned the same (average) per-unit cost.

- Average cost = (beginning inventory + purchases) / units available for sale
- Ending inventory = average cost x units of ending inventory
- COGS = cost of goods available for sale - ending inventory

FIFO

FIFO is the assumption that the first units purchased are the first units sold. Thus inventory is assumed to consist of the most recently purchased units. FIFO assigns current costs to inventory but older (and often lower) costs to the cost of goods sold.

LIFO

LIFO is the assumption that the most recently acquired goods are sold first. This method matches sales revenue with relatively current costs. In a period of inflation, LIFO usually results in lower reported profits and lower income taxes than the other methods. However, the oldest purchase costs are assigned to inventory, which may result in inventory becoming grossly understated in terms of current replacement costs.

LIFO is not allowed under IFRS. In the U.S., however, LIFO is used by approximately 36 percent of U.S. companies because of potential income tax savings.

Comparison of Inventory Accounting Methods

Inventory data is useful if it reflects the current cost of replacing the inventory. COGS data is useful if it reflects the current cost of replacing the inventory items to continue operations.

During periods of stable prices, all three methods will generate the same results for inventory, COGS, and earnings.

During periods of rising prices and stable or growing inventories, FIFO measures assets better (the most useful inventory data) but LIFO measures income better.

- Under LIFO, the cost of ending inventory is based on the earliest purchase prices, and thus is well below current replacement cost. For many firms using LIFO, the cost of inventory may be decades old and almost useless for analysis purposes. However, the cost of goods sold is based on the most recent purchase prices, and thus closely reflects current replacement costs. As a result, LIFO provides a better measurement of current income and future profitability.

- Under FIFO, the cost of ending inventory is based on the most recent purchase prices, and thus closely reflects current replacement cost. However, costs of goods sold are based on the earliest purchase prices, and this is well below the current replacement costs. The gain is actually holding gain or inventory profit. It is debatable whether this should be considered income; at least, analysts can say the underestimated COGS leads to inflated net income.

In an environment of declining inventory unit costs and constant or increasing inventory quantities, the opposite is true.

The usefulness of inventory data reported using the average-cost method lies between LIFO and FIFO.

3. Periodic versus Perpetual Inventory System

```
   The <b>perpetual inventory system</b> updates inventory accounts after each purchase c
```

The **periodic inventory system** records inventory purchase or sale in the "Purchases" account. The "Inventory" account is updated on a periodic basis, at the end of each accounting period (e.g., monthly, quarterly). Cost of goods sold or cost of sale is computed from the ending inventory figure.

With perpetual FIFO, the first (or oldest) costs are the first moved from the Inventory account and debited to the Cost of Goods Sold account. The end result under perpetual FIFO is the same as under periodic FIFO. In other words, the first costs are the same whether you move the cost out of inventory with each sale (perpetual) or whether you wait until the year is over (periodic).

With perpetual LIFO, the last costs available at the time of the sale are the first to be removed from the Inventory account and debited to the Cost of Goods Sold account. Since this is the perpetual system, we cannot wait until the end of the year to determine the last cost. An entry must be recorded at the time of the sale in order to reduce the Inventory account and increase the Cost of Goods Sold account.

If costs continue to rise throughout the entire year, perpetual LIFO will yield a lower cost of goods sold and a higher net income than periodic LIFO. Generally this means that periodic LIFO will result in lower income taxes than perpetual LIFO.

Example

Date..................................Units....Price

12.31.2008........Beginning Inventory....1.......85

1.1.2009..........Purchase..............1.......87

2.1.2009..........Purchase..............2.......89

6.1.2009..........Sales.................1.......89

12.1.2009.........Purchase..............1.......90

Under perpetual LIFO the following entry must be made at the time of the sale: $89 will be credited to Inventory and $89 will be debited from Cost of Goods Sold. If that was the only item sold during the year, at the end of the year the Cost of Goods Sold account will have a balance of $89 and the cost in the Inventory account will be $351 ($85 + $87 + $89 + $90).

Under periodic LIFO we assign the last cost of $90 to the one item that was sold. (If two items were sold, $90 would be assigned to the first item and $89 to the second item.) The remaining $350 is assigned to inventory. The $350 of inventory cost consists of $85 + $87 + $89 + $89. The $90 assigned to the item that was sold is permanently gone from inventory.

4. The LIFO Method

```
   In the U.S., firms that use LIFO must report a LIFO reserve. The <b>LIFO reserve</b> i
```

$$\textbf{Inventory}_{\textbf{FIFO}} = \textbf{Inventory}_{\textbf{LIFO}} + \textbf{LIFO Reserve}$$

It represents the cumulative effect over time of ending inventory under LIFO vs. FIFO.

When adjusting COGS from LIFO to FIFO: $\textbf{COGS}_{\textbf{FIFO}} = \textbf{COGS}_{\textbf{LIFO}}$ **- Change in LIFO Reserve**.

LIFO Liquidations

So far, discussions have been based on the assumptions of rising prices and stable or growing inventory quantity. As a result, the LIFO reserve increases over time. However, LIFO reserves can decline for either of the two reasons listed below. In either case, the COGS will be smaller and the reported income will be higher relative to what they would have been if the LIFO reserve had not declined. However, the implications of a decline in the LIFO reserve on financial analysis vary, depending on the reason for the decline.

- Liquidation of inventories. When a firm reduces its inventory, the old assets flow into income. The COGS figure no longer reflects the current cost of inventory sold. This is called **LIFO liquidation**. Gross profit margin will be abnormally high and unsustainable ("phantom" gross profits). To defer taxes indefinitely, purchases must always be greater than or equal to sales. A LIFO liquidation may signal that a company is entering an extended period of decline (and needs the "profit" to show as income). Analysts should exclude this profit from recurring earnings, as it is not operating in nature; the reported COGS should be restated by adding back the decline in the LIFO reserve to remove the artificial boost to net income.

- Price declines. The lower-cost current purchases enter reported LIFO COGS when purchase prices fall, reducing the cost differences between LIFO and FIFO ending inventories. As a result, the LIFO reserve declines. Such a decline is not considered a LIFO liquidation. Amounts on the balance sheet are still outdated but those on the income statement are still current. However, the tax benefits are lost under LIFO. For analytical purposes, no adjustment is required for declining prices, since price decreases are a normal business situation.

5. Measurement of Inventory Value

```
   Under IFRS, inventories are reported at the lower of cost or net realizable value (NR\
```

- If inventory declines in value below its original cost for whatever reason (obsolescence, price-level changes, damaged goods, etc.), the inventory should be written down to reflect this. If the NRV is lower than the cost, the ending inventory is written down to the NRV. The loss then is charged against revenues *as an expense* in the period in which the loss occurs, not in the period in which it is sold. However, if the NRV is higher than the cost, nothing is done. The increases in the value of the inventory are recognized only at the point of sale.

- A reversal (up to the amount of original write-down) is required if the inventory value goes up later.
- The amount of any reversal is recognized as a reduction in the cost of sales.
- This rule can be applied either directly to each inventory item, to each category, or to the total of the inventory. The most common practice is to price inventory on an item-by-item basis.

IFRS does not apply to the measurement of inventories held by producers of agricultural and forest products, mineral products, or commodity brokers and dealers. Their inventories are measured at net realizable value (above or below cost) in accordance with well-established practices in those industries.

Similarly, GAAP requires the use of the lower-of-cost-or-market valuation basis (LCM) for inventories, with market value defined as replacement cost. Reversal is prohibited, however. The LCM valuation basis follows the principle of conservatism (on both the balance sheet and income statement) since it recognizes losses or declines in market value as they occur, whereas increases are reported only when inventory is sold.

Here are some relevant terms:

- *Net realizable value*: Estimated selling price less estimated costs of completion necessary to make the sale.
- *Historical cost*: The cash equivalent price of goods or services at the date of acquisition.
- *Market value* (*Replacement cost*): The cost that would be required to replace an existing asset.
- *Fair value*: The amount for which an asset could be exchanged, or a liability settled, between knowledgeable, willing parties in an arm's length transaction.

Example

Historical cost: $5,000

Market cost: $2,000

Estimated selling price: $4,000

Estimated costs to complete sale: $1,000

Net realizable value: $4,000 - $1,000 = $3,000

- Inventory Valuation under IFRS: $3,000 (the lower of historical cost and NRV).
- Inventory Valuation under U.S. GAAP: $2,000 (the lower of historical cost and market cost).

Now assume NRV increases from $3,000 to $4,000 and the market cost increases from $2,000 to $3,000.

- Under IFRS, $1,000 of original write-down may be recovered to bring NRV up from $3,000 to $4,000. Note that reversals are limited to the amount of the original write-down ($2,000).
- Under U.S. GAAP, the value of inventory is $2,000 even though the new market value is $3,000. No adjustment is made and reversal is prohibited.

6. Financial Analysis of Inventories

```
  Financial statement disclosures provide information regarding the accounting policies
```

Presentation and Disclosure

Consistency of inventory accounting policy is required under both U.S. GAAP and IFRS. If a company changes an inventory accounting policy, the change must be justifiable and all financial statements accounted for retrospectively. The one exception is for a change to the LIFO method under U.S. GAAP; the change is accounted for prospectively and there is no retrospective adjustment to the financial statements.

Inventory Ratios

Inventory turnover measures how fast a company moves its inventory through the system.

This ratio can be used to measure how well a firm manages its inventories. The lower the ratio, the longer the time between when the good is produced or purchased and when it is sold.

- An abnormally high inventory turnover and a short processing time could mean either effective inventory management or inadequate inventory, which could lead to outages, backorders, and slow delivery to customers (which would adversely affect sales). Revenue growth should be compared with that of the industry to assess which explanation is more likely.
- An extremely low inventory turnover value implies capital is being tied up in inventory and could signal obsolete inventory. Again, the analyst should compare the firm's revenue growth with that of the industry to assess the situation.

Financial Analysis: FIFO versus LIFO

The advantages of LIFO are:

- Matching. Current costs are matched against revenues and inventory profits are thereby reduced.
- Tax benefits. These are the major reason why LIFO has become popular. As long as the price level increases and inventory quantities do not decrease, a deferral of income tax occurs. "Whatever is good for tax is good for financial reporting."
- Improved cash flow. This is related to tax benefits, because taxes must be paid in cash.
- Future earnings hedge. With LIFO, a company's future reported earnings will not be affected substantially by future price declines. Since the most recent inventory is sold first, there isn't much ending inventory sitting around at high prices, vulnerable to a price decline.

The disadvantages of LIFO:

- Reduced earnings. Many managers would just rather have higher reported profits than lower taxes. However, non-LIFO earnings are now highly suspect and may be severely penalized by Wall Street.
- Inventory understated. LIFO may have a distorting effect on a company's balance sheet. It makes the working capital position of the company appear worse than it really is.
- Physical flow. LIFO does not approximate the physical flow of the inventory items except in particular situations.
- Current cost income not measured. LIFO falls short of measuring current cost (replacement cost) income, though not as far as FIFO. Using replacement cost is referred to as the next-in, first-out method; it is not acceptable for purposes of inventory valuation.
- Inventory liquidation. If the base or layers of old costs are eliminated, strange results can occur, because old, irrelevant costs can be matched against current revenues. The income tax problem is particularly severe when involuntary liquidation results from a strike or a shortage of materials; in these situations, companies may incur high tax bills when they can least afford to pay taxes.
- Poor buying habits. A company may attempt to manipulate its net income at the end of the year simply by altering its pattern of purchases.

The choice of inventory system or method affects financial numbers. For example, the following is the comparison between LIFO (Last In, First Out) and FIFO (First In, First Out):

During periods of rising prices and stable or growing inventories:

- COGS and Income. Since LIFO allocates the most recent purchase prices to COGS, the use of LIFO results in higher COGS and lower reported income. In contrast, FIFO allocates the earliest purchase prices to COGS, resulting in lower COGS and higher income.

- Cash Flows. The choice of LIFO vs. FIFO has no effect on pretax cash flows. The pretax cash flow is determined by the cash inflow from sales and cash outflow for purchases, neither of which is affected by the method of inventory accounting. However, the choice of LIFO vs. FIFO affects tax payments. In the U.S., the IRS requires the same inventory methods for financial reporting and tax reporting. Since LIFO generates lower pretax income (when prices are rising), it will result in lower tax payments and, therefore, higher after-tax cash flows than FIFO.

- Working Capital. Working capital is defined as current assets less current liabilities. Since LIFO reports lower inventory than FIFO, working capital will be lower under LIFO.

- Profitability. Profit margin = net income / sales. Sales are not affected by the choice of LIFO or FIFO. Since FIFO results in lower COGS and, therefore, higher net income, profit margins will be higher under FIFO. The net income provided by LIFO is more useful and the lower profit margins reported under LIFO should be used in analysis.

- Liquidity. Current ratio = current assets / current liabilities. Current liabilities are not affected by the choice of FIFO or LIFO. Since LIFO results in lower inventory and, therefore, lower current assets, the current ratio will be lower under LIFO. However, since the inventory provided by FIFO is more useful, the higher current ratio reported under FIFO is better for analytical purposes.

- Activity. Inventory turnover = COGS / average inventory. LIFO provides the more useful COGS while FIFO provides the more useful inventory measure. For analytical purposes, a **current cost** inventory turnover should be computed using LIFO-basis COGS and FIFO-basis inventory: Inventory Turnover (Current Cost) = COGS (LIFO) / Average Inventory (FIFO).

- Solvency. Debt-to-equity ratio = long-term debt/equity. The choice of LIFO or FIFO has no impact on debt. Since FIFO results in higher inventory values, it reports higher equity, so as to reconcile the balance sheet. Therefore, the debt-to-equity ratio will be lower under FIFO. The lower debt-to-equity ratio reported under FIFO should be used in analysis.

  However, a firm that uses FIFO usually does not disclose its equity under FIFO. In this case, the equity under FIFO can be approximated by adding the LIFO reserve to the equity under LIFO: Equity under FIFO = Equity under LIFO + LIFO reserve. Note that the LIFO reserve is not adjusted for taxes because the tax effect would be insignificant during periods of rising prices and stable or growing inventories.

The general guideline is to use LIFO-based numbers for components that are income-related and FIFO-based data for components that are balance-sheet-related. Ideally, firms could have used FIFO to prepare the balance sheet and LIFO to prepare the income statement. In reality this "perfect" combination is not permitted by accounting rules. Analysts should adjust financial statements between FIFO and LIFO to suit their analytic purposes.

Long-lived Assets

1. Capitalizing versus Expensing

```
   The costs of acquiring resources that provide services over more than one operating cy
```

Accounting rules on capitalization are not straightforward. As a result, management has considerable discretion in making decisions such as whether to capitalize or expense the cost of an asset, whether to include interest costs incurred during construction in the capitalized cost, and what types of costs to capitalize for intangible assets. The choice of capitalization or expensing affects the balance sheet, income and cash flow statements, and ratios both in the year the choice is made and over the life of the asset.

Here is a summary of the different effects of capitalization versus expensing:

- Income variability. Firms that capitalize costs and depreciate them over time show "smoother" patterns of reported income. Firms that expense those costs as incurred tend to have higher variability of net income.

- Profitability. In the early years expensing lowers profitability because the entire cost of the asset is expensed. In later years expensing results in higher net income because no more expense is charged in those years. This results in higher ROA and ROE because these expensing firms report lower assets and equity.

- CFO. The net cash flow remains the same, but the compositions of cash flows differ. Cash expenditures for capitalized assets are included in *investing cash flows* and are never classified as *CFO*. In contrast, cash expenditures for expensed outlays are included in *CFO* and are never classified as investing cash flows. Capitalization results in higher CFO but lower investing cash flows, and the cumulative difference increases over time.

- Leverage ratios. Capitalization firms have better (lower) debt-to-equity and debt-to-assets ratios, since they report higher assets and equities.

Under SFAS 34, interest is capitalized for certain assets and only if the firm is leveraged. Therefore, the carrying amount of a self-constructed asset depends on the firm's financial decisions. The capitalized interest cost is added to the value of the asset being constructed.

The amount of interest cost to be capitalized has two components:

- Any interest on borrowed funds made specifically to finance the construction of the asset. The interest rate applicable is the interest rate on each borrowing.
- The interest on other debt of the firm, up to the amount invested in the construction project. The interest rate applicable is the weighted-average interest rate on all outstanding debt not specifically borrowed for the asset under construction.

Therefore, the total interest cost incurred during the accounting period has two parts:

- Capitalized interest cost, which is reported as part of the asset on the balance sheet. Payments for capitalized interest cost are classified as an investing cash outflow and never as CFO.
- Other interest cost, which is charged to expense on the income statement. Payments for such non-capitalized interest cost are reported as CFO.

The total interest cost, along with the amount capitalized, must be disclosed as part of the notes to the financial statements.

Once the construction is complete, capitalized interest costs will be written off as part of depreciation over the useful life of the asset. From now on, any future interest cost on remaining borrowings made for the construction of the asset must be expensed.

For analytical and adjustment purposes, analysts probably need to expense all interest in self-constructed assets (that is, the income statement capitalization of interest should be reversed).

During the construction period this could result in:

- Lower fixed and total assets, as the capitalized interest would be converted to interest expense.
- Lower net income, as the interest expense would be higher.
- Lower CFO and higher CFI as payments for capitalized interest would be classified as investing cash flows and be reversed to be operating cash flows.
- Lower interest coverage ratio, as the adjustment would produce lower earnings before interest and tax but higher interest expense.
- The same net cash flows, as capitalization and expensing are accounting adjustments only. They don't affect net cash flows.

During the useful life of the asset this could result in:

- Higher net income, due to a lower depreciation amount.
- The same interest, as all interest costs would be expensed.
- A higher interest coverage ratio, due to higher earnings before interest and tax (same interest expense).

2. Intangible Assets

```
   Intangible assets are identifiable nonmonetary resources controlled by firms. Examples
```

Accounting for the Acquisition of Long-Lived Intangible Assets

Accounting for an intangible asset depends on how it is acquired.

1. Intangible Assets Purchased in Situations Other than Business Combinations

These are accounted for at acquisition costs. "Cost" includes purchase price, legal fees, and other expenses that make the intangibles ready for use. For example, fees paid to obtain a license or franchise are capitalized. Another example: expenditures on patents and copyrights purchased from another party are capitalized. They are amortized over their remaining legal lives or 40 years, whichever is less. The straight-line method is typically used for amortization.

2. Intangible Assets Developed Internally

For internally generated intangible assets, it is difficult to measure costs, benefits, and economic lives. Generally, internally generated assets (such as costs of R&D, patents and copyrights, brands and trademarks, and advertising and secret processes) must be expensed in the period incurred.

One exception is research and development (R&D) expenditures which add risk to investment with uncertain future economic benefits. As a result, they must be expensed as incurred in most countries (including the U.S.). SFAS 86 requires that all R&D costs to establish the technological and/or economic feasibility of software must be expensed. Subsequent costs that are beyond the point of *technological feasibility* can - but don't have to - be capitalized as part of product inventory and amortized based on revenues or on a straight-line basis. The point of technological feasibility is the point when a software prototype has been proven to be technologically feasible, as evidenced by the existence of a working model of the software.

IFRS also requires research costs be expensed but allows development costs to be capitalized under certain conditions.

As you can see, managers have considerable discretion in making decisions, such as whether or when to capitalize these costs and by how much. For software development costs, one particular risk is that capitalized costs will not be realized and a future write-down may be needed.

If companies apply different approaches to capitalizing software development costs, adjustments can be made to make the two comparable.

3. Intangible Assets Acquired in a Business Combination

Business combinations are accomplished when one entity (investor) acquires "control" over the net assets of another entity. The transaction is accounted for using the purchase method of accounting, in which the company identified as the acquirer allocates the purchase price to each asset acquired (and each liability assumed) on the basis of its fair value.

Any excess of cost over fair value of net assets acquired is recorded as goodwill.

U.S. GAAP requires that in-process R&D (IPRD) of the target company should be expensed at the date of acquisition, which results in a large one-time charge. IFRS requires identifying IFRD as a separate asset with a finite life or including it as part of goodwill.

Amortizing Intangible Assets with Finite Useful Lives

An intangible asset with a finite useful life is amortized over its useful life. The estimates required for amortization calculations are: original valuation amount, residual value at the end of useful life, and the length of useful life.

Example

Torch, Inc. has developed a new device. Patent registration costs consisted of $2,000 in attorney fees and $1,000 in federal registration fees. The device has a useful life of 5 years. The legal life is 17 years. At the end of year 1, what is Torch's amortization expense?

Use the shorter of economic life (5 years) or legal life (17 years): Amortization = Cost / Useful Life = $3,000 / 5 = $600.

Intangible assets without a finite useful life (i.e., with an indefinite useful life) are not amortized, but are reviewed for impairment whenever changes in events or circumstances indicate that the carrying amount of an asset may not be recoverable.

3. Depreciation Methods

```
    For accountants, depreciation is an <b>allocation</b> process, not a <b>valuation</b>
```

The different depreciation methods are:

- **Straight Line Depreciation (SLD)**

  This is the dominant method in the U.S. and most countries worldwide. It is based on the assumption that depreciation depends solely on the passage of time. The amount of depreciation expense is computed as:

  If income is constant, SLD will cause the asset base to decline, causing ROA to increase over time. For assets whose benefit may decline over time, the matching principle supports using an accelerated depreciation method.

- **Accelerated Depreciation Methods**

  Accelerated depreciation methods are consistent with the matching principle because benefits from most depreciable assets are higher in the earlier years as the assets wear out. Therefore, more depreciation should be allocated to earlier years than to later years.

  Under the **sum-of-the-years' digits (SYD)** method, depreciation expense is based on a decreasing fraction of depreciable cost. The numerator decreases year by year but the denominator remains constant. As a result, this method applies higher depreciation expense in the early years and lower depreciation expense in later years.

  Where sum of years = (1 + 2 + 3 + ... + n) = n x (n + 1)/2, and years remaining = n - t + 1 (n: the estimated useful life. t: the index for current year).

  Double decline balance (DDB):

  Note that cost minus accumulated depreciation is the book value at the beginning of the year and that salvage value is not shown in the formula. For each year, however, depreciation is limited to the amount necessary to reduce book value to salvage value.

  With SYD and DDB methods, book value, net income, tax expense, and equity will be lower than with SLD in the earlier years of an asset's life. The percentage effect on net income is usually greater than the effects on assets and shareholders' equity. Consequently:

  - Profit margin is lower as net income is lower.
  - Asset turnover ratio is higher as assets are lower.
  - Debt-to-equity ratio is higher as equity is lower.
  - Return on assets ratio is lower; both net income and total assets are lower, but net income is lower by a larger percentage.
  - Return on equity ratio is lower; both net income and equity are lower, but net income is lower by a larger percentage.

  In later years the situation will reverse and income and book values will increase. This is true for individual assets. For a firm with stable or rising capital expenditures, however, the early-year impact of newly-acquired assets dominates. Therefore, an accelerated depreciation method will continuously result in lower reported earnings and tax expenses for these firms.

- **Units of Production (UOP) and Service Hours Method**

  This method assumes that depreciation depends solely on the use of the asset and bases depreciation on actual service usage:

  - UOP = Depreciation [per period] = Output [per period] x Unit Cost
  - Unit Cost = (Cost - Salvage Value)/Estimated Production Capacity or Estimated Service Life

  Therefore, more depreciation expense is charged in years of higher production. The advantage is that they make depreciation expense a variable rather than a fixed cost, decreasing the volatility of reported earnings as compared to straight-line or accelerated methods. A drawback occurs when the firm's productive capacity becomes obsolete as it loses business to more efficient competitors. These methods will reduce depreciation expense during periods of low production, resulting in overstated reported income and asset value. However, low production is often caused by intensified competition, which tends to reduce the economic value of the asset and thus requires a higher rate of depreciation.

Note that in the U.S. different depreciation methods have the same effect on taxes payable, as the depreciation method (MACRS) used for tax reporting is independent of the method chosen by management for financial reporting. It is taxes payable, not tax expense, that determines cash outlay for tax payment. Therefore, the choice of depreciation methods has no impact on the statement of cash flows.

Estimates Required for Depreciation Calculations

Depreciable life, also called useful life, is the total number of service units expected from a depreciable asset. It can be measured in terms of units expected to be produced, hours of service to be provided by the asset, or years the asset is expected to be used. The longer the depreciable life, the lower the annual depreciation expense.

Reducing the depreciable life of an asset has the following impact on financial statements over its depreciable life:

- Higher depreciation expense.
- Lower book value of the asset.
- Lower net income. The percentage effect on net income is usually greater than the effects on assets and shareholders' equity.
- Lower shareholders' equity (caused by lower retained earnings).

Consequently, a shorter depreciable life tends to reduce profit margin, returns on assets, and returns on equity, while raising asset turnovers and the debt-to-equity ratio. However, changing the depreciable life has no effect on cash flows, since depreciation is a non-cash charge.

Salvage value, also called **residual value**, is the estimated amount that will be received when an asset is sold or removed from service.

- The higher the salvage value, the lower the annual depreciation expense, as salvage value is deducted

from the original cost to compute annual depreciation expense for depreciation methods such as straight-line, units-of-production, service-hour, and sum-of-the-years' digits.
- Salvage value serves as a floor for net book value for depreciation methods such as double-declining-balance, units-of-production, and service-hour depreciation.
- Note that MACRS assumes there is no salvage value.

The effects of choosing a lower salvage value are similar to those of a shorter depreciable life or an accelerated depreciation method. However, the effects do not reverse in the later years of the asset's useful life.

Shorter lives and lower salvage values are considered conservative in that they lead to higher depreciation expense. These factors interact with the depreciation method to determine the expense; for example, use of the straight-line method with short depreciation lives may result in depreciation expense similar to that obtained from the use of an accelerated method with longer lives.

4. The Revaluation Model

```
   Under U.S. accounting standards, it is compulsory to account for impairment in long-li
```

The balance sheet is more informative when assets and liabilities are stated at market value rather than historical cost. IASB and some other non-U.S. GAAP do permit upward revaluations. The purpose of a revaluation is to bring into the books the fair market value of long-lived assets.

- If an asset revaluation *initially decreases* the asset's carrying value, the decrease is recognized as a loss. Later, if there is an increase in the carrying value, the increase is recognized as a profit (up to the amount of the original decrease).
- If an asset revaluation *initially increases* its carrying value, the increase bypasses the income statement and goes to equity (revaluation surplus). Later, if there is a decrease, it first decreases the revaluation surplus, then goes to income.

Financial Statement Analysis Considerations

- The leverage motivation. An upward revaluation may improve a firm's leverage.
- Income manipulation. Revaluations are subjective in nature. For example, a downward revaluation will reduce ROE in the current period but make the firm more profitable in future years, since total assets and shareholders' equity will be lower.
- Revaluation has no impact on cash flows.
- What is the true value of the firm's long-lived assets? Why is the revaluation necessary? Who does the appraisal? How often is it done?

5. Impairment of Assets

```
   Sometimes a long-term asset may lose some of its revenue-generating ability prior to t
```

GAAP and IFRS differ as to the methodology used to determine impairment.

The **GAAP methodology** of determining impairment uses a two-step recoverability test. Occurrence of an impairment differs from recognition of an impairment. An impairment, whether recognized in financial reports or not, occurs as long as an asset's carrying value cannot be fully recovered in the future. However, only impairments that meet certain conditions are recognized in financial reports. SFAS 121 provides a two-step process:

- **Recoverability test**. Impairment must be recognized when the carrying value of the assets exceeds the *undiscounted future cash flows* from their use and disposal.
- **Loss measurement**. The excess of the carrying amount over the fair value of the assets. If the fair value is not available, the *present value* of future cash flows discounted at the firm's incremental borrowing rate should be used. That is:

Impairment Loss = Book Value - Either Fair Value or Present Value of Future Cash Flows

Conversely, **IFRS methodology** uses a one-step approach. This approach requires that impairment loss be calculated if "impairment indicators" exist. This approach does not rely on net undiscounted future cash flows and subsequent comparison to asset carrying value as required in GAAP methodology. In addition, the impairment loss is calculated as the amount by which the carrying amount of the asset exceeds it recoverable amount. The recoverable amount is the higher of the following: 1) fair value less cost to sell, or 2) value in use (i.e., the present value of future cash flows including disposal value).

Impairment of Intangible Assets

- Similar accounting treatment if the intangible asset has a finite life.
- Tested annually for impairment for an intangible asset with an indefinite life.

Among the most interesting intangible assets is *goodwill*. Goodwill is the present value of future earnings in excess of a normal return on net identifiable assets. It stems from such factors as a good reputation, loyal customers, and superior management. Any business that earns significantly more than a normal rate of return actually has goodwill.

Goodwill is recorded in the accounts only if it is purchased by acquiring another business at a price higher than the fair market value of its net identifiable assets. It is not valued directly but inferred from the values of the acquired assets compared with the purchase price. It is the premium paid for the target company's reputation, brand names, customers or suppliers, technical knowledge, key personnel, and so forth.

Goodwill only has value insofar as it represents a sustainable competitive advantage that will result in abnormally high earnings. Analysts need to be aware of the possibility that the goodwill recognized by accountants may, in fact, represent overpayment for the acquired company. Since goodwill is inferred rather than computed directly, it will increase as the payment price increases. It is only after the passage of time that analysts will be able to evaluate the extent to which the purchase price was justified.

Under U.S. GAAP SFAS 142, goodwill is not amortized, but is tested annually for impairment. Goodwill impairment for each reporting unit should be tested in a two-step process at least once a year.

1. The fair value of a reporting unit is compared to its carrying amount (goodwill included) at the date of the periodic review. If the fair value at the review date is less than the carrying amount, then the second step is necessary.

2. The carrying value of goodwill is compared to its implied fair value (and a loss recognized when the carrying value is the higher of the two). To arrive at an implied fair value for goodwill, the FASB specifies that an entity should allocate the fair value of the reporting unit at the review date to all of its assets and liabilities as if the unit had been acquired in a combination with the fair value of the unit as its purchase price. The excess of that fair value (purchase price) over the fair value of the identifiable net asset is the implied fair value of goodwill.

Financial Impacts

The carrying value of long-lived assets should be written down to fair value (less cost of disposal, if intended for sale). The impairment loss is reported pretax as a component of income from continuing operations. Once recognized, the impairment loss cannot be restored.

Some impacts on current financial statements:

- Lower fixed assets and total assets.
- Both net income and tax expense are reduced due to the impairment loss.
- Tax payable: the impairment loss is not recognized for tax purposes until the property is disposed of. It leads to a deferred tax asset (a future tax benefit), not a current refund.
- Stockholders' equity is reduced, and thus the debt-to-equity ratio is increased.
- Impairment write-down has no effect on cash flows, since it is a non-cash charge.

- Asset turnover ratios tend to increase due to the lower asset base.
- Return on assets and return on equity are reduced because of the impairment loss.

The write-down affects future financial statements and ratios in the same way as it affects the current period, except in the following aspects:

- Future depreciation expenses are reduced due to the reduced book value of the asset.
- As a result, future net income and profit margin increases. Note that impairment loss is a one-time loss, and does not affect the income statements of future periods.
- Future return on assets and return on equity will both increase because of higher future profitability and a lower asset and equity base.

Assets Held for Sale

Long-lived assets held for sale are tested for impairment at the time they are categorized as held for sale. If the carrying value > fair value, an impairment is recognized. Assets held for sale should cease to be depreciated after reclassification as held for sale. Subsequent increases in fair value less cost of disposal are recognized as gains only to the extent of previously recognized write-downs.

Reversals of Impairments

U.S. GAAP:

- Not permitted for assets held for use.
- Permitted for assets held for sale.

IFRS:

- Loss may be reversed, but not to exceed the initial carrying amount adjusted for depreciation.

6. Derecognition

```
For accounting purposes, an asset may be disposed of in three different ways. It may k
```

- sold for cash
- exchanged for another asset
- abandoned

When plant assets are disposed of, depreciation should be recorded on the date of disposal. The cost is then removed from the asset account and the total recorded depreciation is removed from the accumulated depreciation account. Normally an asset's market value at the time of sale or disposal will most likely be different than the asset's book value (its original historical cost minus all accumulated depreciation on that asset). The sale of a plant asset at a price above or below book value results in a gain or loss to be reported in the income statement.

Because different depreciation methods are used for income tax purposes, the gain or loss reported on income tax returns may differ from that shown in the income statement. It is the gain or loss shown in the financial statement that is recorded in the company's general ledger accounts.

To illustrate each of these methods consider this. A machine was purchased on 1 January Year 1 for $1,000. The depreciation method was straight-line with a useful-life of 5 years and an estimated residual value of $200. On 31 July Year 3 the firm decides to dispose of the asset. The firm has a December year-end.

The first step irrespective of the method of disposal is to calculate the depreciation up to the date of sale:

Depreciation per year = (Cost - residual value) / Useful life = (1,000 - 200) / 5 = $160

Depreciation for year 1 = $160

Depreciation for year 2 = $160

Depreciation for year 3 = $93 (160 x 7/12)

Total: $413

Remember that the depreciation for year 3 is only for 7 months, as the asset is disposed of on 31 July Year 3.

Sale of Long-Lived Assets

The gain or loss on the sale of long-lived assets is computed as the sales proceeds minus the carrying value of the asset at the time of sale. Assume that the machinery is sold for cash in three scenarios:

a. Sold for $587 cash (Sale of machinery for carrying value)

Debit: Cash (B/S) $587

Debit: Accumulated depreciation (B/S) $413

Credit: Machinery (B/S) $1000

b. Sold for $600 cash (Sale of machinery for above carrying value)

Debit: Cash (B/S) $600

Debit: Accumulated depreciation (B/S) $413

Credit: Machinery (B/S) $1000

Credit: Profit on sale of machinery (I/S) $13

c. Sold for $500 cash (Sale of machinery for below carrying value)

Debit: Cash (B/S) $500

Debit: Accumulated depreciation (B/S) $413

Debit: Loss on sale of machinery (I/S) $87

Credit: Machinery (B/S) $1000

In summary, when disposing of an asset, entries are prepared to:

- eliminate the cost of the asset from the books.
- eliminate the accumulated depreciation from the books.
- record the proceeds on the sale. This is reported as cash from investing activities on the statement of cash flows.
- record the profit or loss on the sale (if applicable). This amount is excluded from net income when the indirect method is used to calculate cash flows from operating activities.

Exchange of Long-Lived Assets

If an asset is exchanged for another asset, the basic accounting is similar to the accounting for sales of plant assets for cash. If the trade-in allowance received is greater than the carrying value of the asset surrendered, there has been a gain. If the allowance is less, there has been a loss.

Level II will cover some special rules for recognizing these gains and losses, depending on the nature of the assets exchanged:

Abandoned

If an asset is discarded, no compensation is received for it and it is taken out of service. As a result, if there is a carrying value left on the accounting records, this would need to be written off as follows:

Debit: Accumulated depreciation (B/S) $413

Debit: Loss on disposal of asset (I/S) $587

Credit: Machinery (B/S) $1000

Discarded machinery is no longer used in the business. The full cost and accumulated depreciation are reversed and the balance taken to the income statement as loss on the disposal of asset account.

Assets that are to be sold are classified as *assets held for sale* instead of *assets held for use*. Long-lived assets to be disposed of other than by sale are classified as *held for use until disposal*; they continue to be depreciated and tested for impairment as required.

7. Presentation and Disclosures

```
<b>Property, Plant, and Equipment (PP&amp;E)</b><p> </p>
```

IAS 16 provides a long list of disclosure requirements for PP&E. For each class of PP&E, the financial statements must disclose the following:

- the measurement bases used for determining the gross carrying amount.
- the depreciation methods and rates or useful lives.
- the gross carrying amount and the accumulated depreciation (aggregated with accumulated impairment losses) at the beginning and end of the period.
- the detailed reconciliation of the carrying amount at the beginning and end of the period (showing, for example, additions, depreciation, impairment losses, revaluation information, foreign currency translation impacts, and so on).

U.S. GAAP require a company to disclose the depreciation expense for the period, the balances of assets, accumulated depreciation and a general description of the depreciation method(s) used.

Intangible Assets

IAS 38 provides a considerable set of disclosure requirements for intangible assets. For each class of intangibles, and distinguishing between internally generated and other assets, the financial statements must disclose:

- whether the useful lives are indefinite or finite and, if finite, the length of the useful lives or the amortization rates used.
- the amortization methods used for intangible assets with finite useful lives.
- the gross carrying amount and any accumulated amortization (aggregated with accumulated impairment losses) at the beginning and end of the period.
- the line item(s) of the statement of comprehensive income in which any amortization of intangible assets is included.
- detailed reconciliation of the carrying amount at the beginning and end of the period (showing, for example, additions, amortization, impairment losses, revaluation information, foreign currency translation impacts, and so on).

Under U.S. GAAP, a company is required to disclose the gross carrying amounts and accumulated amortization, the aggregated amortization expense for the period, and the estimated amortization expense for the next 5 years.

Impairment of Assets

As with most other standards, IAS 36 provides a long list of disclosure requirements. To begin with, for each class of assets, the financial statements must disclose:

- the amount of impairment losses and reversals recognized in profit or loss during the period and the line item(s) of the statement of comprehensive income in which those impairment losses and reversals are included.
- the amount of impairment losses and reversals on revalued assets recognized in other comprehensive income during the period.

U.S. GAAP require a company to disclose a description of the impaired asset, what caused impairment, the method of determining fair value, the amount of impairment loss, and where the loss is recognized on the financial statements.

8. Investment Property

```
   Investment property is defined as property that is owned (or, in some cases, leased ur
```

Under IFRS, companies are allowed to value investment properties using either a cost model or a fair value model.

- The cost model is identical to the cost model used for property, plant, and equipment (PP&E).
- The fair value model differs from the revaluation model used for PP&E. All changes in the fair value of investment property affect income.

Under U.S. GAAP, there is no specific definition of investment property. Investment properties are generally measured using the cost model.

Income Taxes

1. Key Terms

```
   The computation of income taxes poses problems in financial reporting. The major probl
```

Taxes are paid based on tax reporting, but from a financial reporting standpoint, the tax expense in the income statement (IS) is based on the matching principle and is computed on pretax accounting income. In order to achieve matching between taxes based on taxable income and taxes based on pretax income for accounting purposes, deferred tax entries are put through the accounting books.

The differences between the tax expense for tax and the accounting tax expense create deferred tax liabilities (credits) and deferred tax assets (debits or prepaid taxes).

Here are key terms based on *tax return*:

- **Taxable income**: Income subject to tax based on the tax code.

- **Taxes payable**: Tax return liability resulting from current period taxable income. (U.S.) SFAS 109 calls this "current tax expense or benefit."

- **Income tax paid**: Actual cash flow for income taxes, including payments (refunds) for other years.

- **Tax loss carry forward**: Tax return loss that can be used to reduce taxable income in future years.

- The **tax base** of an asset or liability is the amount attributed to that asset or liability for tax purposes.

Here are key terms based on *financial reporting*:

- **Pretax income** or **accounting profit**: Income before income tax expense.

- The **carrying amount** is the amount at which the asset or liability is valued according to accounting principles.

- **Income tax expense**: Expense resulting from current period pretax income; this includes taxes payable (from the tax return) and deferred income tax expense. It is reported in the income statement.

- **Deferred income tax expense**: Accrual of income tax expense expected to be paid (or recovered) in future years (difference between taxes payable and income tax expense). Under (U.S.) SAAS 109, this results from changes in deferred tax assets and liabilities.

- **Deferred tax asset**: Balance sheet item that results from a temporary excess of taxes payable over income taxes expense. It is expected to be recovered from future operations; it is not created if the excess is a permanent difference.

- **Deferred tax liability**: Balance sheet item that results from a temporary excess of income taxes expense over taxes payable. It is expected to result in future cash outflows; it is not created if the excess is a permanent difference.

- **Valuation allowance**: Reserve against deferred tax assets based on likelihood that those assets will be realized.

- **Timing difference**: The result of the tax return treatment (timing or amount) of a transaction that differs from the financial reporting treatment.

- **Temporary difference**: Difference between tax reporting and financial reporting that will affect taxable income when those differences reverse. This is similar to but slightly broader than timing difference. It also considers other events that result in differences between the tax bases of assets and liabilities and their carrying amounts in financial statements.

- **Permanent difference**: Differences between tax reporting and financial reporting that will not reverse in the future.

2. Deferred Tax Assets and Liabilities

```
Tax reporting and financial reporting are based on two different sets of assumptions.
```

In the U.S.:

- **Tax reporting** is based on the Internal Revenue Code (the tax code).
    - The modified cash basis of accounting is used in tax reporting to determine the periodic liability from currently taxable events.
    - Revenue and expense recognition methods used in tax reporting often differ from those used in financial reporting.

- **Financial reporting** is based on GAAP.
    - Accrual accounting is used in financial reporting to provide maximum information to allow evaluation of a firm's financial performance and cash flows.
    - Management is allowed to select revenue and expense recognition methods. A firm has a strong incentive to use methods that allow it to minimize taxable income.

Because of the differences between tax accounting and financial accounting, the financial statements may include tax liabilities or assets - allowances that have been made in the financial statements for taxes that have not yet been or have already been paid.

Deferred tax liabilities generally arise when tax relief is provided in advance of an accounting expense, or when income is accrued but not taxed until received. Deferred tax liabilities on an individual transaction are expected to be reversed when these liabilities are settled, causing *future cash outflows*.

A typical example is depreciation: a company uses the Accelerated Cost Recovery System for tax reporting but uses straight-line depreciation for financial reporting.

- Recall that taxes payable is calculated based on taxable income, and tax expense is calculated based on accounting profit.
- Lower depreciation expense in financial reporting results in accounting profit that is higher than taxable income, and tax expense that is higher than taxes payable.
- Deferred tax liabilities are thus created.

Deferred tax assets generally arise when tax relief is provided after an expense is deducted for accounting purposes. Deferred tax assets on an individual transaction are expected to be reversed when these assets are recovered, causing *future cash inflows*. Different treatments of warranty expenses in tax reporting and financial reporting are a common cause of deferred tax assets:

- For tax reporting, warranty expenses cannot be recognized until they have been incurred. For financial reporting, warranty expenses are recognized each year using accrual accounting, regardless of whether they are incurred or not.
- Lower warranty expense in tax reporting results in taxable income that is higher than accounting profit, and tax payable that is higher than tax expense.
- Deferred tax assets are thus created.

In the U.S., deferred tax assets/liabilities are classified on the balance sheet as current or non-current based on the classification of the underlying asset or liability. However, deferred tax assets/liabilities are always classified as non-current under IFRS.

A deferred tax item cannot be created if it is doubtful that the company will realize economic benefits in the future.

Example

A company purchases an asset for $1,000 at the beginning of Year 1. It depreciates the asset at 33% per annum (straight-line) for financial reporting. The tax depreciation is 50% per annum (straight-line). The pretax income and taxable income are $2,000 before depreciation for Year 1 to 3. Assume a tax rate of 30%.

The company will report the following for tax reporting:

Taxes payable is based on taxable income, not accounting profit.

Taxes payable = Taxable income x Tax rate

The company will report the following for financial reporting:

The deferred taxation can be computed in two ways:

- (Taxable income - accounting profit) x tax rate
- (Tax base - carrying amount) x tax rate

Take Year 1 as an example:

- (Taxable income - accounting profit) x tax rate = (1,500 - 1667) x 30% = -50
- (Tax base - carrying amount) x tax rate = (500 - 667) x 30% = -50

The textbook uses the second approach to calculate deferred tax assets/liabilities. At the end of each year, these are calculated by comparing the tax base and carrying amounts of the balance sheet items.

Note that tax expense can be broken down into taxes payable and deferred taxation.

- Year 1: taxes payable (450) + deferred tax liabilities (50) = Tax expense (500)
- Year 2: taxes payable (450) + deferred tax liabilities (50) = Tax expense (500)
- Year 3: taxes payable (600) + deferred tax liabilities (-100) = Tax expense (500)

In Year 1 and 2 the deferred tax is a liability in the balance sheet and an additional expense in the income statement. This occurs because the tax depreciation (500) is greater than the accounting depreciation charge (333). If the situation had been the other way around, a deferred tax asset would have resulted. This is the case when the pretax accounting profit is less than the taxable income.

In Year 3 the deferred tax is a reversal of the deferred tax liability on the balance sheet and a saving in the income statement. This occurs because the tax depreciation (0) is less than the depreciation charge for financial reporting (334).

As can be seen above, at the end of the three years there is no deferred tax on the balance sheet. This is why it is referred to as a temporary difference. The total tax (1,500) and total net income (3,500) are the same for tax and financial reporting. It is just the timing of their recognition that is different. At the end of the day, the tax collector and the accountant arrive at the same result.

Here are some important "tricks" you will need to know well for the exams.

A common question on the exam asks you to compute taxes payable or the tax expense. You may be given the pretax income, but need the taxable income figure to compute taxes payable. To adjust pretax income to taxable income and vice versa when assets have different depreciable lives for tax and accounting, remember the following:

- Pretax income + Accounting depreciation - Tax depreciation = Taxable Income, or
- Taxable Income + Tax depreciation - Accounting depreciation = Pretax Income, or
- Taxable Income + Temporary differences creating deferred tax liabilities - Temporary differences creating deferred tax assets = Pretax Income

Once you have determined the pretax income or taxable income, you can determine the taxes payable and tax expense as follows:

- Pretax income x tax rate = tax expense in the income statement
- Taxable income x tax rate = taxes payable
- Taxes payable + deferred tax effect on income statement = tax expense in the income statement

3. Determining the Tax Base of Assets and Liabilities

```
The tax base of an asset or liability is the amount attributed to that asset or liabil
```

The Tax Base of an Asset

An asset's tax base is the amount that will be deductible for tax purposes against any taxable economic benefits that will flow to an entity when it recovers the asset's carrying amount. It is the amount that would be tax deductible if the asset was sold on the balance sheet date.

For example, a firm has total accounts receivable of $100,000. At the end of the year, management recognized a specific doubtful debt division of $3,000 for financial reporting. However, provisions for doubtful debts are not allowed for tax purposes in the firm's tax jurisdiction. A tax deduction is received when the receivable is written off as bad debt.

The carrying amount of the accounts receivable becomes $97,000. The tax base of the asset still remains $100,000. The firm has a deductible temporary difference of $3,000. Management should recognize a deferred tax asset in respect to the deductible temporary difference.

If the economic benefit will not be taxable, the tax base of the asset will be equal to the carrying amount of the asset. An example is dividends receivable from a subsidiary. If it is not taxable, the tax base and the carrying amount of the dividends receivable are equal.

The Tax Base of a Liability

The tax base of a liability is its carrying amount, less any amount that will be deductible for tax purposes with respect to that liability in future periods.

- An unearned revenue item is treated as a liability for financial reporting but tax authorities often recognize it as taxable income. The tax base of such a liability is the carrying amount less any amount of the revenue that will not be taxable in the future. Examples are prepaid rent, prepaid subscriptions, etc.
- If an item has already been expensed, then its tax base and carrying amount are both zero. One example is interest paid on long-term loans.

Example

At the beginning of the year a firm received a lump sum of $5 million for rent from a lessee. The rent was for the use of an office building for the next 5 years. Local tax authorities require 70% of rent received in advance to be taxable income.

At the end of the year, $4 million should be treated as a liability for financial reporting purposes. That's the carrying amount. The tax base of the liability is $1.2 million (30% of $4 million) and $2.8 million should be treated as taxable income.

Changes in Income Tax Rates

When tax rates change, the deferred tax liability or asset has to be adjusted immediately to the new amount that is now expected, based upon the new expected tax consequences. The effect of this change in estimate will be included in the income from continuing operations.

The effect of an income tax rate increase:

- It raises deferred tax liabilities and thus increases tax expense.
- It raises deferred tax assets and thus decreases tax expense.
- If deferred tax liabilities exceed deferred tax assets, the net effect is to increase tax expense, and vice versa.

4. Temporary versus Permanent Differences

```
   Numerous items create differences between accounting profit and taxable income. These
```

Permanent differences do not cause deferred tax liabilities or assets. These occur if a revenue or expense item:

- is recognized for tax reporting but never for financial reporting, or
- is recognized for financial reporting but never for tax reporting.

Therefore, permanent differences result from revenues and expenses that are reportable on either tax returns or in financial statements but not both. Permanent differences arise because the tax code excludes certain revenues from taxation and limits the deductibility of certain expenses.

- In the U.S., for example, interest income on tax-exempt bonds, premiums paid on officer's life insurance, and amortization of goodwill (in some cases) are included in financial statements but are never reported on tax returns.
- Similarly, certain dividends are not fully taxed, and tax or statutory depletion may exceed cost-based depletion reported in the financial statements.
- Tax credits are another type of permanent difference. Such credits directly reduce taxes payable and are

different from tax deductions that reduce taxable income.

These differences are permanent because they will not reverse in future periods.

No deferred tax consequences are recognized for permanent differences; however, they result in a difference between the effective tax rate and the statutory tax rate that should be considered in the analysis of effective tax rates.

Example

A company owns a $50,000 municipal bond with a 4% coupon and has an effective tax rate of 50% and a statutory tax rate of 40%. Calculate the deferred tax created by this bond.

Solution

The bond does not result in deferred tax, as the difference it causes is a permanent difference that will not reverse. As a result, no deferred tax is recognized.

Temporary differences result in deferred tax liabilities or assets. Different depreciation methods or estimates used in tax reporting and financial reporting are a common cause of temporary differences.

There are two categories of temporary differences.

Taxable Temporary Differences (TTD)

- These will result in *taxable* amounts when an asset is recovered or a liability is settled.
- Hence, these result in deferred tax liabilities. This means the company will pay more tax in the future.

Items that give rise to taxable temporary differences are:

- Receivables resulting from sales.
- Prepaid expenses.
- Tax depreciation rates > accounting rates.
- Development costs capitalized and amortized.

Deductible Temporary Differences (DTD)

- These will result in *deductible* amounts when an asset is recovered or a liability is settled.
- Hence, these result in deferred tax assets. This means the company will pay less tax in the future.

Items that give rise to deductible temporary differences are:

- Accrued expenses.
- Unearned revenue.
- Tax depreciation rates < accounting rates.
- Tax losses.

5. Recognition and Measurement of Current and Deferred Tax

```
   Deferred tax assets and liabilities are re-assessed on each balance sheet date.<p> </p
```

- They are measured against the criteria of probable future economic benefits.
- The tax rate used to calculate them should be the one that is expected to apply when the asset is realized or liability settled.
- They are not discounted to present value although they are related to amounts at some future date.

Valuation Allowance

Deferred tax assets are reduced by a valuation allowance to amounts that are "more likely than not" to be realized, taking into account all available positive and negative evidence about the future. For determining whether deferred tax assets must be reduced by a valuation allowance, all available positive and negative evidence must be considered. Information concerning recent pretax accounting earnings generally is critical. For example, if a firm has been recording material cumulative losses recently, it will be hard to justify a conclusion that tax credits can be realized in the near future. This will be evidence supporting the use of a valuation allowance ("negative evidence"). It is not necessary to quantify positive evidence for the conclusion that a valuation allowance is not required unless significant negative evidence exists. Where both positive and negative evidence exist, judgment must be used in evaluating what evidence is more persuasive. More weight should be given to objectively verifiable evidence.

Recognition of Current and Deferred Tax Charged Directly to Equity

A firm's deferred tax liability during an accounting period represents the portion of income tax expense that has not been paid. Therefore, from a pure accounting perspective, deferred tax liabilities are an accounting liability. However, from a financial analyst's perspective, whether deferred tax liabilities should be considered liabilities or not depends on whether they will reverse in the future. If they will, resulting in a cash outflow, then they should be treated as liabilities. If not, then they should be treated as equity! As deferred tax liabilities are created by temporary differences, reversal of a deferred tax liability depends on the reversal of the temporary difference that created it.

Changes in a firm's operations or tax law may result in deferred taxes that are never paid or recovered. For example, the use of accelerated depreciation methods for tax reporting creates a temporary difference. Normally, when there is less depreciation in later years, the deferred tax liability created by more depreciation in earlier years will be reversed. However, for firms with high growth rates, increased investments in fixed assets result in ever-increasing new deferred tax liabilities, which replace the reversing one. That is, a firm's growth may continually generate deferred tax liabilities. In this case, the deferred taxes are unlikely to be paid. Therefore, for such high-growth firms, deferred tax liabilities will not reverse and should be treated as equity.

Deferred tax liabilities are recorded at their stated value. Even if deferred taxes are eventually paid, payments typically occur far in the future. The present value of those payments is considerably lower than the stated amounts. Thus, the deferred tax liability should be discounted at an appropriate interest rate and the difference should be treated as equity.

In some cases, financial statement depreciation understates the value of economic depreciation. Instead, the accelerated depreciation in tax reporting is a better measure. Examples of such cases include equipment obsolescence due to technology innovation and rising price levels. Deferred tax liabilities are neither liabilities nor equity if they are not expected to reverse, and should be ignored by financial analysts.

- They are not liabilities since they will not reverse.
- They are not equity since adding the entire tax liabilities to equity overstates the value of the firm.

In practice, the financial analyst must decide on the appropriate treatment of deferred taxes on a case-by-case basis.

Treatment of Operating Losses

Tax losses can be carried back and applied to prior years to obtain refunds of taxes paid. They can also be carried forward to future periods. Because the realization of tax loss carry forward depends on future taxable income, the expected benefits are recognized as deferred tax assets. Such assets are recognized in full but a valuation allowance may be required if recoverability is unlikely.

6. Presentation and Disclosure

```
   The best way to approach this subject is to study the example presented in the reading
```

7. Comparison of IFRS and U.S. GAAP

<b>Similarities</b><p> </p>

FAS 109 *Accounting for Income Taxes* and IAS 12 *Income Taxes* provide the guidance for income tax accounting under U.S. GAAP and IFRS, respectively. Both pronouncements require entities to account for both current tax effects and expected future tax consequences of events that have been recognized (that is, deferred taxes) using an **asset and liability approach**. Further, deferred taxes for temporary differences arising from non-deductible goodwill are not recorded under either approach and the tax effects of items directly accounted for as equity during the current year are also allocated directly to equity. Finally, neither principle permits the discounting of deferred taxes.

Significant Differences and Convergence

Below we discuss the significant differences in the current literature.

Tax basis:

- U.S. GAAP: Tax basis is a question of fact under the tax law. For most assets and liabilities there is no dispute on this amount; however, when uncertainty exists, it is determined in accordance with FIN 48 *Accounting for Uncertainty in Income Taxes*.
- IFRS: Tax basis is generally the amount deductible or taxable for tax purposes. The manner in which management intends to settle or recover a carrying amount affects the determination of tax basis.

Uncertain tax positions:

- U.S. GAAP: FIN 48 requires a two-step process, separating recognition from measurement. A benefit is recognized when it is "more likely than not" to be sustained based on the technical merits of the position. The amount of benefit to be recognized is based on the largest amount of tax benefit that is greater than 50% likely of being realized upon ultimate settlement. Detection risk is precluded from being considered in the analysis.
- IFRS: There is no specific guidance; IAS 12 indicates tax assets/liabilities should be measured at the amount expected to be paid. In practice, the recognition principles on provisions and contingencies in IAS 37 are frequently applied. Practice varies regarding consideration of detection risk in the analysis.

Initial recognition exemption:

- U.S. GAAP: No similar exemption for non-recognition of deferred tax effects for certain assets or liabilities.
- IFRS: Deferred tax effects arising from the initial recognition of an asset or liability are not recognized when the amounts did not arise from a business combination and, upon occurrence, the transaction affects neither accounting nor taxable profit (for example, acquisition of nondeductible assets).

Recognition of deferred tax assets:

- U.S. GAAP: Recognized in full (except for certain outside basis differences), but valuation allowance reduces assets to the amount that is more likely than not to be realized.
- IFRS: Amounts are recognized only to the extent it is probable (similar to "more likely than not" under U.S. GAAP) that they will be realized.

Calculation of deferred asset or liability:

- U.S. GAAP: Enacted tax rates must be used.
- IFRS: Enacted or "substantively enacted" tax rates (as of the balance sheet date) must be used.

Classification of deferred tax assets and liabilities in balance sheet:

- U.S. GAAP: Current or non-current classification, based on the nature of the related asset or liability, is required.
- IFRS: All amounts are classified as non-current in the balance sheet.

Recognition of deferred tax liabilities from investments in subsidiaries or joint ventures (JVs) (often referred to as outside basis differences):

- U.S. GAAP: Recognition is not required for investment in foreign subsidiary or corporate JVs that are essentially permanent in duration, unless it becomes apparent that the difference will reverse in the foreseeable future.
- IFRS: Recognition is required unless the reporting entity has control over the timing of the reversal of the temporary difference and it is probable ("more likely than not") that the difference will not reverse in the foreseeable future.

Non-current (Long-term) Liabilities

1. Accounting for Bond Issuance, Bond Amortization, Interest Expense, and Interest Payments

```
Debt is classified as short-term (ST) and long-term (LT).<p> </p><ul class="notes">
```

- Current liabilities result from both operating and financing activities.
  - Those caused by operating activities include accounts payable and advances from customers. Operating and trade debt is reported at the expected (undiscounted) cash flow and is an important exception to the rule that liabilities are recorded at present value. Note that advances from customers are the consequence of operating decisions, the result of normal activity. They should be distinguished from other payables when analyzing a firm's liquidity. Advances are a prediction of future revenues rather than cash outflows.
  - Those resulting from financing activities include short-term (ST) debt and the current portion of long-term (LT) debt. They are recorded at present value. Note that the current portion of LT debt is the consequence of financing activity and indicates a need for cash or refinancing. A shift from operating to financing indicates the beginning of liquidity problems, and inability to repay ST credit is a sign of financial distress.
- Long-term debt results from financing activities. It may be obtained from many sources that may differ in interest and principal payments. Some claims are below or subordinated to others while other claims may be senior or have priority. Whatever the different payment terms are, there are two basic principles:
  - Debt equals present value of the future interest and principal payments. For book values the discount rate is the rate when debt was incurred. For market values the discount rate is the current rate.
  - Interest expense is the amount paid to the creditor in excess of the amount received. Though the total to be paid is known, allocation to specific time periods may be uncertain. The coupon rate is just the *stated* cash interest rate.

Below we will focus on debt resulting from financing activities.

Bond premiums and discounts:

- A **bond premium** represents the amount over the face value of the bond that the issuer never has to return to the bondholders. In effect it *reduces* the *higher-than-market interest rate* that the issuer is paying on the bond. It must be allocated over the life of the bond as a *reduction* of interest expense each period.
- A **bond discount** represents the amount in excess of the issue price that must be paid by the issuer at the time of maturity. In effect it *increases* the *lower-than-market interest rate* the issuer is paying on the bond. It must be allocated over the life of the bond as an *increase* of interest expense each period.

At the time of issuance, the firm receives proceeds from issuing the bond. A bond payable is valued at the present value of its future cash flows (periodic coupon payments and principal repayment at maturity). These cash flows are discounted at *the market rate of interest at issuance*. Therefore, the value of the bond depends on the market rate of interest. For example, if the market rate of interest is higher than the coupon rate, the bond value will be less than its face value, and the bond is issued at a discount.

- Balance sheet. Initial liability is the amount paid to the issuer by the lender. The amount may not equal to the face value of the bond.

  - Issued at par on Interest Date:
    If $800,000 of bonds were issued on January 1, 2015 at 100, the issuance would be recorded as follows:

  - Issued at Discount on Interest Date:
    If $800,000 of bonds were issued on January 1, 2015 at 97, the issuance would be recorded as follows:

  - Issued at Premium on Interest Date:
    If $800,000 of bonds were issued on January 1, 2015 at 103, the issuance would be recorded as follows:

- Cash flow statement. Cash flow from financing (CFF) increases by the amount received.

At the end of each semi-annual payment period, the firm makes a coupon payment:

- Income statement. Interest expense is reported here. The effective interest rate is the market rate at the time of issuance, and the interest expense is market rate multiplied by the bond liability at the beginning of this six-month period.

- Cash flow statement. Cash flow from operations (CFO) decreases by the coupon payment. The coupon rate and face value are used to calculate actual cash flows only.

- Balance sheet. The bond liability is adjusted if necessary. Liability over time is a function of (1) initial liability and the relationship of (2) interest expense to (3) the actual cash payments. That is, the difference between interest expense and coupon payment represents the change in bond liability during this period: change in bond liability = interest expense - coupon payment. The ending bond liability = beginning bond liability + change in bond liability.

  The bond premium or discount is amortized over the life of the bond by what is known as the **interest method**. This results in a constant rate of interest (not a constant interest expense) over the life of the bond. Bond interest expense is increased by amortization of a discount and decreased by amortization of a premium.

  - If the bond is issued at a premium, interest expense is always lower than coupon payment, and decreases over time. In this case the interest expense is only one component of the coupon payment. The rest of the coupon payment is used to amortize the bond's premium.
  - If the bond is issued at a discount, interest expense is always higher than coupon payment, and increases over time. In this case the interest expense has two components: the coupon payment and amortization amount of the bond's discount.
  - If the bond is issued at par, interest expense equals coupon payment.

  At any point in time the liability on the balance sheet will equal the present value of the remaining cash flow payments to the creditor discounted at the effective market interest rate.

At the maturity date, the firm repays the face value of the bond. The treatment and effects of the last coupon payment are the same as those shown above.

- Balance sheet: the bond liability is reduced by the face value.
- Cash flow statement: similar to the treatment of initial cash received, the final face value payment is treated as CFF.

Total interest expense is equal to amounts paid by the issuer to the creditor in excess of the amount received.

Summary of the Effective Interest Method

- Bond interest expense is computed first by multiplying the carrying value of the bonds at the beginning of the period by the effective interest rate:

  Interest Expense = Beginning Carrying Value x Market Rate of Interest

- The bond discount or premium amortization is then determined by comparing the bond interest expense with the interest to be paid. Note that Interest Payment = Face Value x Coupon Rate x 1/2 (assume semi-annual coupon payment).

- The carrying value of the bond at the end of the period = Beginning Carrying Value - Amortization of Bond Premium (or + Amortization of Bond Discount).

This method produces a periodic interest expense equal to a constant percentage of the carrying value of the bonds. Since the percentage is the effective rate of interest incurred by the borrower at the time of issuance, the effective interest method results in a better matching of expense with revenues than the straight-line method.

Example

Evermaster Corporation issued $100,000 of 8% term bonds on January 1, 2015, due on January 1, 2020, with interest payable each July 1 and January 1. Since investors required an effective interest rate of 10%, they paid $92,278 for the $100,000 of bonds, creating a $7722 discount.

1. To record the issuance:

2. To record the first interest payment of $4,000 on July 1, 2015:

Note: 92,278 x 0.10 x 0.5 = $4,614. Now the discount balance is 7,722 - 614 = $7,108.

1. To record the second interest payment of $4,000 on January 1, 2016:

Note: (100,000 - 7,108) x 0.10 x 0.5 = $4,645. Now the discount balance is 7,108 - 645 = $6,463.

And so on.

- The carrying value of a premium (discount) bond decreases (increases) over time.
- At the maturity date, the carrying value of both a premium bond and a discount bond equals the face value.
- The interest expense of a discount (premium) bond increases (decreases) over time due to the increasing (decreasing) carrying value.

Zero-Coupon Bond

Zero-coupon bond has no periodic interest payments and is issued at a large discount from par. The proceeds at issuance equal the present value of the face value, discounted at the market value of interest at issuance. Repayment at maturity includes all the (implied) interest expense (equal to face value minus the proceeds) from the time of issuance: Total Implied Interest = Par Value - Proceeds Received.

In essence, zero-coupon bonds are a special type of discount bonds. Therefore, their effects on financial statements are similar to those of discount bonds.

- The *interest expense* on a zero-coupon bond never reduces operating cash flow. Reported CFO is systematically "overstated" when a zero-coupon (or deep-discount) bond is issued, while CFF is understated by the amortization amount of the discount and should be adjusted accordingly.

- Unlike discount bonds (whose reported CFO is reduced by the coupon payments), they make no *coupon payments* so they have no effect on reported CFO.
- Solvency ratios, such as cash-basis interest coverage, are improved relative to the issuance of par bonds. The cash eventually required to repay the obligations may become a significant burden.

For example, assume a market interest rate of 5%, a $100,000 face value zero-coupon bond payable in three years will be issued at $86,229.68 (semi-annual compounding). This is computed by pressing the following calculator keys: 100000 FV; 2.5I/Y; 6N; CPT PV.

The journal entry (of the issuer) to record this issue will be:

Bank (balance sheet): $86,229.68 (debit)

Bond liability (balance sheet): $86,229.68 (credit).

When the bond is paid at maturity, the repayment of $100,000 includes $13,770.32 of interest. The major difference between a zero-coupon and a par value bond is that the interest of $13,770.32 is never reported as a cash flow from operations for a zero-coupon bond.

This is because the full $100,000 is reported as a cash flow from financing.

2. Accounting for Bonds at Fair Value

```
   The market value of debt is dependent on future interest rates. This is because the ma
```

Debt reported on the balance sheet is equal to the present value of future cash payments discounted at the market value rate on the date of issuance. Increases (decreases) in the current market value rate decrease (increase) the market value of the debt. This economic gain or loss is NOT reflected in either the income statement or balance sheet. For some analytic purposes the market value of a company's debt may be more relevant than its book value.

A bond that pays a fixed interest rate is more susceptible to changes in interest rates than a bond that pays a variable interest rate.

- As the interest rates change, the bond that pays the variable interest rate will not really change in value, as the future cash flows will fluctuate with the change in the interest rate curve. Therefore, *for adjustable-rate debt, book value approximates market value and no adjustment is required.*
- A bond that pays a fixed interest rate will change in value because the discount rate of each cash flow fluctuates as the market yield curve fluctuates. This is especially true of zero-coupon and other discount debt, due to their longer duration relative to debt of the same maturity issued at par.

It is important, if a company has issued debt that pays a fixed interest rate, to consider whether the market value of the debt is different from the value of the debt in the company's books.

Consider what would happen to the value of the company's debt if market interest rates change:

- When interest rates increase, the market value of fixed-rate debt decreases.
- When interest rates decrease, the market value of fixed-rate debt increases.
- For floating-rate debt, increases or decreases in interest rates do not have an impact on market value, as future cash flows will change along with the changes in interest rates.

Consider how this will affect the financial statements of company that has issued fixed-rate debt (it will have no effect on companies that issue floating-rate debt):

- When interest rates increase, and the market value of the debt decreases, this will decrease the company's liabilities. As the liabilities will be lower, the debt-to-equity ratio will be more favorable.

- When interest rates decrease, and the market value of the debt increases, this will increase company's liabilities. As the liabilities will be higher, the debt-to-equity ratio will be less favorable.

The Market Value of Debt

The market value of a company's debt is either the market price, if it is traded, or the present value of the future cash flows. The discount rate that is used is a risk-free rate, plus an appropriate spread for the risk of the particular company.

Typical Exam Question

A firm has variable-rate long-term debt outstanding. All other factors being equal, what effect will a rise in interest have on the firm's debt-to-equity ratio and net income?

Interest has increased, which means income decreases. Retained earnings are therefore lower and the debt-to-equity ratio will increase.

3. Derecognition of Debt

```
   Early retirements of debt may occur because a company has generated sufficient cash re
```

Whether debt is being retired or refinanced in some other way, accounting rules dictate that the retired debt be removed from the books and the difference between the debt's net carrying value and the funds paid to retire the debt be recognized as a gain or loss.

Example

Assume that Cabano Corporation is retiring $200,000 face value of its 6% bonds payable. The last semi-annual interest payment occurred on April 30 and the bonds are being retired on June 30, 2010. The unamortized discount on the bonds on April 30, 2010, was $6,000, and there was a 5-year remaining life on the bonds as of that date. Further, Cabano is paying $210,000, plus accrued interest, to retire the bonds; this "early call" price was stipulated in the original bond covenant.

The first step to account for this bond retirement is to bring the accounting for interest up to date:

Interest Expense (debit): 2,200

Discount on Bonds Payable (credit): 200

Interest Payable (credit): 2,000

Then, the actual bond retirement can be recorded, with the difference between the up-to-date carrying value and the funds utilized recorded as a loss (debit) or gain (credit).

Bonds Payable (debit): 200,000

Interest Payable (debit): 2,000

Loss on Bond Retirement (debit): 15,800

Discount on Bonds Payable (credit): 5,800

Cash (credit): 212,000

Notice that Cabano's loss relates to the fact that it took a lot more cash ($210,000) to pay off the debt than was the debt's carrying value ($200,000 - $5,800 = $194,200).

4. Debt Covenants

```
<b>Bond covenants</b> are a bond issuer's enforceable promises to perform or refrain f
```

- **Affirmative covenants** set forth certain actions that issuers must take, such as:
  - paying interest and principal on a timely basis.
  - paying taxes and other claims when due.
  - keeping assets in good conditions and in working order.
  - submitting periodic reports to a trustee so the trustee can evaluate the issuer's compliance with the indenture.

- **Negative covenants** set forth certain limitations and restrictions on the borrower's activities, such as limitations on:
  - the borrower's ability to incur additional debt or other liabilities.
  - dividend payments and stock repurchases.
  - production and investment (mergers and acquisitions, sales and leaseback, or outright disposal of certain assets).
  - payoff patterns (sinking fund requirements and the priorities of claims on assets).

In addition to direct restrictions on activities, covenants may require maintenance of certain levels of such accounting-based financial variables as stockholders' equity (or retained earnings), working capital, interest coverage, and debt-to-equity ratios.

Auditors and management must certify that the firm has not violated the covenants. If any covenant is violated, the firm is in technical default of its leading agreement, and the creditor can demand repayment of the debt after the stated grace period.

Example

The following exhibit contains information from NorAm Energy's 1994 Annual Report regarding debt covenants imposed by its creditors. The covenants restrict borrowings and dividend payments.

Note 5: Restrictions on Stockholders' Equity and Debt:

Under the provision of the Company's revolving credit facility as described in Note 3, and under similar provisions in certain of the Company's other financial arrangements, the Company's total debt capacity is limited and it is required to maintain a minimum level of stockholders' equity.

- The required minimum level of stockholders' equity was initially set at $650 million at December 31, 1993, increasing annually thereafter by (1) 50% of positive consolidated net income and (2) 50% of the proceeds (in excess of the first $50 million) of any incremental equity offering made after June 30, 1994.
- The Company's total debt is limited to $2,055 million.

Based on these restrictions, the Company has incremental debt issuance and dividend capacity of $321.2 million and $43.3 million, respectively, on December 31, 1994. The Company's revolving credit facility also contains a provision that limits the Company's ability to reacquire, retire, or otherwise prepay its long-term debt prior to its maturity, to a total of $100 million.

The exhibit states that as of December 31, 1994 the company has a dividend capacity equal to $43.3 million. This amount was computed after reflecting the annual dividend of $42 million declared in 1994.

Question 1. How was the figure $43.3 million computed?

Question 2. State whether the debt covenants restrict NorAm's ability to maintain its annual dividend through 1998. Justify your answer by preparing a schedule for the years 1995 - 1998 showing NorAm's expected and minimum shareholders' equity given current income and dividend levels.

Without any increase in income, the current dividend can be maintained for only two years:

As the table above shows, in 1997 the minimum equity requirement will be violated.

Question 3. Compare the level of income that would be required to maintain current dividend levels through 1998.

To maintain dividend payments at the 1994 level through 1998, income would have to increase. The required income for 1997 and 1998 is $69.4 million and $84 million respectively. These amounts result in the following table for those years:

These amounts can be calculated as follows:

1997: Increase = 2 x shortfall in equity = 2 x (746.0 - 735.3) = 21.4 million

1998: Income must equal $84 million, twice the dividend, to maintain equity at the required level.

Question 4. In 1995, NorAm approached its shareholders with a proposal to issue new shares. What are possible reasons the company was motivated to make this proposal? Would you, as a shareholder, have supported the proposal?

The answer would depend on the shareholder' view of the market price of NorAm's shares. Issuance of new shares to maintain the current dividend makes no sense given finance theory, which states that the two are equivalent. In an imperfect market, however, NorAm's shares may have been fully valued but the shareholder may not have wished to sell and incur capital gains taxes. If NorAm had attractive investment opportunities not reflected in its stock price, issuing new shares to increase the firm's borrowing capacity would be desirable.

5. Presentation and Disclosure of Long-Term Debt

```
    <b>Balance Sheet Disclosure of Bonds</b><p> </p>
```

Bonds payable and unamortized discounts or premiums are typically shown on the balance sheet as *long-term liabilities*. Bond discount is reported as a direct deduction from - and bond premium as a direct addition to - the face value of the bond. If a bond issue will mature within a year, it should be reported as a *current liability* if the issue will be retired using current assets. However, the issue should continue to be reported as a long-term liability if it is to be replaced with another bond issue, converted into stock, or paid off with noncurrent assets.

Financial statement footnotes provide more information on the type and nature of a company's debt: stated and effective interest rates, maturity dates, restrictions imposed by creditors, and collateral pledged (if any). An MD&A also provides other information on a company's capital resources, such as debt financing and off-balance-sheet financing.

For bonds issued at a premium or discount, reporting coupon payments as cash outflow from operations is inappropriate. For example, if a bond is sold at a premium, part of the coupon payment is used to amortize the premium and reduce the principal, and therefore should be treated as a financing cash outflow. As a result, CFO is understated and CFF is overstated by the amortization amount of the bond's premium.

Debt with Equity Features

A company can issue a bond that is convertible into shares or a bond with a warrant attached. The reason for doing this is to add a sweetener to a bond issue. Because of the benefit to the investor (the ability to obtain shares of the company), the issuer is able to issue the bonds at a lower interest rate than would be the case for a straight bond.

6. Advantages of Leasing

```
    A lease in its most basic form is the renting of some sort of property. For instance,
```

When purchasing an asset, the buyer acquires ownership of the asset and all benefits and risks embodied in the asset. A firm may acquire use of an asset, including some or all of its benefits and risks, for specified periods of time by making payments through a contractual arrangement called a **lease**. Using leases, a firm can avoid tying up too much capital in fixed asset investment.

General incentives for leasing:

- Lessee ownership is closely held; risk reduction is important.
- Lessor has market power and can generate higher profits by leasing the asset than selling it.
- Asset is not specialized to the firm.
- Asset's value is not sensitive to use or abuse (as owner takes better care of asset than lessee).
- **Tax incentives**. If the lessee is in a low tax bracket and the lessor in a high tax bracket, there are tax benefits to structuring the lease as an operating lease. The reason for this is that by leasing the asset the lessor can retain greater tax benefits from owning the asset (such as accelerated depreciation methods for tax purposes).

Factors favoring an operating lease:

- Period of use is short relative to the overall life of the asset.
- Lessor has a comparative advantage in reselling the asset.
- Corporate bond covenants contain specific covenants relating to financial policies that firms must follow. An operating lease results in lower leverage ratios and higher asset turnover ratios.
- Management compensation contracts contain provisions expressing compensation as a function of returns on invested capital.

7. Lease Classification

```
    <b>Operating leases</b> (<b>OL</b>) allow the lessee to use the property for only a po
```

- The lessee reports only the required lease payments as they are made. There is no balance sheet recognition of the property.
- All of the risks and rewards of ownership remain with the lessor. For the lessor:

  - Payments received are recognized as income.
  - The property remains on the balance sheet and is depreciated over time.

In an operating lease the lessee is not seen as becoming the owner of the asset. In a capital lease the lessee is seen as becoming the owner.

Capital leases (**CL**) involve effective transfer of all risk and benefits of property to the lessee. For accounting purposes the transaction is treated as though the lessor has granted the lessee a loan to purchase the asset. The lessee recognizes the asset as a loan on its balance sheet and treats the lease payments as part of interest expense and part of repayment of the principal on the loan. The lessee then also recognizes depreciation on the asset.

Benefits

Generally, *lessees* prefer leases to be classified as *operating leases*. They do not have to recognize the asset and loan on their balance sheet, although they still receive all of the benefit of using the asset. Therefore, operating leases result in higher profitability ratios and reduce reported leverage for lessees.

Generally, *lessors* prefer leases to be classified as *capital leases*. This allows them to potentially recognize a profit on the sale of the asset, though the substance of the transaction is similar to installment sales or financing. This also allows them to be able to derecognize the asset on their balance sheet.

Capital or Operating Leases?

To determine whether a lease is a capital or operating lease, you need to follow the rules in (U.S.) SFAS 13 or IAS 17. Even if a lease contract states that it is an operating lease, if the lease meets the requirements for a capital lease it needs to be accounted for as such.

Under U.S. GAAP, a lease meeting *any of the following criteria at inception* must be classified as a capital lease by the lessee:

- The lease transfers ownership of the property to the lessee at the end of the term.
- The lease contains a bargain purchase option that the lessee may purchase the leased asset for a price that is significantly below its fair market value at the end of the lease term.
- The lease term is equal to 75% or more of the asset's economic life.
- The present value of the minimum lease payments (MLPs) equals or exceeds 90% of the asset's fair market value, using the lessee's incremental borrowing rate or the implicit rate of the lease.

The provisions of IAS 17 are less precise. That standard defines a finance lease (the IFRS term for a capital lease) as one that transfers substantially all the risks and rewards incident to ownership of an asset. Title may or may not eventually be transferred.

If a lease does not meet any of the four criteria above, it is classified as an operating lease by the lessee. SFAS 13 mandates the use of the straight-line method of recognizing periodic payments unless another systematic basis provides a better approach. As a result, for leases with rising rental payments, lease expense and cash flow will not be identical.

From a lessor's perspective, a lease is classified as a capital lease if, at its inception, it meets any of the above four criteria and *both* of the following two revenue recognition criteria:

- Collectability of lease payments is reasonably predictable.
- There are no significant uncertainties regarding the amount of un-reimbursed costs yet to be incurred by the lessor. That is, future costs are reasonably predictable, so the lessor's performance is substantially complete.

A lease that does not meet any of the four criteria is classified by the lessor as an operating lease.

Example

A company leases a machine with a six-year life and a cost of $10,000 for 4 years with annual payments of $2,500 due at the end of each year. The lessee can borrow at 4% per annum. Calculate if the lease should be treated as an operating lease or capital lease.

1. Does the lease transfer ownership of the property to the lessee at the end of the lease term? No, the asset is only leased for 4 years out of its possible 6-year useful life and the asset is given back to the lessor. This requirement is not met.

2. Does the lease contain a bargain purchase option? Not from the information available in the question. Thus this requirement is not met.

3. Is the lease term equal to 75% or more of the estimated economic life of the leased property? The lease is for 4 of the 6-year useful life, which equates to 66.6% (4/6) of the estimated economic life. This requirement is not met.

4. Does the present value of the minimum lease payments (MLPs) equal or exceed 90% of the fair value of leased property to the lessor? The cost of the asset is $10,000. In order to determine the present value of the minimum lease payments you need to know the borrowing cost of the lessee. If you assume that the lessee is able to borrow at 4% for the 4 years, the present value is $9,074. This is 90.7% (9074/10000) of the fair value of the leased asset. The present value of the minimum lease payment exceeds 90% of the fair value of the leased asset to the lessor and thus qualifies for recognition as a capital lease.

The lease is recognized as a capital lease by the lessee.

8. Accounting and Reporting by the Lessee

Suppose a non-cancelable lease begins on Dec. 31, 20X0 with annual lease payments of $

Operating Lease for a Lessee

- No entry is made at the inception of the lease.
- Over the life of the lease, only the annual rental expense (an operating expense) of $10,000 will be charged to income and CFO.

Operating Lease for a Lessor

- The leased asset is reported on the lessor's balance sheet, because the lessor still retains ownership of the asset. Accordingly, the lessor will depreciate the asset during the term of lease.
- Over the life of the lease, periodic lease payments from the lessee are reported as rental revenue on the income statement. They are recorded as operating cash flows.

Capital Lease for a Lessee

- At the inception of the lease, an asset (leasehold asset) and liability (leasehold liability) equal to the present value of the lease payments, $31,700, is recognized. The **implicit interest rate of the lease** is the discount rate that sets the aggregate present value of lease payments to be equal to the fair market value of the leased asset.
- Over the life of the lease:
  - The annual rental expense of $10,000 will be allocated between interest and principal payments on the $31,700 leasehold liability according to the amortization schedule.
    - Interest payment = Beginning lease obligation x implicit discount rate. It is accounted for as interest expense, which reduces income and cash flow from operating activities.
    - Principal repayment = Lease payment - Interest payment. It reduces lease liability and cash flows from financing activities. It has no effect on the income statement.
    - Ending lease obligation = Beginning lease obligation - principal repayment
  - The $31,700 cost of the leasehold asset is charged to operating expense (annual depreciation is $7,925) using the straight-line method over the term of the lease. Depreciation is charged to income but has no effect on cash flows.

Amortization schedule:

Financial effects:

- <u>Balance sheet effects</u>
  - When a lease is a CL, gross ($31,700) and net amounts are reported at each BS date. CL increases asset balances, resulting in lower asset turnover and lower ROA than OL classification.
  - The current liability is the principal portion of the first lease payment. Non-current liability is the rest of the principal. At the lease's inception, leased assets and liabilities (A&L) are equal at $31,700.
  - The most important effect of a CL is the impact on leverage ratios, which results in an increase in debt to equity and other leverage ratios. As lease obligations aren't recognized for OL, leverage ratios are understated.
- <u>Income statement (IS) effects</u>

  An OL charges constant rental payments to expense as accrued, whereas a CL recognizes and apportions depreciation and interest expense over the term of the lease.

    - Operating income:

      Capitalization results in a higher EBIT, as the straight-line depreciation expense of $7,925 is lower than the OL rental expense of $10,000.

    - Total expense & net income:

      CL interest expense falls over time and depreciation expense is constant (straight-line depreciation) or declining (accelerated depreciation method). Total expense for CL declines over the lease term. Initially, it's higher than OL expense but over time it becomes lower. Tax expense and net income for an OL are constant over time. Tax expense and net income for a CL increase; for a CL, one also reports an accumulating deferred tax expense.

      Compared to a CL, firms using an OL generally report higher profits, interest coverage, and ROA. Lease expense (for CL=$11,095) exceeds lease payments (for OL=$10,000), so there will be a deferred tax credit. The deferred tax amount increases until the lease expense is less than the lease payments, and then the account declines and is eliminated by the end of the lease.

      No deferred tax is reported for OL, since the amount deductible for taxes and reported lease expense are always the same. Total (interest and depreciation) expense for a CL must equal total rental expense for an OL, over the life of the lease. Net income is not affected by CL but CL reports lower income earlier in the lease term and higher income later.

- Cash flows effects

  Under an OL, all cash flows are operating and there is an operating cash outflow of $10,000 per year. Annual payments of $10,000 create a tax benefit of $3,500 per year, which is deductible regardless of the lease method used. CL produces operating cash flows (CFO) and financial cash flows (FCF). The $10,000 paid under the CL is allocated between interest and amortization of the lease obligation (reported as cash from financing).

  For a CL, as interest expense declines over the term of lease and an increasing portion is allocated to the lease obligation, the difference in CFO increases over the lease. CL therefore decreases operating cash outflow while increasing financing cash outflow.

  Note that under IFRS the interest expense portion of the lease payment can be treated as either operating or financing cash outflow, but under U.S. GAAP it is treated as operating cash outflow.

Summary (For a Lessee):

All other factors remaining equal, firms reporting operating leases will report better performance because:

- Their balance sheet will report less debt.
- They will report higher profits, which appear to be generated by a relatively smaller investment in assets.

9. Accounting and Reporting by the Lessor

```
   How a lessee accounts for a lease has been described in the previous subject. In this
```

Under IFRS, leases are classified as **finance leases** when substantially all the risks and rewards of legal ownership are transferred to the lessee.

Under U.S. GAAP, for a lessor, the lease must be capitalized if it meets *any* one of the four criteria specified for capitalization by the lessee *and both* of the following revenue recognition criteria:

- Collectability of the minimum lease payments is reasonably predictable.
- There are no uncertainties regarding the amount of unreimbursable cost yet to be incurred by the lessor.

Leases not meeting these criteria must be reported as operating leases.

Under U.S. GAAP there are two alternative types of **capital leases** when dealing with lessors: a sales-type lease and a direct financing lease. IFRS does not distinguish between them. The accounting is the same for IFRS and U.S. GAAP.

1. **Sales-Type Lease**

   Sale-leaseback (**S-L**) transactions are sales of property by the owner, who then leases the property back from the buyer-lessee. The seller (lessor) leases its own property, rather than selling it outright. Such leases involve two transactions:

- Selling the property at the time the lease is initiated.
- Providing financing to the lessee.

At the inception of the lease, a manufacturer treats the transaction as if it sold the asset in exchange for an investment in a capital lease. It recognizes a gross profit from the sale of the asset.

- Sales = the lower of the present value of minimum lease payments or the fair value of the asset. The total lease payments (un-discounted) are known as the lessor's **Gross Investment in Lease**.
- Cost of goods sold = cost of leased asset - present value of residual value
- Gross profit (manufacturer's profit) = sales - cost of goods sold.
- The balance sheet also records an asset: **net investment in lease** = present value of lease payment + present value of residual value. The difference between gross investment in a lease and net investment in a lease is the unearned interest income, which is the financing income component of the manufacturer's total profit.

Periodic Transactions

- Each year over the life of the lease, interest income is recognized using the following formula: interest income = the year's beginning value of net investment in lease x implicit interest rate of the lease. *It is a cash flow from operations under U.S. GAAP and either an operating or investing cash flow under IFRS.*
- The difference between lease payment and interest income represents the portion of Net Investment in Lease recovered during the year. Recovery in Net Investment in Lease is an investment cash flow.
- The Net Investment in Lease at the end of the year is: the year's beginning value of Net Investment in Lease - Net Investment Recovery.

1. **Direct Financing Lease**

In a **direct financing lease**, a leasing company purchases a property from a manufacturer and then leases the equipment to the lessee. The distinction between a sales-type lease and a direct financing lease is the presence/absence of a manufacturer's or dealer's profit. In a direct financing lease, the cost of the leased asset equals its market value, so only financing income is involved. In a sales-type lease, the cost of the leased asset is less than its market value (the present market value of lease payments), creating a manufacturer's or dealer's profit in addition to financing income.

As in a direct financing lease, the lessor's original cost or the carrying value (prior to the lease) of the asset approximates the market value of the leased asset (the present value of the MLPs); such leases are pure financing transactions and financial reporting for direct financing leases reflects this fact. No sale is recognized at the inception of the lease and there is no manufacturing or dealer profit. Only financing income is reported.

Effects of Operating and Capital (Finance) Leases on a Lessor's Financial Statements

Balance Sheet:

- Operating Lease: The book value of the asset is reported on the balance sheet as a long-term asset, net of accumulated depreciation.

- Capital Lease: Lease receivable, instead of the asset, is reported on the balance sheet, and is amortized over the life of the lease.
- For a sales-type lease, the gain from the sale of the asset recognized at the inception of the lease increases shareholders' equity. Therefore, a sales-type lease results in higher total assets and higher shareholders' equity.

Income Statement:

- Operating Lease: At the inception of the lease, no income is recognized. Over the life of the lease, income tends to be constant if straight-line depreciation is used.
- Capital Lease: Over the life of the lease, interest income tends to decline over time.
- At the inception of the lease, the gain from the sale of the lease asset is recognized for a sales-type lease. The sales-type lease reports substantially higher income at the inception of the lease, thus recognizing income earlier than an operating lease. However, total income over the life of the lease is the same for both. Although IFRS does not have a provision for sales-type leases, the accounting treatment is the same if there's a profit at the inception of the lease.

Cash Flow Statement:

- Operating Lease: At the inception of the lease, no cash flow occurs. Over the term of the lease, the entire lease payment is reported as an operating cash inflow.
- Capital Lease: Over the term of the lease, the lease payment from the lessee is allocated to interest income (an operating cash inflow under U.S. GAAP, and either an operating or an investing cash flow under IFRS) and net investment recovery (investing inflow).
- Sales-type leases report higher operating cash flow at the inception of the lease, but lower operating cash flow over the lease term.

Example

The lessor has an asset on the books with a cost of $1,000. The lessor leases the asset out as a sales-type lease with the present value of the minimum lease payments equal to $1,300. The total minimum lease payments are $2,000 over the lease term.

The lessor recognizes a sale of $1,300 and cost of sales of $1,000, which results in a gross profit of $300 ($1,300 - $1,000).

The corresponding entry is the recognition of a net investment in leases of $1300. The net investment in leases is made up of a gross investment - unearned finance income. Gross investment is equal to the remaining minimum lease payments (not discounted) and the unearned finance income is the interest income that will be earned over the remainder of the lease. In this instance these are as follows:

Gross investment in lease: $2,000

Unearned finance income: ($700)

Net investment: $1,300

As you can see, the unearned finance income is the difference between the minimum lease payments and the present value of the minimum lease payments. This amount amortizes on the income statement as interest income as the lease is repaid.

10. Defined Contribution and Defined Benefit Pension Plans

```
  A <b>pension plan</b> provides benefits to retirees for services provided during emplo
```

There are two types of pension plans: defined-benefit plans and defined-contribution plans.

Defined Contribution Plans

In a defined contribution plan, the employer agrees to contribute a certain sum each period based on a formula. This formula may consider such factors as age, length of employee service, employer's profits, and compensation level. Only the employer's contribution is defined; no promise is made regarding the ultimate benefits paid out to the employees.

The accounting for a defined contribution plan is straightforward. The employer's responsibility is simply to make a contribution each year based on the formula established in the plan. If the contribution is less than the pension expense, the employer accrues a liability. If the contribution is more than then pension expense, the employer accrues an asset.

Defined Benefit Plans

A **defined benefit plan** defines the benefits that employees will receive at the time of retirement.

- The employer is committed to specified retirement benefits.
- The trust accumulates assets, and the employer is the trust-beneficiary. That is, the employer assumes the risk of the investment. As long as the plan continues, the employer is responsible for the payment of the defined benefits, regardless of what happens in the trust.
- Retiree benefits are a fixed amount. Any shortfall in the accumulated assets held in the trust must be made up by the employer, and any excess accumulated in the trust can be recaptured by the employer.

The employer needs to determine what the contribution should be today to meet the pension benefit commitments that will arise at retirement. It is at risk because it must be sure to make enough contributions. The liability is often controversial because its measurement and recognition relate to unknown future variables.

The plan's return objective is to meet the **actuarial rate of return**, which is the discount rate used to find the present value of the plan's future obligations and therefore determines the size of the firm's annual contribution to the pension plan.

The principal elements of a defined benefit plan are (1) the obligation for benefits to be paid to retirees and (2) the plan assets that will be used to meet that obligation.

The determination of pension cost is a function of the following components:

- **Service cost**

  Service cost is the actuarial present value of pension benefits attributed to employee service in a period, based on the pension benefit formula.

  Example

  Assume that the annual retirement benefit earned by one employee in the current period is $5,000. Determine the service cost for the current year assuming an 8% discount rate, 15 years until retirement, and 10 years of retirement payments.

  Annual retirement benefit: $5,000
  Present value of an annuity factor (PVA, 8%, 10 periods of payments): 6.71008
  Present value of $1 factor (PV1, 8%, 15 periods until retirement): 0.31524
  Service cost = $5,000 x 6.71008 x 0.31524 = $10,576

- **Interest cost**

  Interest cost is the growth in Projected Benefit Obligation (PBO) during a reporting period. It is calculated as $\text{PBO}_{\text{Beginning}}$ x discount rate of the previous period.

  Example

Assume that 1995 is the first year of a retirement plan. PBO is $650,000 on 12/31/1995 and $740,000 on 12/31/1996. If the discount rate is 8%, what is the interest cost for 1995 and 1996, respectively?

Interest cost is calculated by multiplying the PBO at the beginning of the year by the discount rate.
PBO 1/1/1995 = $0 x 8% = $0
PBO 1/1/1996 = $650,000 x 8% = $52,000

- **Expected return on plan assets**
- **Actuarial gains and losses**

  These occur when changes in assumptions about future events, such as quit rates, retirement dates, mortality, and the discount and compensation increase rates, decrease or increase PBO.

- **Prior service costs**

  Prior service costs (**PSC**) result from the granting of pension benefits for service rendered before the pension plan began or from plan amendments granting increased pension benefits for service rendered before the amendment. It is the present value of the retroactive benefits.

11. Leverage and Coverage Ratios

```
Major leverage and coverage ratios are discussed in Reading 20 [Financial Analysis Tec
```

Financial Reporting and Analysis (4)

Financial Reporting Quality

1. Reporting Quality and Results Quality

```
<b>Financial reporting quality</b>: a subjective evaluation of the extent to which fir
```

Earnings are considered to be high quality if they exhibit persistence and are unbiased. Sustainable earnings enable better forecasts of future cash flows or earnings. This is referred to as results quality or **earnings quality**.

Financial reporting quality is different from earnings quality. The two concepts are, however, interrelated because a correct assessment of earnings quality is possible only if we have some basic level of financial reporting quality. Low financial reporting quality makes it hard to assess earnings quality.

2. Quality Spectrum of Financial Reports

```
Financial reporting quality varies across companies.<p> </p>
```

GAAP, Decision-Useful, Sustainable, and Adequate Returns

- GAAP compliance.
- Useful: helpful in decision-making. Relevant, faithful representation and material.
- Sustainable earnings indicate an adequate level of return on investment.

GAAP, Decision-Useful, but Sustainable?

- GAAP compliance and useful.
- But not sustainable earnings.

Biased Accounting Choices

- Within GAAP.
- Biased choices such as aggressive/conservative accounting, income smoothing, hidden reserves, and earnings management.

Departures from GAAP

It is difficult or impossible to assess earnings quality. Engaging in fraudulent financial reporting provides no quality of earnings.

Conservative and Aggressive Accounting

An aspect of financial reporting quality is the degree to which accounting choices are conservative or aggressive. "Aggressive" typically refers to choices that aim to enhance a company's reported performance and financial position by inflating the amount of revenues, earnings, and/or operating cash flow reported in the period or by decreasing the amount of expenses reported in the period and/or the amount of debt reported on the balance sheet.

Conservatism in financial reports can result from either (1) accounting standards that specifically require a conservative treatment of a transaction or an event or (2) judgments necessarily made by managers when applying accounting standards that result in more or less conservative results.

An example of conservatism in the oil and gas industry is the revenue recognition accounting standard. This standard permits recognition of revenue only at the time of shipment rather than closer to the time of actual value creation (which is the time of discovery).

Big Bath Accounting

The strategy of manipulating a company's income statement to make poor results look even worse. The big bath is often implemented in a bad year to artificially enhance next year's earnings. The big rise in earnings might result in a larger bonus for executives.

Cookie Jar Reserve Accounting

Companies shift earnings around by creating overly large reserve accounts in good years then drawing them down in bad years.

3. Context for Assessing Financial Reporting Quality

```
    <strong>Motivations</strong> for managers to issue less than high quality financial re
```

- Mask poor performance
- Boost stock price
- Improve incentive compensation
- Meet debt covenants

Management might have an incentive to manipulate earnings lower as well, possibly to smooth higher earnings in the current quarter into weaker quarters.

Conditions conductive to issuing low-quality financial reports:

- **Opportunity** is generally provided through weaknesses in internal controls.
- **Motivation** can be imposed due to personal financial problems or unrealistic deadlines and performance goals.
- **Rationalization** occurs when an individual develops a justification for fraudulent activities.

Mechanisms that discipline financial reporting quality:

- **The free market.** A company seeking to minimize its long-term cost of capital should aim to provide

high-quality financial reports.
- Enforcement by **market regulatory authorities**, which plays a central role in encouraging high-quality financial reporting.
- **Auditors.** An audit is intended to provide assurance that a company's financial reports are presented fairly. There are, however, inherent limitations. Auditors are only able to offer "reasonable assurance" of the truth and fairness of financial statements rather than absolute assurance.
- **Private contracts.** External parties such as lenders and investors are motivated to ensure the quality of financial reports is high.

4. Detection of Financial Reporting Quality Issues

```
   There is really nothing new in this reading, just a review of the previous material. A
```

Presentation Choice

If a company uses a non-GAAP financial measure in an SEC filing, it is required to provide the most directly comparable GAAP measure with equivalent prominence in the filing. In addition, the company is required to provide a reconciliation between the non-GAAP measure and the equivalent GAAP measure.

Similarly, IFRS require that any non-IFRS measures included in financial reports must be defined and their potential relevance explained. The non-IFRS measures must be reconciled with IFRS measures.

Accounting Choices and Estimates

Managers' considerable flexibility in choosing their companies' accounting policies and formulating estimates provides opportunities for aggressive accounting.

Examples include:

- Revenue recognition policies.
- Inventory cost flow assumptions.
- Capitalization policies.
- Estimates of uncollectible account receivable.
- Estimated realizability of deferred tax assets.
- Depreciation method, estimated salvage value of depreciable assets, and estimated useful life of depreciable assets.

Cash flow, especially operating cash flow and free cash flow, are always at the heart of any discussion of financial performance and valuation. Investors, creditors, and analysts are all interested in whether a firm is generating cash flow and where that cash flow can be expected to recur.

Operating cash flow is usually unaffected by estimates and judgments. However, firms can still create the perception that sustainable operating cash flow is greater than it actually is. One technique is to misrepresent a firm's cash-generating ability by classifying financing activities as operating activities and vice versa. Additionally, management has discretion over the timing of cash flows and where to report cash flows.

Warning Signs

Analysts should pay attention to:

- Revenue. Check revenue recognition policies and revenue relationship.
- Inventories. Look at inventory relationships.
- Capitalization policies and deferred costs.
- The relationship of cash flow and net income.
- Other warning signs.

Applications of Financial Statement Analysis

1. Evaluating Past Financial Performance

This reading describes selected applications of financial statement analysis. In all c

Evaluating a company's historical performance addresses not only what happened but also the causes behind the company's performance and how the performance reflects the company's strategy. The analyst needs to create common-size financial statements, calculate the financial ratios of the company, its competitors, and the industry, and make necessary adjustments. After processing the data, the analyst should perform:

- **time series analysis** to compare the company's performance to itself over time to examine the trend of its ratios (e.g., profitability, efficiency, liquidity, and solvency ratios).
- **cross-sectional analysis** to compare these ratios to those of its competitors or the industry.

When examining the data, the analyst should try to find answers to critical questions, including:

- What are the key performance indicators of the company, in light of its competitive strategy?
- What is driving the company's current performance? Specifically, what factors are causing the changes of a particular ratio over time? Why?
- What aspects of performance are critical for the company to succeed in the market? How did the company do in the past?
- What strategy does the company have and what were its impacts on the company's performance in the past?

Two examples are presented in the textbook to illustrate the application.

2. Projecting Future Financial Performance

The projection of a company's future net income and cash flow often begins with a top-

By projecting profit margins and expenses, and the level of investment in working capital and fixed capital needed to support projected sales, the analyst can forecast net income and cash flow. When projecting profit margins:

- For relatively mature companies operating in non-volatile product markets, historical information on operating profit margins can be used to estimate future operating profits. Non-recurring items should be removed from computations.
- For a new company, or a company in a volatile market or a capital intensive industry, historical operating profit margins are usually less reliable in projecting future margins.

Sensitivity analysis is often used to assess the impact of different assumptions on income and cash flow. These assumptions include sales forecasts, working capital requirements, profit margins, etc.

3. Assessing Credit Risk

Credit risk is risk due to uncertainty about a counterparty's ability to meet its obli

Moody's ratings focus primarily on four factors:

1. Company profile - Scale and diversification.

These elements are indicative of other characteristics that mitigate risk and are a good indicator of market leadership, purchasing power, operational flexibility, the potential for enhanced access to financing and the capital markets, etc.

2. Financial policies - Tolerance for leverage.

Cash flow available to service indebtedness is considered the most fundamental measure of credit stature. Various solvency ratios are used for that purpose:

- Retained Cash Flow (RCF)/ Total Debt (TD)
- (RCF - CapEx) / TD
- TD / EBITDA
- (EBITDA - CapEx) / Interest
  *CapEx: Capital expenditure
- EBITDA / Interest

3. Operational efficiency.

This factor is analogous to operating leverage. Since they can generate larger levels of cash flow, companies with low operating leverage (i.e., superior profit margins) can afford to have larger debt loads. Owing to the fact that debt loads can be restructured, low-cost companies have better prospects than high-cost companies when faced with financial stress/distress and forced reorganizations.

4. Margin stability.

Lower volatility in margins would imply lower risk relative to economic conditions.

4. Screening for Potential Equity Investments

```
   A <b>bottom-up</b> manager is one who looks for stocks company by company. These are t
```

- Which metrics to use as screens?
- How many metrics to include?
- What values of those metrics to use as cutoff points?
- What weighting to give each metric?

Many studies have been done to determine the most effective accounting ratios for screening equity investments.

Backtesting is the process of testing a trading strategy on prior time periods. Instead of applying a strategy for the time period forward (which could take years), an analyst can do a simulation of his or her trading strategy on relevant past data in order to gauge its effectiveness. However, as frequently heard, "past performance does not necessarily guarantee future returns"; backtesting may not provide a reliable indication of future performance because of survivorship bias, look-ahead bias, or data-snooping.

5. Analyst Adjustments to Reported Financials

```
   Analysts' adjustments to a company's reported financial statements are sometimes neces
```

Investments Adjustments

Different categories of investment securities have different treatments regarding unrealized holding gains and losses. Depending on management's intention, investment securities can be classified as:

- Trading securities. Any unrealized gains and losses are recognized on the income statement as part of the net income.
- Available-for-sale securities. Any unrealized gains and losses are recognized on the balance sheet as part of other comprehensive income.

Adjustments may be needed to facilitate the comparison of two otherwise comparable companies that have significant differences in the classification of investments.

Inventory Adjustments

IAS No.2 does not permit the use of LIFO. If a company not reporting under IFRS uses LIFO but another company uses FIFO, comparison of the two companies may be difficult. Reading 21 [Inventories] illustrates how to make an inventory adjustment and its impact.

Property, Plant and Equipment

Companies may choose different depreciation methods (e.g., a straight-line method or an accelerated method) and accounting estimates (e.g., salvage value or useful life) related to depreciation. Disclosures required for depreciation often do not facilitate specific adjustments. Analysts may evaluate the relationships between various depreciation-related items (e.g., gross PPE, accumulated depreciation, depreciation expense, cash flows for capital expenditure, and asset disposals).

- Relative Age (in %) = Accumulated Depreciation / Ending Gross Investment. This equation suggests how much of the useful life of the company's overall asset base has passed.
- Average Depreciable Life = Ending Gross Investment / Depreciation Expense.
- Average Age (in years) = Accumulated Depreciation / Depreciation Expense. This equation indicates how many years' worth of depreciation expense has already been recognized.

The above three indicators are discussed in Reading 22 [Long-Lived Assets].

- The ratio of Net PPE / Depreciation Expense suggests how many years of useful life remain for a company's overall asset base.
- CapEx / (Gross PPE + CapEx) signifies what percentage of the asset base is being renewed through new capital investment.
- CapEx / Asset Disposal indicates the growth of the asset base.

Goodwill

Goodwill is recorded as an asset if one company purchases another for a price that is more than the fair value of the assets acquired. Internally generated goodwill is not recorded on the balance sheet. Adjustments are needed to compare two otherwise comparable companies when one has a recorded goodwill asset. The textbook provides an excellent example of the ratio comparisons for goodwill.

Off-Balance-Sheet Financing

This topic is covered in Reading 24 [Non-current (Long-term) Liabilities].

Corporate Issuers

Corporate Issuers (1)

Introduction to Corporate Governance and Other ESG Considerations

1. Corporate Governance Overview

```
Corporate governance is "the system by which companies are directed and controlled."<p
```

According to the textbook, corporate governance is "the system of internal controls and procedures by which individual companies are managed. It provides a framework that defines the rights, roles and responsibilities of various groups ... within an organization. At its core, corporate governance is the arrangement of checks, balances, and incentives a company needs in order to minimize and manage the conflicting interests between insiders and external shareowners."

Corporate governance is about promoting corporate fairness, transparency, and accountability. Its purpose is to prevent one group from expropriating the cash flows and assets of one or more other groups.

There are many systems of corporate governance, most reflecting the influences of either shareholder theory or stakeholder theory, or both. Current trends point to increasing convergence.

2. Company Stakeholders

```
    Stakeholders are individuals or groups with an interest, or stake, in a firm. They are
```

The corporate governance structure specifies the distribution of rights and responsibilities among stakeholders, and spells out the rules and procedures for making decisions on corporate affairs.

Shareholders

Shareholders provide funds and expect returns. They are the legal owners of a firm and their wealth is directly related to the value of the company. They typically focus on growth in company profitability.

- The money provided by stockholders is called risk capital, because the stockholders are making a risky investment in the firm with no guarantee of returns or even the preservation of their original investment.
- Because of their willingness to assume risk, managers are obliged to reward stockholders by pursuing strategies that maximize returns to them.
- There are controlling shareholders and minority shareholders.

Creditors

Creditors provide funds and expect repayment and interest. They typically don't have much control over a firm.

Managers and Employees

Managers and employees provide labor, skills, and ideas, and expect income, job satisfaction and security, and good working conditions. Managers can best serve the interests of stockholders by increasing profitability. Higher profits generate more funds for paying high salaries and offering more benefits to employees.

Board of Directors

A board of directors is a group of individuals that are elected by shareholders to protect shareholder interests, provide strategic direction, and monitor company and management performance. There are one-tier and two-tier structures.

Customers

Customers provide sales revenues and expect products that provide value for money.

Suppliers

Suppliers provide inputs and expect revenues and dependable buyers.

Governments/Regulators

Governments provide regulation and expect companies to adhere to the rules.

3. Principal-Agent and Other Relationships in Corporate Governance

```
    <b>Shareholder and Manager/Director Relationships</b><p> </p>
```

Problems can arise in a business relationship when one person delegates decision-making authority to another. The **principal** is the person delegating authority, and the **agent** is the person to whom the authority is delegated.

Agency theory offers a way to understand why managers do not always act in the best interests of stakeholders.

- Managers and shareholders may have different goals. They may also have different attitudes towards risk.
- **Information asymmetry**. Managers almost always have more information than shareholders. Thus, it is difficult for shareholders to measure managers' performance or to hold them accountable for their

performance.

Controlling and Minority Shareholder Relationships

Ownership structure is one of the main dimensions of corporate governance. For firms with controlling shareholders, separation of ownership and control generates a two-level agency problem: between controlling shareholders and management and between minority shareholders and controlling shareholders. The interests of controlling and minority shareholders are often not aligned.

For example, if a company has two classes of common shares (dual classes of common equity):

- Class A shareholders have all the voting rights.
- Class B shareholders don't have any voting rights.

The management team and the board are more likely to focus on the interests of Class A shareholders. The rights of Class B shareholders may suffer as a consequence of the ownership structure.

Minority shareholders have less influence on the board composition than controlling shareholders. Controlling shareholders may receive special attention from management. They are often in the position to facilitate third-party takeovers by splitting the large gains on their own shares with the bidder.

Manager and Board Relationships

This is another example of agency theory (discussed above).

Shareholder versus Creditor Interests

These two parties have different relationships to the company, accompanied by different rights and financial returns. For example, shareholders have an incentive to take on riskier projects than creditors do, as creditors are more interested in strategies that will increase the chances of getting their investment back. Shareholders also prefer that the company pay more out in dividends than creditors would like.

Other Stakeholder Conflicts

There are conflicts among other stakeholders, such as those between:

- customers and shareholders;
- customers and suppliers;
- shareholders and government or regulators.

4. Stakeholder Management

```
<b>Stakeholder management</b> involves identifying, prioritizing, and understanding th
```

Mechanisms of Stakeholder Management

Mechanisms of stakeholder management may include:

- General meetings.
  - The right to participate in general shareholder meetings is a fundamental shareholder right. Shareholders, especially minority shareholders, should have the opportunity to ask questions of the board, to place items on the agenda and to propose resolutions, to vote on major corporate matters and transactions, and to participate in key corporate governance decisions, such as the nomination and election of board members.
  - Shareholders should be able to vote in person or in absentia, and equal consideration should be given to votes cast in person or in absentia.

- A board of directors, which serves as a link between shareholders and managers, acts as the shareholders' monitoring tool within the company.

- The audit function. It plays a critical role in ensuring the corporation's financial integrity and consideration of legal and compliance issues. The primary objective is to ensure that the financial information reported by the company to shareholders is complete, accurate, reliable, relevant, and timely.
- Company reporting and transparency. It helps reduce of information asymmetry and agency costs.
- Related-party transactions. Related-party transactions involve buying, selling, and other transactions with board members, managers, employees, family members, and so on. They can create an inherent conflict of interest. Policies should be established to disclose, mitigate, and manage such transactions.
- Remuneration policies. Does the company's remuneration strategy reward long-term or short-term growth? Are equity-based compensation plans linked to the long-term performance of the company?
  - **Say on pay** is the ability of shareholders in a company to actively vote on how much executives employed by the company should be compensated.
- Contractual agreements with creditors; indentures, covenants, collaterals and credit committees are tools used by creditors to protect their interests.
- Employee laws, contracts, codes of ethics and business conduct, and compliance offer(s) are all means a company can use to manage its relationship with its employees.
- Contractual agreements with customers and suppliers.
- Laws and regulations a company must follow to protect the rights of specific groups.

5. Board of Directors and Committees

```
<b>Composition of the Board of Directors</b><p> </p>
```

A board of directors is the central pillar of the governance structure, serves as the link between shareholders and managers, and acts as the shareholders' internal monitoring tool within the company.

The structure and composition of a board of directors vary across countries and companies. The number of directors may vary, and the board typically includes a mix of expertise levels, backgrounds, and competencies. Board members must have extensive experience in business, education, the professions and/or public service so they can make informed decisions about the company's future. If directors lack the skills, knowledge and expertise to conduct a meaningful review of the company's activities, and are unable to conduct in-depth evaluations of the issues affecting the company's business, they are more likely to defer to management when making decisions.

Executive (internal) directors are employed by the company and are typically members of senior management. Non-executive (external) directors have limited involvement in daily operations but serve an important oversight role.

In a **classified** or **staggered** board, directors are typically elected in two or more classes, serving terms greater than one year. Proponents argue that by staggering the election of directors, a certain level of continuity and skill is maintained. However, staggered terms make it more difficult for shareholders to make fundamental changes to the composition and behavior of the board and could result in a permanent impairment of long-term shareholder value.

Functions and Responsibilities of the Board

Two primary duties of a board of directors are *duty of care* and *duty of loyalty*. Among other responsibilities, the board is to:

- establish long-term strategic objectives for the company with a goal of ensuring that the best interests of shareholders come first and that the company's obligations to others are met in a timely and complete manner.
- establish clear lines of responsibility and a strong system of accountability and performance measurement in all phases of a company's operations.
- hire the chief executive officer, determine the compensation package, and periodically evaluate the officer's performance.
- ensure that management has supplied the board with sufficient information for it to be fully informed and prepared to make the decision that are its responsibility, and to be able to adequately monitor and oversee the company's management.

Board of Directors Committees

A company's board of directors typically has several committees that are responsible for specific functions and report to the board.

- The **audit committee** plays a critical role in ensuring the corporation's financial integrity and consideration of legal and compliance issues. The primary objective is to ensure that the financial information reported by the company to shareholders is complete, accurate, reliable, relevant, and timely.
- The **governance committee** tries to ensure that the company adopts good corporate governance practices.
- The **remuneration (compensation) committee** develops and implements executive compensation policies. Incentives should be provided for actions that boost long-term share profitability and value.
- The **nomination committee** searches for and nominates board director candidates, and establishes the nomination policies and procedures.
- Other common committees include those responsible for overseeing management's activities in certain areas, such as mergers and acquisitions, legal matters, and risk management.

6. Factors Affecting Stakeholder Relationships and Corporate Governance

```
Stakeholder relationships and corporate governance are continually shaped and influenc
```

Market Factors

Shareholder engagement involves a company's interactions with its shareholders. It can provide benefits that include building support against short-term activist investors, countering negative recommendations from proxy advisory firms, and receiving greater support for management's position.

Shareholder activism encompasses a range of strategies that may be used by shareholders seeking to compel a company to act in a desired manner. It can take any of several forms: proxy battles, public campaigns, shareholder resolutions, litigation, and negotiations with management.

Corporate takeovers can happen in different ways: proxy contest or proxy fight, tender offer, hostile takeover, etc. The justification for the use of various anti-takeover defenses should rest on the support of the majority of shareholders and on the demonstration that preservation of the integrity of the company is in the long-term interests of shareholders.

Non-Market Factors

These factors include the legal environment, the media, and the corporate governance industry itself.

7. Corporate Governance and Stakeholder Management Risks and Benefits

```
    From a corporation's perspective, risks of poor governance include:<p> </p><ul class="
```

- weak control systems or inefficient monitoring tools;
- ineffective decision making;
- legal, regulatory, and reputational risks;
- default and bankruptcy risks.

Benefits of effective governance and stakeholder management include:

- better operational efficiency and control brought by effective monitoring tools and control mechanisms;
- better operating and financial performance;
- lower default risk and cost of debt.

8. Analyst Considerations in Corporate Governance and Stakeholder Management

```
    <b>Economic Ownership and Voting Control</b><p> </p>
```

What is the company's ownership and voting structure among shareholders? Why do some shareholders own a small portion of a company's stock but get most of the voting power? Does the practice really insulate managers from Wall Street's short-term mindset? Dual-class structures create an inferior class of shareholders, and may allow management to make bad decisions with few consequences.

Board of Directors Representation

Analysts should assess whether the experience and skill sets of board members match the needs of the company. Are they truly independent? Are there inherent conflicts of interest?

Remuneration and Company Performance

What are the main drivers of the management team's remuneration and incentive structure? Does the remuneration plan reward long-term or short-term growth? Is it based on the performance of the company relative to its competitors or other peers? Equity-based remuneration plans can offer the greatest incentives. Are they linked to the long-term performance of the company? What is the impact on the income statement?

Investors in the Company

What is the composition of investors in a company? Are there any significant investors in the company? Any sizable affiliated stockholders that can block the votes of the majority? Any activist shareholders that could bring rapid changes for the company?

Strength of Shareholders' Rights

How robust are the shareholder rights at the company? How robust compared to those of peers?

Managing Long-Term Risks

How effectively is the company managing long-term risks, such as securing access to necessary resources, managing human capital, exhibiting integrity and leadership, and strengthening the long-term sustainability of the enterprise?

9. ESG Considerations for Investors

```
    ESG integration is the practice of considering environmental, social, and governance f
```

ESG Factors in Investment Analysis

Environmental factors include natural resource management, pollution prevention, water conservation, energy efficiency and reduced emissions, the existence of carbon assets, and adherence to environmental safety and regulatory standards.

Social factors generally pertain to human rights and welfare concerns in the workplace, product development, and, in some cases, community impact.

ESG Implementation Methods

Asset managers and asset owners can incorporate ESG issues into the investment process in a variety of ways.

- **Negative screening** is a type of investment strategy that excludes certain companies or sectors from investment consideration because of their underlying business activities or other environmental or social concerns.

- **Positive screening** and **best-in-class** strategies focus on investments with favorable ESG aspects.

- **Thematic investing** focuses on a single factor, such as energy efficiency or climate change.

- **Impact investing** strategies are targeted investments, typically made in private markets, aimed at solving social or environmental problems.

Uses of Capital

1. The Capital Allocation Process

```
<p> </p>The <b>capital allocation process</b> is the process of planning expenditures
```

- "Capital" refers to long-term assets.
- The "budget" is a plan which details projected cash inflows and outflows during a future period.

Steps in the Capital Allocation Process

The typical steps:

1. Idea Generation: Generating good investment ideas to consider.
2. Investment Analysis: Analyzing individual proposals (forecasting cash flows, evaluating profitability, etc.).
3. Capital Allocation Planning: How does the project fit within the company's overall strategies? What's the timeline and priority?
4. Monitoring and Post-Auditing: It is a follow-up of capital allocation process and a key element. By comparing actual results with predicted results and then determining why differences occurred, decision-makers can improve forecasts (based on which good capital allocation decisions can be made). Otherwise, you will have the GIGO (garbage in, garbage out) problem. Improve operations, thus making capital decisions well-implemented.

Types of Capital Projects

Replacement Projects. There are two types of replacement decisions:

- Replacement decisions to maintain a business. The issue is twofold: should the existing operations be continued? If yes, should the same processes continue to be used? Maintenance decisions are usually made without detailed analysis.
- Replacement decisions to reduce costs. Cost reduction projects determine whether to replace serviceable but obsolete equipment. These decisions are discretionary and a detailed analysis is usually required.

The cash flows from the old asset must be considered in replacement decisions. Specifically, in a replacement project, the cash flows from selling old assets should be used to offset the initial investment outlay. Analysts also need to compare revenue/cost/depreciation before and after the replacement to identify changes in these elements.

Expansion Projects. Projects concerning expansion into new products, services, or markets involve strategic decisions and explicit forecasts of future demand, and thus require detailed analysis. These projects are more complex than replacement projects.

New Products and Services. Such investments usually involves more stakeholders and higher degrees of risk.

Regulatory, Safety and Environmental Projects. These projects are mandatory investments, and are often non-revenue-producing.

Others. Some projects need special considerations beyond traditional capital budgeting analysis (for example, a very risky research project in which cash flows cannot be reliably forecast).

Capital Allocation Assumptions

Capital allocation decisions are based on incremental after-tax cash flows discounted at the opportunity cost of capital. Assumptions are:

- Capital budgeting decisions must be based on cash flows, not accounting income.

  Accounting profits only measure the return on the invested capital. Accounting income calculations reflect non-cash items and ignore the time value of money. They are important for some purposes, but for capital allocation process, cash flows are what are relevant.

  Economic income is an investment's after-tax cash flow plus the change in the market value. Financing costs are ignored in computing economic income.
- The opportunity cost should be charged against a project. Remember that just because something is on hand does not mean it's free. See below for the definition of opportunity cost.
- Expected future cash flows must be measured on an after-tax basis. The firm's wealth depends on its usable after-tax funds.
- Cash flow timing is critical because money is worth more the sooner you get it. Also, firms must have adequate cash flow to meet maturing obligations.
- Ignore how the project is financed. Interest payments should not be included in the estimated cash flows since the effects of debt financing are reflected in the cost of capital used to discount the cash flows. The existence of a project depends on business factors, not financing.

Important Capital Allocation Concepts

A **sunk cost** is a cash outlay that has already been incurred and which cannot be recovered regardless of whether a project is accepted or rejected. Since sunk costs are not increment costs, they should not be included in the capital budgeting analysis.

For example, a small bookstore is considering opening a coffee shop within its store, which will generate an annual net cash outflow of $10,000 from selling coffee. That is, the coffee shop will always be losing money. In the previous year, the bookstore spent $5,000 to hire a consultant to perform an analysis. This $5,000 consulting fee is a sunk cost; whether the coffee shop is opened or not, the $5,000 is spent.

An **opportunity cost** is the return on the *best alternative use* of an asset or the highest return that will not be earned if funds are invested in a particular project. For example, to continue with the bookstore example, the space to be occupied by the coffee shop is an opportunity cost - it could be used to sell books and generate a $5,000 annual net cash inflow.

An **incremental cash flow** is the net cash flow attributable to an investment project. It represents the change in the firm's total cash flow that occurs as a direct result of accepting the project.

- Forget sunk costs.
- Subtract opportunity costs.
- Consider side effects on other parts of the firm: externalities and cannibalization.
- Recognize the investment and recovery of net working capital.

Externalities are the effects of a project on cash flows in other parts of a firm. Although they are difficult to quantify, they should be considered. Externalities can be either positive or negative:

- **Positive externalities** create benefits for other parts of the firm. For example, the coffee shop may generate some additional customers for the bookstore (who otherwise may not buy books there). Future cash flows generated by positive externalities occur with the project and do not occur without the project, so they are incremental.

- **Negative externalities** create costs for other parts of the firm. For example, if the bookstore is considering opening a branch two blocks away, some customers who buy books at the old store will switch to the new branch. The customers lost by the old store are a negative externality. The primary type of negative externality is **cannibalization**, which occurs when the introduction of a new product causes sales of existing products to decline.

Future cash flows represented by negative externalities occur regardless of the project, so they are non-incremental. Such cash flows represent a transfer from existing projects to new projects, and thus should be subtracted from the new projects' cash flows.

Conventional versus non-conventional cash flows.

- A conventional cash flow pattern is one with an initial outflow followed by a series of inflows.

- In a non-conventional cash flow pattern, the initial outflow can be followed by inflows and/or outflows.

Project Interactions

Independent projects versus **mutually exclusive projects.** Mutually exclusive projects are investments that compete in some way for a company's resources - a firm can select one or another but not both. Independent projects, on the other hand, do not compete for the firm's resources. A company can select one or the other or both, so long as minimum profitability thresholds are met.

Project sequencing. How does one sequence multiple projects over time, since investing in project B may depend on the result of investing in project A?

Unlimited funds versus **capital rationing.** Capital rationing occurs when management places a constraint on the size of the firm's capital budget during a particular period. In such situations, capital is scarce and should be allocated to the projects most likely to maximize the firm's aggregate NPV. The firm's capital budget and cost of capital must be determined simultaneously to best allocate the firm's capital. On the other hand, a firm can raise the funds it wants for all profitable projects simply by paying the required rate of return.

2. Investment Decision Criteria

```
    <p> </p>When a firm is embarking upon a project, it needs tools to assist in making th
```

Net Present Value (NPV)

This method discounts all cash flows (including both inflows and outflows) at the project's cost of capital and then sums those cash flows. The project is accepted if the NPV is positive.

where CF_t is the expected cash flow at period t, k is the project's cost of capital, and n is its life.

- Cash outflows are treated as negative cash flows since they represent expenditures of the company to fund the project.
- Cash inflows are treated as positive cash flows since they represent money being brought into the company.

The NPV represents the amount of present-value cash flows that a project can generate after repaying the invested capital (project cost) and the required rate of return on that capital. An NPV of zero signifies that the project's cash flows are just sufficient to repay the invested capital and to provide the required rate of return on that capital. If a firm takes on a project with a positive NPV, the position of the stockholders is improved.

Decision rules:

- The higher the NPV, the better.
- Reject if NPV is less than or equal to 0.

NPV measures the dollar benefit of the project to shareholders. However, it does not measure the rate of return of the project, and thus cannot provide "safety margin" information. Safety margin refers to how much the project return could fall in percentage terms before the invested capital is at risk.

Assuming the cost of capital for the firm is 10%, calculate each cash flow by dividing the cash flow by $(1 + k)^t$ where k is the cost of capital and t is the year number. Calculate the NPV for Project A and B above.

$NPV = CF_0 + CF_1 + CF_2 + CF_3 + CF_4$

Project A's NPV = $-1{,}000 + 750/1.10^1 + 350/1.10^2 + 150/1.10^3 + 50/1.10^4 = -1{,}000 + 682 + 289 + 113 + 34$ = \$118 (rounded)

Project B's NPV = $-1{,}000 + 100/1.10^1 + 250/1.10^2 + 450/1.10^3 + 750/1.10^4 = -1{,}000 + 91 + 207 + 338 + 512$ = \$148 (rounded)

Internal Rate of Return (IRR)

This is the discount rate that forces a project's NPV to equal zero.

Note that this formula is simply the NPV formula solved for the particular discount rate that forces the NPV to equal zero. The IRR on a project is its expected rate of return. The NPV and IRR methods will *usually* lead to the same accept or reject decisions.

Decision rules:

- The higher the IRR, the better.
- Define the **hurdle rate**, which typically is the cost of capital.
- Reject if IRR is less than or equal to the hurdle rate.

IRR does provide "safety margin" information.

Calculate Project A's and B's IRR.

Project A: $-1000 + 750/(1 + IRR)^1 + 350/(1+IRR)^2 + 150/(1+IRR)^3 + 50/(1+IRR)^4 = 0$

Since it is difficult to determine by hand, the use of a financial calculator is needed to solve for IRR.

The IRR for Project A is 18.32% and for Project B is 15.03%.

NPV Profile

This section is not required.

A **NPV profile** is a graph showing the relationship between a project's NPV and the firm's cost of capital. The point where a project's net present value profile crosses the horizontal axis indicates a project's internal rate of return.

Some observations:

- The IRR is the discount rate that sets the NPV to 0.
- The NPV profile declines as the discount rate increases.
- Project A has a higher NPV at low discount rates, while Project B has a higher NPV at high discount rates. The NPV profiles of Project A and B join at the crossover rate, at which the projects' NPVs are equal.
- The slope of Project A's NPV profile is steeper. This indicates that Project A's NPV is more sensitive to changes in the discount rates.

Comparison of the NPV and IRR Methods

The IRR formula is simply the NPV formula solved for the particular rate that sets the NPV to 0. The same equation is used for both methods.

The NPV method assumes that cash flows will be reinvested at the firm's cost of capital, while the IRR method assumes reinvestment at the project's IRR. Reinvestment at the cost of capital is a better assumption in that it is closer to reality.

For **independent projects**, the NPV and IRR methods indicate the same accept or reject decisions. Assuming that Project A and B are independent, consider their NPV profiles.

- The IRR criterion for accepting an independent project is IRR > hurdle rate. That is, the cost of capital must be less than (or to the left of) the IRR.
- Whenever the cost of capital is less than the IRR, the project's NPV is positive. Recall the decision rule for independent projects: accept if NPV > 0. Thus, both projects should be accepted based on the NPV method.

However, for **mutually exclusive projects**, ranking conflicts can arise. Assuming that Project A and B are mutually exclusive, consider their NPV profiles.

- If the cost of capital > crossover rate, then NPVB > NPVA and IRRB > IRRA. Thus, both methods lead to the selection of Project B.
- If the cost of capital < crossover rate, then NPVB < NPVA and IRRB > IRRA. Thus, a conflict arises because now the NPV method will select Project A while the IRR method will choose B.
- Therefore, for mutually exclusive projects, the NPV and IRR methods lead to same decisions if the cost of capital > the crossover rate and different decisions if the cost of capital < the crossover rate.

For mutually exclusive projects, the NPV and MIRR methods will lead to the same accept or reject decision when:

- Two projects are of equal size and have the same life.

- Two projects are of equal size but differ in lives.

However, the projects can generate conflicting results if the NPV profiles of two projects cross (and there is a crossover rate):

- As long as the cost of capital (k) is larger than the crossover rate, the two methods both lead to the same decision;
- A conflict exists if k is less than the crossover rate.

Two conditions cause the NPV profiles to cross:

- When project size (or scale) differences exist. The cost of one project is larger than that of the other.
- When timing differences exist. The timing of cash flows from the two projects differs in that most of the cash flows from one project come in the early years while most of the cash flows from the other project come in the later years.

The root cause of the conflict between NPV and IRR is the rate of return at which differential cash flows can be reinvested. Both the NPV and IRR methods assume that the firm will reinvest all early cash flows. The NPV method implicitly assumes that early cash flows can be reinvested at the cost of capital. The IRR method assumes that the firm can reinvest at the IRR.

Whenever a conflict exists, the NPV method should be used. It can be demonstrated that the better assumption is the cost of capital for the reinvestment rate (Hint: don't focus too much on this topic, as it is beyond the scope of the CFA exam).

Multiple IRRs is a situation where a project has two or more IRRs. This problem is caused by the non-conventional cash flows of a project.

- Conventional cash flows means that the initial cash outflows are followed by a series of cash inflows.
- Non-conventional cash flows means that a project calls for a larger cash outflow either sometime during or at the end of its life. Thus, the signs of the net cash flows flip-flop during the project's life.

In fact, non-conventional cash flows can cause other problems, such as negative IRR or an IRR that leads to an incorrect accept or reject decision. However, a project can have only one NPV, regardless of its cash flow patterns, so the NPV method is preferable when evaluating projects with non-normal cash flows.

Corporate Usage of Capital Allocation Methods

The usefulness of various capital budgeting methods depends on their specific applications. Although financial textbooks often recommend the use of NPV and IRR methods, other methods are also heavily used by corporations.

Capital budgeting is also relevant to external analysts in estimating the value of stock prices. Theoretically, if a company invests in positive NPV projects, the wealth of its shareholders should increase.

The integrity of a firm's capital budgeting processes can also be used to show how the management pursues its goal of shareholder wealth maximization.

3. Real Options

```
    <p> </p>A <b>real option</b> is an economically valuable right to make or else abandor
```

Example 1: An industrialist who owns a factory with excess capacity

has an option to increase production that she may exercise at any time. This option might be of particular value when demand for the factoryâ€™s output increases.

Example 2: The owner of an oilfield has an option to drill for oil that he may exercise at any time. In fact since he can drill for oil in each time period he actually holds an entire series of options. On the other hand if he only holds a lease on the oilfield that expires on a specified date, then he holds only a finite number of drilling options.

Example 3: A company is considering to invest in a new technology. If we ignore the optionality in this investment, then it may have a negative NPV so that it does not appear to be worth pursuing. However, it may be the case that investing in this new technology affords the company the option to develop more advanced and profitable technology at a later date. As a result, the investment might ultimately be a positive NPV value project that is indeed worth pursuing.

Types of Real Options

Sizing options are options relating to the size of a project. Depending on the ROI analysis, options may exist to expand, contract, or expand and contract the project over time, given various contingencies.

Timing options relate to the lifetime of a project - to initiate one, delay starting one, abandon an existing one, or plan the sequencing of the project's steps.

Flexibility options involves the project's operations: the process flexibility, product mix, price setting, and operating scale, among others.

Fundamental Options

There are four common approaches to evaluating capital projects with real options:

1. Using the NPV without considering options. If the NPV is positive, the firm goes ahead with the investment.

1. Using the formula: Project's NPV = NPV (based on discount cash flows alone) â€“ Cost of options + Value of options

1. Using decision trees.

1. Using option pricing models.

Valuation techniques for real options do often appear similar to the pricing of financial options contracts, where the spot price or the current market price refers to the current net present value (NPV) of a project. The net present value is the cash flow that's expected as a result of the new project, but those flows are discounted by a rate that could otherwise be earned for doing nothing.

The precise value of real options can be difficult to establish or estimate. Itâ€™s difficult to pin an exact financial value on benefits of a real option.

4. Common Capital Allocation Pitfalls

```
    Here are some of the common mistakes that managers make:<p> </p><ul class="notes">
```

- **Economic Responses.**

Economic responses must be correctly anticipated when performing an investment analysis. Can competitors enter the market easily if the project turns out to be very profitable? Will vendors, suppliers and employees want to gain more from a profitable enterprise?

- **Template Errors.**

Due to large number of projects to be analyzed over time, companies have developed standardized templates for capital budgeting. Are these templates misused?

- **Pet Projects.**

Pet projects are projects that influential managers want the corporation to invest in. They should undergo normal capital budgeting analysis. However, sometimes insufficient analysis is performed or overly optimistic projections are used to inflate the profitability of a pet project.

- **Basing Decisions on short-run accounting numbers.**

Investment decisions should be based on *long-term* EPS, Net Income or ROE.

- **Basing Decisions on the IRR.**

If projects are mutually exclusive, or they have non-conventional cash flow patterns, the IRR criterion will not be economically sound. The NPV method should be used.

- **Incorrectly Accounting for Cash Flows.**

It's easy to omit or double count cash flows, and mishandle taxes in the analysis of a complex project.

- **Overhead Costs.**

It's hard to estimate appropriate overhead costs for a project in large companies.

- **Discount Rate Errors.**

The required rate of return for a project should be based on its risk, not the cost of debt, equity or weighted average of capitals involved. The longer a project's life, the bigger the impact of the discount rate errors on a project.

- **Overspending and underspending the Capital Allocation.**

Managers tend to argue, for political reasons, that their budget is small, and they tend to spend the entire investment budget just because it is available.

- **Failure to Consider Investment Alternatives.**

Many good investment alternatives are never considered.

- **Sunk Costs and Opportunity Costs.**

It is often difficult to ignore sunk costs and identifying opportunity costs.

Sources of Capital

1. Corporate Financing Options

```
<p> </p>There are different types of financing methods for companies to consider. Cons
```

Internal Financing

operating cash flows = net income + depreciation - dividend payments.

Two countering forces should be considered when managing **accounts payable**:

- Paying too early is costly unless the company can take advantage of discounts.
- Postponing payment beyond the end of the net (credit) period is known as "stretching accounts payable" or "leaning on the trade." Trade discounts should be evaluated carefully.

Accounts receivable is just the opposite of accounts payable. The most popular measures to evaluate receivables are receivable turnover and number of days of receivables.

Managing **inventory** is a juggling act. Excessive stocks can place a heavy burden on the cash resources of a business. Insufficient stocks can result in lost sales, delays for customers, etc. The goal of inventory management is to identify the level of inventory which allows for uninterrupted production but reduces the investment in raw materials - and minimizes reordering costs - and hence increases cash flow. Just-In-Time (JIT) is an inventory strategy implemented to improve a business's return on investment by reducing in-process inventory and its associated costs.

Cash does not pay interest. Companies should invest funds that are not needed in daily transactions, in **marketable securities** that can be sold quickly if they need cash.

Financial Intermediaries

Banks and non-bank lenders can offer the following means of financing.

Line of credit (L/C):

- A bank provides a letter of credit, for a fee, guaranteeing the investor that the company's obligation will be paid. It is a promise from a bank for payment in the event that certain conditions are met.
- It is frequently used to guarantee payment of an obligation.
- **Committed lines of credit** are stronger than those that are **uncommitted** because of the bank's formal commitment.

Revolving credit agreement: A formal, legal commitment to extend credit up to some maximum amount over a stated period of time (e.g. three to fix years).

Asset-based loans are forms of debt for money borrowed in which specific assets have been pledged to guarantee payment. In **assignment of accounts receivable**, the borrower pays interest, a service charge on the loan, and the assigned receivables serve as collateral. **Factoring** is the selling of receivables to a financial institution, the factor, usually "without recourse."

Capital Markets

Commercial paper:

- Short-term, unsecured promissory notes, generally issued by large corporations (unsecured corporate IOUs).
- Cheaper than a short-term business loan from a commercial bank.
- Dealers often require a line of credit to ensure that the commercial paper is paid off.

A firm can finance its operations from three main sources of capital:

- **equity or common stock**
- **preferred stock**
- **debt**

Please examine the exhibit 2 in the textbook for detailed comparisons between debt and equity financing.

A key point to note is *interest on debt is tax deductible*; therefore, to calculate the cost of debt, the tax benefit is deducted. There is no tax savings associated with the use of preferred stock or common stock.

Other Financing

A **lease** in its most basic form is the renting of some sort of property. Using leases, a firm can avoid tying up too much capital in fixed asset investment.

Considerations Affecting Financing Choices

Firm-specific financing considerations include company size, riskiness of assets, assets for collateral, public vs. private equity, asset liability management, debt maturity structure, currency risks and agency costs.

General economic considerations include taxation, inflation, government policy and monetary policy.

2. Defining Liquidity Management

```
    <p> </p><b>Liquidity</b> refers to the ability of a company to satisfy its short-term
```

Even though long-term assets may also be converted to cash to improve liquidity, it has other costs for a company, for example, it may impair a company's financial strength.

There are two sources of liquidity. The main difference between the two sources is whether or not the company's normal operations will be affected.

- Primary sources of liquidity include cash, short-term funds, and cash flow management. These resources represent funds that are readily accessible at relatively low cost.
- Secondary sources include negotiating debt contracts, liquidating assets, and filing for bankruptcy and reorganization. They provide liquidity at a higher price and may impact a company's financial and operating positions.

Drags and Pulls on Liquidity

The timing of cash receipts and disbursements can significantly affect a company's liquidity position.

A **drag on liquidity** exists when cash inflows lag.

- *Uncollected receivables*: For an analyst, the drags are often visible from an analysis of balance sheet trends and ratios. For example, a deterioration in days sales outstanding (DSO) is often an indication of negative developments acting as drags on liquidity. Increasing levels of bad debt expenses are also a useful indicator to identify issues in the collection of receivables.
- *Inventory obsolescence*: If a company's inventory is turning obsolete, it will experience a drag on liquidity as the value of such inventory declines, turning into lower cash inflows than planned. A good indication of increasing inventory obsolescence is often given by slowing inventory turnover ratios.
- *Tight credit*: If access to capital worsens or becomes more expensive, a company's liquidity may worsen.

A **pull on liquidity** is generated when cash outflows happen too quickly or when a company's access to commercial or financial credit is limited.

- *Early payments*: A company that pays its suppliers, creditors, or employees before the payment is due is creating a pull on liquidity. It is a commonplace among companies to hold payments until the due date without any anticipation of payments.
- *Reduced credit limits*: Consider a company fails to pay its obligations to its suppliers on a timely basis, or willingly takes advantage of its suppliers by paying after a long delay. In such cases, suppliers may decide to reduce the amount of trade credit to the company - impacting its liquidity.
- *Reduced lines of credit*: As a supplier can reduce the amount of credit to a customer, banks can also reduce the amount of credit available to their customers. It can be due to company-specific reasons, such as deteriorating business trends in the company or in the bank itself. In other cases, it can be a response to a customer's poor track record of debt repayment. The reductions may be mandated by governments or may be due to conditions in the credit markets, such as tighter access to funds from central banks.
- *Low liquidity positions.*

3. Measuring Liquidity

```
    <p> </p>Almost all liquidity/activity ratios are covered in Reading 18 [Understanding
```

Liquidity Ratios

Liquidity ratios measure the ability of a company to meet future short-term financial obligations from current assets and, more importantly, cash flows. Each of the following ratios takes a slightly different view of cash or near-cash items.

- **Current Ratio** is a measure of the number of dollars of current assets available to meet current obligations. It is the best-known liquidity measure. A current ratio of less than 1 indicates the company has negative working capital.

- **Quick Ratio** (**Acid-Test Ratio**) eliminates less liquid assets, such as inventory and pre-paid expenses, from the current ratio. If inventory is not moving, the quick ratio is a better indicator of cash and near-cash items that will be available to meet current obligations.

- **Cash Ratio** is the most conservative liquidity ratio, determined by eliminating receivables from the quick ratio. As with the elimination of inventory in the quick ratio, there is no guarantee that the receivables will be collected.

Activity Ratios

A company's operating activities require investments in both short-term (inventory and accounts receivable) and long-term (property, plant, and equipment) assets. Activity ratios describe the relationship between the company's level of operations (usually defined as sales) and the assets needed to sustain operating activities. They measure how well a company manages its various assets.

- **Receivables turnover** measures the liquidity of the receivables - that is, how quickly receivables are collected or turn over. The lower the turnover ratio, the more time it takes for a company to collect on a sale and the longer before a sale becomes cash.

  receivables turnover = credit sales / average receivables

  This ratio provides a better level of detail than the current or quick ratio. A company could have a favorable current or quick ratio, but if the receivables turn over very slowly, these ratios would not be a good measure of liquidity. The same applies for the inventory turnover below.

  This ratio also implies an average collection period (the number of days it takes for the company's customers to pay their bills): **number of days of receivables = average accounts receivable / (sales on credit/365)**
- **Inventory turnover** measures how fast the company moves its inventory through the system. The lower the turnover ratio, the longer the time between when the good is produced or purchased and when it is sold.

  inventory turnover = COGS / average inventory

  An abnormally high inventory turnover and a short processing time could mean inadequate inventory, which could lead to outages, backorders, and slow delivery to customers, adversely affecting sales. An extremely low inventory turnover value implies capital is being tied up in inventory and could signal obsolete inventory.

  number of days of inventory = average inventory / (COGS/365)
- **Payables turnover** measures the length of time a company has to pay its current liabilities to suppliers.

This ratio examines the use of trade credit. The longer the time, the better it is for the company, since it is an interest-free loan and offsets the lack of cash from receivables and inventory turnovers

payables turnover = purchases / average trade payables

The following measures the number of days it takes for the company to pay its bills.

number of days of payables = average accounts payable / (Purchases/365)

The **cash conversion cycle** is the time period that exists from when the company pays out money for the purchase of raw materials to when it gets the money back from the purchasers of the company's finished goods.

cash conversion cycle = days of inventory + days of receivables - days of payables

Example

Average accounts receivable: $25,400

Average inventory: $48,290

Average accounts payable: $37,510

Credit sales: $325,700

Cost of goods sold: $180,440.

Total purchases: $188,920

How many days are in the operating cycle? How many days are in the cash cycle?

1. The receivable turnover rate tells you the number of times during the year that money is loaned to customers. Credit sales / Average accounts receivable = 325,700 / 25,400 = 12.8228.

Receivables period = 365 days / 12.8228 = 28.46 days. This tells you that it takes customers an average of 28.46 days to pay for their purchases.

1. The inventory turnover rate indicates the number of times during the year that a firm replaces its inventory. COGS / Average inventory = 180,440 / 48,290 = 3.7366

Inventory period = 365 days / 3.7366 = 97.68 days. This means inventory sits on the shelf for 97.68 days before it is sold. That's ok for a furniture store but you should be highly alarmed if a fast food restaurant has a 98-day inventory period.

1. The accounts payable is matched with total purchases to compute the turnover rate because these accounts are valued based on the wholesale, or production, cost of each item.

Payables turnover = Total purchases / Average accounts payables = 188,920 / 37,510 = 5.0365

Payables period = 365 / 5.0365 = 72.47 days. On average, it takes 72.47 days to pay suppliers.

The operating cycle begins on the day inventory is purchased and ends when the money is collected from the sale of that inventory. This cycle consists of both the inventory period and the accounts receivable period.
Operating cycle = 97.68 + 28.46 = 126.14 days

The cash cycle is equal to the operating cycle minus the payables period. It is the number of days for which the firm must finance its own inventory and receivables. During the cash cycle, the firm must have sufficient cash to carry its inventory and receivables.

Cash cycle = 126.14 - 72.47 = 53.67 days

In this example, the firm must pay for its inventory 53.67 days before it collects the payment from selling that inventory. Controlling the cash cycle is a high priority for financial managers.

4. Evaluating Short-Term Financing Choices

```
    <p> </p>The major objectives of a short-term borrowing strategy:
```

- Ensuring that sufficient capacity exists to handle peak cash needs.
- Maintaining sufficient sources of credit to be able to fund ongoing cash needs.
- Ensuring that rates obtained are cost-effective and do not substantially exceed market averages.

Several addition factors to consider:

- Size and creditworthiness. In general, the larger (the borrower or the lender), the better terms.
- Legal and regulatory considerations. Regulated industries may impose borrowing restrictions, for example.
- Sufficient access. It's always good to have alternatives (lenders, rates etc), when it comes to borrowing.
- Flexibility of borrowing options. i.e. active maturity management.

There are active and passive borrowing strategies.

Corporate Issuers (2)

Cost of Capital - Foundational Topics

1. Cost of Capital

```
    <p> </p>Capital is a necessary factor of production, and has a cost. The providers of
```

Calculating the cost of capital is important for a firm, as this is the rate of return that must be used when evaluating capital projects. The return from the project must be greater than the cost of the project in order for it to be acceptable.

In general, a firm can finance its operations from three main sources of capital:

- Equity or common stock
- Preferred stock
- Debt

Each of these sources of capital has a cost. The cost of capital used in capital budgeting should be calculated as a weighted average, or composite, of the various types of funds a firm generally uses.

The **weighted average cost of capital** (**WACC**) is defined as the weighted average cost of the component costs of debt, preferred stock, and common stock or equity. It is also referred to as the **marginal cost of capital** (**MCC**), which is the cost of obtaining another dollar of new capital.

- w_d = the weight for debt
- w_p = the weight for preferred stock
- w_e = the weight for common stock
- r = required rate for each component
- t = the marginal tax rate

Taxes and the Cost of Capital

Interest on debt is tax deductible; therefore, to calculate the cost of debt, the tax benefit is deducted. This means that after-tax cost of debt = interest rate - tax savings (the government pays part of the cost of debt as interest is tax deductible).

There is no tax savings associated with the use of preferred stock or common stock.

Weights of the Weighted Average

The **target capital structure** is the percentage of debt, preferred stock, and common equity that a firm is striving to maintain and that will maximize the firm's stock price. Each firm has a target capital structure, and it should raise new capital in a manner that will keep the actual capital structure on target over time.

- If the target capital structure is known, it should be used.
- If not, market values of debt and stocks should be used to calculate weights. That is, the company's current capital structure is assumed to represent the company's target capital structure.
- If this is not possible, then trends in the company's capital structure or averages of comparable companies' capital structures should be used as targets.

Example

Firm A has a capital structure consisting of 40% debt, 5% preferred stock, and 55% common equity (made up of retained earnings and common stock). Firm A pays 10% interest on its debt and has a marginal tax rate of 35%. If Firm A's component cost of preferred stock is 12.5% and the component cost of common stock equity from retained earnings is 13.5%, calculate Firm A's WACC.

WACC = 0.4 x 10% (1 - 0.35) + 0.05 x 12.5% + 0.55 x 13.5% = 0.026 + 0.00625 + 0.07425 = 10.65%

2. Cost of Debt and Preferred Stock

```
    <p> </p>Capital components are the types of capital used by firms to raise fund. They
```

The **cost of debt** is defined as the cost to the firm in terms of the interest rate that it pays for ordinary debt (r_d) less the tax savings that are achieved. Interest on debt is tax-deductible and therefore to calculate the cost of debt the tax benefit is deducted.

Two methods to estimate the before-tax cost of debt (r_d) are discussed.

Yield-to-Maturity Approach

This approach uses the familiar bond valuation equation. Assuming semi-annual coupon payments, the equation is:

The six-month yield ($r_d/2$) is derived and then annualized to arrive at the before-tax cost of debt, r_d.

Debt-Rating Approach

This approach can be used if there isn't a reliable market price for a firm's debt. Based on the company's debt rating, the before-tax cost of debt is estimated by using the yield on comparably rated bonds for maturities that are a close match to those of the firm's existing debt.

For example, assume that:

- A firm's debt has an average maturity of 5 years.
- Its credit rating is AAA.
- The yield on debt with the same debt rating and similar maturity is 6%.

- The marginal tax rate is 30%.

Then the company's after-tax cost of debt is 6% x (1 - 30%) = 4.2%.

Other factors, such as debt seniority and security, may complicate the calculation, so analysts must take care when determining the comparable debt rating and yield.

Issues in Estimating the Cost of Debt

- *Fixed-rate debt versus floating-rate debt*

  Estimating the cost of floating-rate debt is difficult because the cost depends not only on the current yield but also on the future yields. The term structure of interest rates may be used to calculate an average rate.

- *Debt with option-like features*

  Be aware that some debt can have call or put options. (Valuing such debts is a topic for Level II candidates.)

- *Non-rated debt*

  The yields of a firm's debt may not be available, or a firm may not have rated bonds.

- *Leases*

  If a company uses leasing as a source of capital, the cost of these leases should be included in the cost of capital (long-term debt).

Cost of Preferred Stock

The cost of preferred stock is calculated by dividing the dollar amount of the dividend (which is normally paid on an annual basis) by the preferred stock current price.

It is important to note that tax does not affect the calculation of the cost of preferred stock, since preferred dividends are not tax deductible.

3. Cost of Common Equity

```
<p> </p>The <b>cost of common equity</b> (r<sub>e</sub>) is the rate of return stockhc
```

Estimating the cost of common equity is challenging due to the uncertain nature of the amount and timing of future cash flows.

The CAPM Approach

where R_F is the risk-free rate, $E(R_M)$ is the expected rate of return on the market, and Î²$_i$ is the stock's beta coefficient. $[E(R_M) - R_F]$ is called the **equity risk premium** (**ERP**). Both $E(R_M)$ and Î²$_i$ need to be estimated.

For example, firm A has a Î²$_i$ of 0.6 for its stock. The risk-free rate, R_F, is 5%. The expected rate of return on the market, $E(R_M)$, is 10%. The firm's cost of common equity is therefore calculated as 5% + (10% - 5%) x 0.6 = 8%.

There are several ways to estimate the equity risk premium.

- The **historical equity risk premium approach** examines the historical data of realized returns from a country's market portfolio and uses the average rate for both the market portfolio and risk-free assets. One study, cited in the textbook, found that the annualized U.S. equity risk premium relative to U.S. Treasury bills was 5.6% (geometric mean). However, there are some limitations to this approach. For example, the level of risk of the stock index and risk aversion of investors may change over time.

- The **survey approach** is a direct one: ask a panel of financial experts for their estimates and take the mean response.

- The **dividend discount model approach** (not covered in the textbook), or **implied risk premium approach**, analyzes how the market prices an index using the Gordon growth model:

  where r_e is the required rate of return on the market, D_1 is the dividends expected next period on the index, P_0 is the current market value of the equity market index, and g is the expected growth rate of the dividends.

Bond Yield Plus Risk Premium Approach

Because the cost of capital of riskier cash flows is higher than that of less risky cash flows:

This is a subjective, ad hoc procedure: bond yield is the interest rate on the firm's long-term debt, and the risk premium is a judgmental estimate (usually 3-5%). For example, suppose that ABC, Inc.'s interest rate on long-term debt is 10%. Assume the risk premium is 5%. ABC's cost of retained earnings is 10% + 5% = 15%.

4. Estimating Beta

```
<p> </p>The determination of cost of capital under the CAPM approach involves the esti
```

Estimating Beta for Public Companies

The historical Î² is the first step in the determination of the ex-ante Î².

The standard procedure for estimating Î²s is to regress stock returns (R_i)against market returns (R_{mi} = a + b R_m, where b is the slope of the regression and corresponds to the Î² of the stock. It measures the riskiness of the stock.

To estimate Î²$_i$ an analyst needs to make several choices:

- Which index should be used to represent the market portfolio? The S&P 500 is a traditional choice to represent U.S. equities.
- What should be the length of data period and the frequency of observations?

According to Blume, there is a tendency of betas to converge towards the mean of all betas.

$$\text{Adjusted beta} = (2/3)\ (\text{Unadjusted beta}) + (1/3)\ (1.0)$$

It corrects the estimated Î² for its tendency to revert to 1. It adjusts Î² in such a way that it is closer to the expected Î² in the future.

Estimating Beta for Thinly Traded and Nonpublic Companies

If we are thinking of a new company for a single project, we will have no historical records to go by. We would then compute the Î² of companies of the same size and about the same lines of business and after making necessary adjustments, take this as the Î² for the firm. The **pure-play method** can be used to take a comparable publicly traded company's beta and adjust it for financial leverage differences.

The Î² that we impute to a project is likely to undergo changes with changes in the capital structure of the company. If the company is entirely equity-based, its Î² is likely to be lower than it would be if it undertakes borrowing.

Let us call the Î² of a firm that is levered "levered Î²" and that of a firm on an all-equity structure "unlevered Î²."

Î² of a levered firm:

where:

Î²$_{L}$ = Î² of a levered firm

Î²$_{U}$ = Î² of an unlevered firm

T = tax rate

D = component of debt in capital structure

E = component of equity in capital structure

If the Î² of a firm is available and that Î² has been estimated on the premise that the firm is unlevered, we can now ascertain the Î² of the firm should it undertake some borrowing by using the following formula:

Î² of an unlevered firm:

In the same way, given the Î² of a firm which is already levered, we can ascertain what its Î² would be if it chooses an all-equity structure. This also means that if the target firm has leverage different from the structure assumed in estimating the levered Î², this can first be converted into an unlevered Î² and then re-converted into a levered Î² using the leverage parameters relevant to the firm.

As a first step, we have to identify firms that reasonably resemble the project for which the beta is to be estimated. The stock Î² of these firms is then taken. Their respective leverage position (ratio of debt to equity) is also considered. After duly adjusting the tax factor and applying the above formula, we can determine the proxy Î² of the project assuming that it is unlevered.

The procedure is illustrated below:

Suppose there are three firms, P, Q, and R, which closely resemble project X (that is to be embarked upon). The stock betas of the three firms are taken and found to be 2.73, 2.23, and 1.73 respectively. The ratio of debt to equity for the three firms averages to 0.67. The marginal tax rate is 36%.

The average stock Î² works out to 2.23. Translating these numbers into the formula for unlevered firms, we get: Î²$_{U}$ = Î²$_{L}$ / (1 + (1 - T)(D/E)) = 2.23/(1+0.64 x 0.67) = 1.56.

This suggests that on an all-equity basis the Î² of the project would be 1.56. Now, if the project is proposed to be financed by 50% equity and 50% debt, we can modify the above Î² by applying the formula for levered firms:

Î²$_{L}$ = Î²$_{U}$ (1 + (1 - T) D/E) = 1.56 (1 + 0.64 x 0.5/0.5) = 2.56

So, on a 1:1 debt equity ratio, the Î² will be 2.56. This Î² can be used now for determining the cost of equity for the project and its weighted average cost of capital, to make a more meaningful appraisal.

5. Flotation Costs

```
<p> </p><b>Flotation costs</b> are the costs of issuing a new security, including the
```

The amount of flotation costs is generally quite low for debt and preferred stock (often 1% or less of the face value), so we ignore them here. However, the flotation costs of issuing common stocks may be substantial, so they must be accounted for in the WACC. Generally, we calculate this by reducing the proceeds from the issue by the amount of the flotation costs and recalculating the cost of equity.

Example 1

XYZ is contemplating issuing new equity. The current price of their stock is $30 and the company expects to raise its current dividend of $1.25 by 7% indefinitely. If the flotation cost is expected to be 9%, what would be the cost of this new source of capital?

Cost of external equity = (1.25 x 1.07) / (30 x (1 - 0.09)) + 0.07 = 11.9%

Without the flotation cost, the cost of new equity would be (1.25 x 1.07) / 30 + 0.07 = 11.46%.

Note that flotation costs will always be given, but they may be given as a dollar amount or as a percentage of the selling price.

This is a typical example found in most textbooks. One problem with this approach is that the flotation costs are a cash flow at the initiation of the project and affect the value of any project by reducing the initial cash flow. It is not appropriate to adjust the present value of the future cash flows by a fixed percentage. An alternative approach is to make the adjustment to the cash flows in the valuation computation.

Example 2

Continue with the above example. Assume that XYZ is going to raise $10 million in new equity for a project. The initial investment is $10 million and the project is expected to produce cash flows of $4.5 million each year for 3 years.

Ignoring the flotation cost of issuing new equity, the NPV of the project will be $-10 + 4.5/1.1146^1 + 4.5/1.1146^2 + 4.5/1.1146^3 = \$0.9093$ million.

Now consider the flotation cost of 9%. The NPV, considering the flotation costs, is 0.9093 - 0.9 = $0.0093 million.

However, if we use the "typical" approach, the NPV. considering the flotation costs, will be $-10 + 4.5/1.119^1 + 4.5/1.119^2 + 4.5/1.119^3 = \$0.8268$ million.

Capital Structure

1. Capital Structure and Company Life Cycle

```
<p> </p>A company's stage in the life cycle, its cash flow characteristics, and its ak
```

Generally speaking, as companies mature and move from start-up, through growth, to mature, their business risk declines as operating cash flows turn positive with increasing predictability, allowing for greater use of leverage at more attractive terms.

Start-Ups

A company at this stage is a cash consumer, with negative revenues, negative cash flows and high business risks. Its main funding source is private equity, with almost no debt.

Growth Businesses

Revenue growth is high, and cash flows may turn positive and predictable. Asset-backed debt may be needed to finance its revenue growth, but equity remains the predominant source of capital.

Mature Businesses

Revenue growth slows down or even starts to decline. Cash flows are stable and predictable. Debt financing is preferred over higher-cost equity financing, partly due to the tax-deductibility of interest expense.

De-leveraging may occur, and share buybacks may be executed if the company has enough cash and the share price is right.

Unique Situations

Some businesses, such as real estate and other capital-intensive businesses, employ a lot of leverage regardless of their development stage. Some mature, large businesses may use little debt because they don't own large amount of fixed assets. Companies in cyclical industries and capital light businesses tend to have little debt in their capital structures.

2. Modigliani-Miller Propositions

```
    <p> </p>Modigliani and Miller (MM) proved, under a very restrictive set of assumptions
```

Assumptions:

- Homogenous expectations: Investors agree on a given investment's expected cash flows.
- Perfect capital markets: There are no brokerage costs, no taxes and no bankruptcy costs;
- Risk-free rate: Investors can borrow and lend at the same rate as corporations;
- No agency costs: Managers always act to maximize shareholder wealth.
- Independent decisions: Financing and investment decisions are independent of each other.

MM Propositions without Taxes

There are two almost equivalent ways of looking at the same problem:

- Is the value of a levered firm ($V_L = D + E$) different from the value of an unlevered firm (V_U)? (MM1)
- Is the WACC of a levered firm different from the WACC of an unlevered firm? (MM2)

The M&M results are based on absence of arbitrage. Assume there are two firms identical in investments but different capital structure. If they are all equity financed, they should have identical required rate of returns. But, if one firm is levered (V_L) and the other is unlevered (V_U), V_U has to be equal to V_L otherwise an arbitrage profit could be made!

MM Proposition 1: A firm's value is determined by its assets, not its capital structure. This implies that there is no optimal capital structure!

WACC equals the required rate of return of the firm's assets (ROA), which is determined by the firm's investment decisions, not capital structure.

MM Proposition 2: WACC is invariant to capital structure under the previous assumptions.

The cost of equity of a levered firm is equal to the cost of equity of an unlevered firm plus a risk premium which depends on the degree of financial leverage: $r_e = r_0 + (r_0 - r_d)(D/E)$, where r_0 is the cost of all-equity firm.

Reductions in capital costs as a result of using more lower cost debt (r_d) are exactly offset by increases in the cost of levered equity (r_e) due to added financial risk.

The equity's beta, Î²$_e$, rises as more debt is used: Î²$_e$ = Î²$_a$ + (Î²$_a$ - Î²$_d$)(D/E)

The important thing here is not the formulas presented in the textbook, but the intuition of the model. Capital structure irrelevance theory is all about slicing a pie: the size of the pie represents the value of a firm. It is determined by the size of the pie pan, not how it is sliced. With a given pie pan, the size of the pie will be always the same no matter how you slice it. Similarly, the value of a firm depends on the firm's assets, not its capital structure. With a given asset base, the value of the firm remains the same no matter how the firm finances its investments. Capital structure only affects the distribution of the firm's value among debt holders and equity holders.

MM Propositions with Taxes

The value of a firm is still determined by the firm's assets, which generate cash flows. By levying taxes, the government joins debt-holders and shareholders to share the cash flows (and thus, the value) of the firm.

MM Proposition 1 with Taxes:

- The value of an unlevered firm is equal to EBIT (1-T) capitalized at the cost of equity. V_U = EBIT (1 - T)/r_e.
- The value of a levered firm is equal to the value of an unlevered firm of the same risk class, plus the value of the interest tax savings capitalized at the cost of debt: $V_L = V_U + T \times D$.

The deductibility of interest expense favours the use of debt financing for companies. Since this tax shelter accrues to shareholders, using debt will increase the value of the entire equity. The more the firm borrows, the greater the tax shelter, and thus the higher the share value of the stock. Therefore, if other MM assumptions hold, firms will maximize debt in their capital structures: the optimal capital structure in a tax world will be infinitely close to 100% debt!

MM Proposition 2 with Taxes:

The cost of equity of a levered firm is equal to the cost of equity of an unlevered firm plus a risk premium which depends on both the degree of leverage and the corporate tax rate: $r_e = r_0 + (r_0 - r_d)\ (1-T)\ (D/E)$

As debt-equity ratio (thus the financial risk) rises, so does the cost of equity. However, the weight of equity declines as the firm uses more debt. The reduction in the cost of debt can more than offset the effect of rising cost of equity. Therefore the cost of capital (WACC) declines as the firm uses more debt.

The existence of (higher) personal tax rate on interest income than on dividend income, however, reduces the advantage of debt financing to a company. In the Miller model, debt financing may add or lower value.

Costs of Financial Distress

More leverage can magnify financial losses, which can put a company into **financial distress**.

- Direct costs are the various costs of filing for bankruptcy, hiring lawyers and accountant, etc.
- Indirect costs include higher borrowing costs (Lenders charge higher interest rates to firms in financial trouble.), a loss of employee morale and productivity, agency costs of debt, etc.

The threat of bankruptcy and the bankruptcy costs discourage firms from pushing their use of debt to excessive levels. Firms whose earnings are more volatile, all else equal, face a greater chance of bankruptcy and should use less debt than more stable firms.

3. Optimal and Target Capital Structure

```
    <p> </p>The <b>optimal capital structure</b> is the mix of debt, preferred stock and c
```

The gain from the tax shield on debt is offset by financial distress costs. As a firm continues to add more debt, financial distress costs start to rise slowly at first and then more rapidly, increasing the effective cost of debt and thus the WACC. Therefore, the WACC will first fall, then bottom out and finally start to rise. The conclusion is: in a world with taxes and financial distress costs, there is an optimal capital structure where the WACC is minimized and the share value of the stock is maximized.

The **static trade-off theory** suggests that the optimal capital structure is reached at the point where marginal distress costs exceed the marginal tax benefit from adding debt in the MM model.

The formula looks as follows:

$$V_L = V_U + T \text{ x } D - PV \text{ (costs of financial distress)}$$

An optimal capital structure exists that just balances the additional gain from leverage against the added financial distress costs.

The WACC falls initially because of the tax advantage of debt. Beyond the point D/*E*, it begins to rise because of financial distress costs.

At any point in time, management has a specific **target capital structure** in mind:

- If the debt ratio is below the target, expansion capital may be raised by issuing more debt.
- If the debt ratio is above the target, the firm may raise expansion capital by retaining earnings or issuing new equity.

4. Factors Affecting Capital Structure Decisions

```
<p> </p>Is there an optimal capital structure?
```

In financial terms, debt is a good example of the proverbial two-edged sword. Astute use of leverage (debt) increases the amount of financial resources available to a company for growth and expansion. The assumption is that management can earn more on borrowed funds than it pays in interest expense and fees on these funds. A company considered too highly leveraged (too much debt versus equity) may find its freedom of action restricted by its creditors and/or may have its profitability hurt as a result of paying high interest costs.

Capital Structure Policies and Target Capital Structure

Companies should set guidelines to establish the borrowing limit based on the company's risk appetite and its ability to support debt. Such policies and guidelines tend to be debt-oriented. For example, debt/equity < 0.5. There is no magic proportion of debt that a company can take on.

The debt-equity relationship varies according to industries involved, a company's line of business and its stage of development. A common goal is to finance at the lowest cost of capital. A scenario analysis, given a particular combination of assumptions, can be used to assess this point.

Not all companies have capital structure policies. In some industries, a capital structure mix may be dictated by regulators.

Debt Ratings

The static tradeoff theory argues that a value-maximizing firm will balance the value of interest tax shields and other benefits of debt against the costs of financial distress and other costs of debt to determine an interior optimal leverage for the firm. The discrete costs/benefits associated with debt ratings should be included in capital structure decisions.

Debt ratings are formal risk evaluations by credit-rating agencies of a company's ability to repay principal and interest on debt obligations.

The influence of debt ratings on capital structure is economically significant. Corporate financial managers care about debt ratings when making capital structure decisions. In their survey of CFOs, Graham and Harvey (2001) find that debt ratings are the second highest concern for CFOs when considering debt issuance: debt ratings are ranked higher than many of the factors suggested by traditional capital structure theories (such as the "tax advantage of interest deductibility").

A research found that firms near a rating upgrade or downgrade issue less debt relative to equity than firms not near a rating change. The result persists in the context of empirical tests of the pecking order and tradeoff capital structure theories.

Market Value vs. Book Value

Although market value of equity and debt should be used, in practice, however, book value is often used instead.

- Market values can fluctuate substantially without a corresponding change in the company's borrowing ability.
- Management is more concerned to invest the capital in different projects, and less concerned about prevailing share price.
- Lenders, investors and rating agencies generally use book value for calculating various measures.

Financing Capital Investments

Financing decisions are typically tied to investment spending. Considerations include: Are assets suitable to leverage? Are maturity structures and cash flows match between assets and liabilities? When considering capital structures of companies in different countries, are revenues and debt in the same currency?

Market Conditions

When considering to raise capitals, managers pay very close attention to the share price of the company's and market interest rates on its loans. Can their debt or equity issue be included in a benchmark index? What will the risk premium be for a new debt issue? Any significant impact on incremental costs of borrowing? Any macro-economic factors such as interest rates and inflation rates to consider?

Information Asymmetries and Signaling

Information can be used to change the cost of financing.

One of MM's assumptions is that investors and managers have the same information about the firm's prospects. This is called **symmetric information**. In reality, managers often have better information than outside investors. This is called **asymmetric information**. Managers often have better information than outside investors, and this has an important effect on the optimal capital structure.

One would expect a firm with very favorable prospects to try to avoid selling stock and, rather, to raise any required new capital by other means, including using debt beyond the normal target capital structure. Why? Assume that managers act in the best interest of shareholders.

- If the firm sells new stock, then as the firm expands successfully, the stock price will rise and the new stockholders will make a fortune.
- Had the company not sold more stocks, the current shareholders would have avoided sharing the success with new shareholders.
- So there is a motivation on the firm's part to avoid selling new stocks when it has exceptional prospects.

On the other hand, a firm with unfavorable prospects would want to sell stock which would mean bringing new investors to share the losses.

The **pecking order hypothesis** says that managers prefer the following ordering of financing:

- Retained earnings (avoid investor skepticism).
- Debt.
- Equity.

Therefore, a stock issue sets off a negative signal (the firm's prospects as seen by its management are not bright), while using debt is a positive, or at least a neutral, signal.

As a result, companies try to avoid having to issue stock by maintaining a **reserve borrowing capacity**, and this means using less debt in "normal" times than the MM trade-off theory would suggest. Why? If you don't have reserve borrowing capacity, and you have to issue new stocks for good projects, your stock price will be penalized: investors cannot see your good project (asymmetric information), and all they see is your negative signal - bad companies sell stocks.

Agency Costs

Equity agency problem is that managers might:

- use corporate funds for non-value maximizing purposes (perks, acquisitions, value-destroying growth), or;
- seek low risk due to undiversified interest in firm.

The **agency cost of equity** is the cost associated with monitoring a companyâ€™s management by the company shareholders when the shareholders believe that the management is diverging from working towards obtaining shareholderâ€™s interests. The problem is most significant in large firms with diffuse stockholders where management ownership is low.

To mitigate this risk, shareholders can take actions such as requiring audited financial statements, holding an annual meeting, using noncompete employment contracts and insurance to guarantee performance.

5. Stakeholder Interests

```
<h4>Shareholder vs. Stakeholder Theory</h4>
```

The shareholder theory states that the most important responsibility of managers is to serve the interests of shareholders.

The stakeholder theory states other parties' interest should be considered too. These other parties include the company's customers, suppliers, employees, managers, debt-holders and shareholders.

Capital structure decisions impact stakeholders differently. Increased leverage increases risk for all stakeholders, but could benefit the group of shareholders only.

Debt vs. Equity Conflict

The agency cost of debt is the conflict that arises between shareholders and debtholders of a public company.

Debtholders are typically interested in safe investments. Their potential return is limited. Equityholders, however, have a much higher upside potential. They often prefer higher leverage that offer them greater return potential.

Seniority and Security: Secured and senior debt lenders can recover more in a default scenario than unsecured or junior debt owners. They can therefore better tolerate risky decisions by management.

Long-term vs. Short-Term Debt: the longer term, the greater the equity/debt conflict.

Safeguards for Debtholders: Debt covenants are the primary tools for debtholders. They state what the borrower must do (positive covenants) and cannot do (negative covenants). Companies also want to keep good track record for future borrowings. Substantial financial distress costs is also a consideration.

Preferred Shareholders

Preferred shares have debt and equity-like characteristics. Preferred shareholders provide long-term capital without maturity date and covenant protection. They are vulnerable to decisions that increase financial leverage and risk.

Private Equity Investors/Controlling Shareholders

Majority shareholders may have objectives that conflict with minority shareholders. They may take a short-term view on financing, or long-term view on strategic growth. They can be different from what the minority shareholders want.

Bank and Private Lenders

Different lenders have different agendas. Public market debtholders may not hold the debt to maturity and may make decisions on bond prices.

Bank and private lenders generally hold company debt to maturity. They usually have direct access to firm management and non-public company information, which decreases information asymmetry in theory. Bank lending policies are generally conservative. Private lenders vary widely in risk appetite, approach, behavior, and relationships with companies to whom they have provided capital.

Other Stakeholders

Customers and Suppliers: Suppliers are short-term creditors. They both naturally prefer long-term stability of the company.

Employees: Most employees in most businesses prefer the company's stability and growth to equity ownership.

Management and Directors: Compensation can be used to motivate managers to work hard to maximize the shareholder value, which typically results in debt/equity conflict. However, such compensation can be excessive, or motivate risk-taking behavior.

Regulator and Government: They are often key stakeholders. For example, regulators may set the max price charged by utility companies. Certain solvency levels must be kept for financial institutions.

Measures of Leverage

1. Business Risk and Operating Leverage

```
<b>Leverage</b><p> </p>
```

Leverage is the extent to which fixed costs are used in a company's cost structure.

- **Operating leverage** is the extent to which fixed operating costs (e.g., depreciation, rent) are used in a firm's operations.
- **Financial leverage** is the extent to which fixed-income securities (debt and preferred stock) are used in a firm's capital structure.

Leverage affects a firm's risk, as it can magnify earnings both up and down. The bigger the leverage, the more volatile the firm's future earnings and cash flows, and the greater the discount rate applied in the firm's valuation (by bondholders and stockholders).

Business Risk and its Components

Business risk is the uncertainty (variability) about projections of future operating earnings. It is the single most important determinant of capital structure. If other elements are the same, the lower a firm's business risk, the higher its optimal debt ratio.

Business risk is the combined risk of sales and operations risks.

- The **sales risk** is the uncertainty regarding the price and quantity of the firm's goods and services. If the demand for and the price of a firm's goods and services are stable, its sales risk is considered low.

- The **operating risk** is the uncertainty caused by a firm's operating cost structure. If a high percentage of operating costs are fixed costs, operating risk is considered to be high.

In general, management has more opportunity to manage and control operating risk than sales risk.

Operating Risk

A company that has high operating leverage is a company with a large proportion of fixed input costs, whereas a company with largely variable input costs is said to have low operating leverage (due to its small amount of fixed costs).

A company with a high degree of operating leverage that has a small change in sales will experience a large change in profits and rate of return. This is due to the fact that because the company has a large fixed cost component, any increase in sales will cause an even greater increase in net income, since the fixed costs have already been incurred.

In many respects operating leverage is determined by technology. High (low) operating leverage is usually associated with capital (labor) intensive industries.

The **degree of operating leverage** (**DOL**) is defined as the percentage change in EBIT (operating income) that results from a given percentage change in sales. It measures the impact of a change in sales on EBIT.

Here Q is the number of units, P is the average sales price per unit of output, V is the variable cost per unit, F is fixed operating cost, S is sales in dollars, and VC is total variable costs.

P - V is referred to as the **per unit contribution margin**, which is the amount that each unit contributes to covering fixed costs. S - VC is called the **contribution margin**.

For example, assume that a firm has sales of $100,000, variable costs of $50,000, and fixed costs of $20,000. Its DOL is (100,000 - 50,000) / (100,000 - 50,000 - 20,000) = 1.67.

2. Financial Risk and Financial Leverage

```
    <b>Financial risk</b> is the additional risk placed on the common stockholders as a re
```

The questions are: Is the increased rate of return sufficient to compensate shareholders for the increased risk? What is the optimal financial structure to maximize stock price and the firm's value?

Financial risk depends on two factors:

- Cash flow volatility. The more volatile (stable) a firm's cash flows, the higher (lower) the financial risk.
- Financial leverage. The higher the financial leverage, the higher the financial risk.

As a general proposition, financial leverage raises the expected rate of return, but at the cost of increased financial risk (and thus total risk). So, you are faced with a trade-off: if you use more financial leverage, you increase the expected rate of return, which is good, but you also increase risk, which is bad.

The degree of financial leverage (DFL) measures the financial risk.

It shows how a given percentage change in EBIT per share will affect EPS.

The equation above is developed as follows:

where:

I = interest paid

T = marginal tax rate

N = number of shares outstanding

I is a constant so Î”I = 0, therefore:

Now the percentage change in EPS is the change in EPS divided by the original EPS, which is:

DFL is defined as the percentage change in earnings per share (EPS) divided by the percentage change in EBIT.

Consider a company with EBIT = $100,000 and interest = $20,000. Its DFL = 100,000 / (100,000 - 20,000) = 1.25. Therefore, a 100% increase in EBIT would result in a 125% increase in EPS.

Unlike operating leverage, the degree of financial leverage is most often a choice by the company's management. Companies with a higher ratio of tangible assets to total assets may have higher degrees of financial leverage because lenders may feel more secure that their claims would be satisfied in the event of a downturn.

3. Total Leverage and Breakeven Points

```
   Operating leverage (first-stage leverage) affects EBIT, while financial leverage (secc
```

Both operating leverage and financial leverage contribute to the risk associated with a firm's future cash flows. The **degree of total leverage** (**DTL**) combines DOL and DFL, and measures the impact of a given percentage change in sales on EPS.

If both DOL and DFL are high, a small change in sales leads to wide fluctuations in EPS.

The **breakeven point** is the volume of sales at which total costs equal total revenues, causing net income to equal zero: PQ - VQ - F - I = 0. The breakeven number of units, Q_{BE}, is:

The **operating breakeven point** is the number of outputs at which revenues = operating costs: $PQ_{OBE} = VQ_{OBE} + F$. Q_{OBE} is:

Consider a project where the fixed costs are $10,000, the variable costs are $2 per unit, the selling price per unit is $4, and the interest expense is $1,000. The breakeven sales quantity is 11,000 / (4 - 2) = 5,500 units and the operating breakeven sales quantity is 10,000 / (4 - 2) = 5,000 units.

In general, the farther unit sales are from the breakeven point for high-leverage companies, the greater the magnifying effect of this leverage.

4. The Risks of Creditors and Owners

```
    Creditors and stockholders bear different risks because they have different rights anc
```

- Creditors get pre-determined returns and principals back when due, regardless of the profitability of the firm.
- Stockholders get what is left over after all expenses, including interest paid to creditors, have been paid. In exchange for this uncertainty of returns (which is the risk that stockholders face), stockholders exercise decision-making power over the business. They can also declare what portion of the business earnings they will take out as dividends.

The use of greater amounts of debt in the capital structure can raise both the cost of debt and the cost of equity capital.

- The higher the percentage of debt, the riskier the debt, hence the higher the interest rate creditors will charge.
- In general, increasing the use of debt increases the expected rate of return, but more debt also means that the firm's stockholders must bear more risk. The cost of equity capital must be higher now than before.

Creditors have priority over stockholders in a bankruptcy proceeding. When a firm files for bankruptcy, its leverage often determines the final outcome.

- Reorganization. A firm with high financial leverage uses bankruptcy laws and protection to reorganize its capital structure to remain in business.
- Liquidation. A firm with high operating leverage cannot use such bankruptcy protection, as it would not reduce operating costs. This means that the firm's business is terminated, all the assets are sold and distributed to the holders of claims on the organization, and no corporate entity should survive. Stockholders generally lose all value in such a case, and creditors typically receive a portion of their capital.

Equity Investments

Equity Investments (1)

Market Organization and Structure

1. The Functions of the Financial System

```
    <b>Helping People Achieve Their Purposes Using the Financial System</b><p> </p>
```

The financial system helps people:

- *Save money for the future.* Saving here means buying notes, CDs, bonds, stocks, mutual funds, or real estate assets.
- *Borrow money for current use.* This is the opposite of the first purpose (above). Individuals, companies, and governments may need money to spend now (consumption, investment, paying taxes, expenses, etc).
- *Raise equity capital.* Companies can sell ownership rights to raise the equity capital they need.
- *Manage risks.* People can use financial contracts to offset risks.
- *Exchange assets for immediate (in spot markets) and future (in futures markets) deliveries.*
- *Trade on information.* **Information-motivated traders** can (or they believe they can) use the financial

system to earn a return in excess of the fair rate of return because they have information whose value declines over time (as it becomes recognized by other market participants).

Determining Rate of Return

The price in the financial system is the rate of return. It is the result of interaction of the broad forces of supply and demand.

There are as many different prices (rates of return) as there are different types of assets in the financial system. For example, equities have higher rates of return than T-bills. All of the rates are determined in the financial system.

Prices rapidly adjust to new information. The prevailing price is fair because it reflects all available information regarding the asset.

Capital Allocation Efficiency

In the financial markets investors distinguish good firms from bad firms. This lets the market channel capital to good firms and away from problem firms.

Timely and accurate information is available on the price and volume of past transactions and the prevailing bid-price and ask-price. Such information facilitates the rapid flow of capital to its highest value uses.

2. Assets and Contracts

There are many different ways one can classify assets and contracts. The most common w

Fixed-Income Investments

These have a contractually mandated payment schedule. Their investment contacts promise specific payments at predetermined times. Investors who acquire fixed-income securities are really lenders to the issuers. Specifically, you lend some amount of money (the principal) to the borrower. In return, the borrower promises to make periodic interest payments and to pay back the principal at the maturity of the loan.

Bonds, notes, bills, CDs, commercial paper, repo agreements, loan agreements, and mortgages are examples of fixed-income investments.

Preferred stock is classified as a fixed-income security because its yearly payment is stipulated as either a coupon (e.g., 5% of the face value) or a stated dollar amount. Although preferred dividends are not legally binding (as are the interest payments on a bond), they are considered practically binding because of the credit implications of a missed dividend.

Equities

Equities differ from fixed-income securities because their returns are not contractual. They represent residual ownership in companies after all claims-including any fixed-income liabilities of the company - have been satisfied.

Common stocks represent ownership of a firm. Owners of the common stock of a firm share in the company's successes and problems.

A **warrant** allows the holder to purchase a firm's common stock from the firm at a specified price for a given time period. It provides the firm with future common stock capital when the holder exercises the warrant.

Pooled Investments

Rather than directly buying an individual stock or bond, you may choose to acquire these investments indirectly by buying shares in an investment company that owns a portfolio of individual stocks, bonds, or a combination of the two. People invest in pooled investment vehicles to benefit from the investment management services of their managers. Examples of these pooled investments include money market funds, bond funds, stock funds, balanced funds, etc.

Currencies

The currency market is a worldwide decentralized over-the-counter financial market for the trading of currencies. Market participants include commercial banks, central banks, retail brokers, etc.

Contracts

Financial contracts include the following:

- **Forward contracts** allow buyers and sellers to arrange for future sales at pre-determined prices. They represent a commitment to buy or sell.
- **Futures contracts** are standardized forward contracts guaranteed by clearing houses. They are traded on a futures exchange.
- **Swap contracts** are derivative securities in the form of agreements between two counterparties to exchange cash flows over a period of time, depending on the values of specified market variables.
- **Options** are rights to buy or sell an underlying instrument at a specified price within a designated time period.

Commodities

Commodities include agricultural products, energy, metals, etc. Commodities complement investment opportunities offered by shares of corporation that extensively use these raw materials in their production processes.

Real Assets

Real assets include tangible assets such as real estate, airplanes, machinery, or lumber stands. They are often illiquid and have high transaction costs compared to stocks and bonds.

3. Financial Intermediaries

```
    <b>Financial intermediaries</b> are institutions that function as the line of communic
```

Brokers, Exchanges, and Alternative Trading Systems

A **broker** executes trade orders on behalf of a customer. A **block broker** helps fill larger orders.

Investment banks help their corporate clients raise capital by issuing shares or bonds. They also help their corporate clients identify and acquire other companies.

An **exchange** is like a market where stocks, bonds, options and futures, and commodities are traded. Most exchanges offer different categories of membership and regulate their members' behavior when trading on the exchange. They also regulate the issuers that list their securities on the exchange.

Alternative trading systems (ATSs) are non-exchange trading venues that bring together buyers and sellers of securities. ATSs do not exercise regulatory authority over their subscribers and do not discipline subscribers other than by exclusion from trading. For example, an **electronic communication network** (ECN) connects major brokerages and individual traders so that they can trade directly between themselves without having to go through a middleman. **Dark pools** are ATSs that don't display their orders (which are usually very large).

Dealers

A **dealer** trades for its own accounts. Individual dealers provide liquidity to investors by trading the securities for themselves. They buy or sell with one client and hope to do the offsetting transaction later with another client.

In practice, most brokerages are in fact broker-dealer firms. That is, as a broker, the brokerage conducts transactions on behalf of clients, and, as a dealer, it trades on its own account.

In the U.S. most broker-dealers must register with the SEC.

Securitizers

Securitization is a structured finance process that distributes risk by aggregating assets in a pool (often by selling assets to a special purpose entity) then issuing new securities backed by the assets and their cash flows. The securities are sold to investors who share the risk and reward from those assets.

In most securitized investment structures, the investors' rights to receive cash flows are divided into "tranches": senior tranche investors lower their risk of default in return for lower interest payments while junior tranche investors assume a higher risk in return for higher interest.

Financial intermediaries securitize many assets, such as mortgages, car loans, credit card receivables, and banks loans.

Depository Institutions and Other Financial Corporations

They accept monetary deposits from savers and investors, and then lend these deposits to borrowers. Both the depositors and borrowers benefit from the services they provide. Depository institutions also provide other services, such as transaction services, credit services, etc.

Insurance Companies

Insurance involves pooling funds from many insured entities (e.g., policyholders) in order to pay for relatively uncommon but severely devastating losses which can occur to these entities. The insured entities are therefore protected from risk for a fee. In other words, risks are transferred from these entities to the insurance company. The insurance company connects customers who want to insure against risks with investors who are willing to bear those risks.

Insurance companies make money in two ways:

- Through underwriting, the process by which insurers select the risks to insure and decide how much in premiums to charge for accepting those risks;
- By investing the premiums they collect from insured parties.

Arbitrageurs

Arbitrage is the practice of taking advantage of a price difference between two or more markets (the law of one price). Simply put, it is the possibility of a risk-free profit at zero cost.

Arbitrage is not simply the act of buying a product in one market and selling it in another for a higher price at some later time. The transactions must occur *simultaneously* to avoid exposure to market risk, or the risk that prices may change on one market before both transactions are complete.

Arbitrage has the effect of causing prices in different markets to converge.

Settlement and Custodial Services

A **clearinghouse** is a financial institution that provides clearing and settlement services for financial and commodities derivatives and securities transactions.

A clearinghouse stands between two clearing firms (also known as member firms) and its purpose is to reduce the risk of one (or more) clearing firm failing to honor its trade settlement obligations. A clearinghouse reduces the settlement risks by netting offsetting transactions between multiple counterparties, by requiring collateral deposits (a.k.a. margin deposits), by providing independent valuation of trades and collateral, by monitoring the creditworthiness of the clearing firms, and, in many cases, by providing a guarantee fund that can be used to cover losses that exceed a defaulting clearing firm's collateral on deposit.

Depositories or custodians hold securities on behalf of their clients.

4. Positions

```
   A <b>long</b> position is owning or holding securities or contracts. For example, an c
```

A **short sale** allows investors to profit from a decline in a security's price if they believe the security is overpriced. In this procedure an investor (the seller) borrows shares of stock from another investor (the lender) through a broker and sells the shares. The lender keeps the proceeds of the sale as collateral. Later, the investor (the short seller) must repurchase the shares in the market in order to return the shares that were borrowed (*covering the short position*) to the lender. If the stock price has fallen, the shares will be repurchased at a lower price than that at which they were initially sold, and the short seller reaps a profit equal to the drop in price times the number of shares sold short.

For options, to be long means you are the buyer of the option. To be short means you are the seller of the option. Since the put option contract holder (*long*) has the right to sell the underlying to the option writer, he or she is actually *short* the underlying instrument.

The profit in short selling is limited to the value of the security but the loss is theoretically unlimited. In practice, as the price of a security rises, the short seller will receive a margin call from the broker, demanding that the short seller either cover his short position (by purchasing the security) or provide additional cash in order to meet the margin requirement for the security (which effectively places a limit on the amount that can be lost).

Leveraged Positions

Margin transactions occur when investors who purchase stocks borrow part of the purchase price of the stock from their brokers and leave purchased stocks with the brokerage firm because the securities are used as collateral for the loan. The interest rate of the margin credit charged by the broker is typically 1.5% above the rate charged by the bank making the loan. The bank rate (the **call money rate**) is normally about 1% below the prime rate. The market value of the collateral stock minus the amount borrowed is called the **investor's equity**.

Investors can achieve greater upside potential, but they also expose themselves to greater downside risk. The leverage equals 1/margin%.

Buying stocks on margin increases the investment's financial risk and thus requires a higher rate of return.

- **Percentage margin.** The ratio of the net worth or "equity value" of the account to the market value of the securities.
- **Maintenance margin.** The required proportion of equity to the total value of the stock. It protects the broker if the stock price declines.
- **Margin call.** If the percentage margin falls below the maintenance margin, the broker issues a margin call requiring the investor to add new cash or securities to the margin account. If the investor fails to provide the required funds in time, the broker will sell the collateral stock to pay off the loan.

Example

Suppose an investor initially pays $6,000 toward the purchase of $10,000 worth of stock ($100 shares at $100 per share), borrowing the remaining from the broker. The maintenance margin is set at 30%. The initial percentage margin is 60%. If the price of the stock falls to $57.14, the value of his stock will be $5,714. Since the loan is $4,000, the percentage margin now is (5,714 - 4,000) / 5714 = 29.9%. The investor will get a margin call.

When investors acquire stock or other investments on margin, they are increasing the financial risk of the investment beyond the risk inherent in the security itself. They should increase their required rate of return accordingly.

Return on margin transaction = (change in investor's equity - interest - commission) / initial investor's equity.

Example

Suppose an investor is bullish (optimistic) on Microsoft stock, which is currently selling at $100 per share. The investor has $10,000 to invest and expects the stock to go up in price by 30% during the next year. Ignoring any dividends and commissions, the expected rate of return would thus be 30% if the investor spent only$10,000 to buy 100 shares.

What will happen if the investor borrows $10,000 from his broker and invests it in the stock (along with his own $10,000)? Assume the interest rate is 9% per year.

- If the stock goes up 30%, his 200 shares will be worth $26,000. Paying off $10,000 of principal and interest on the margin load leaves $15,100. The rate of return, therefore, will be ($15,100 - $10,000) / $10,000 = 51%. Good investment, huh?
- Borrowing to invest, however, magnifies the downside risk. Suppose the stock actually goes down by 30%: his 200 shares of stock are worth $14,000 now. After paying off $10,900 he is left with only $3,100. The result is a disastrous rate of return of -69%!
- If there is no change in the stock price, he will lose 9%, the cost of the loan.

5. Orders

```
   Orders are instructions to trade. They always specify instrument, side (buy or sell),
```

- **Bid price**: the highest price that a buyer wants to pay for the instrument. The best bid is the highest bid in the market.
- **Ask price**: the lowest price a seller is willing to accept for the instrument. Also called **offer price**. The best offer is the lowest in the market.
- **Bid-ask spread**: the difference between the best bid and the best offer.

Orders usually also provide several other instructions.

Execution Instructions

These indicate how to fill the order.

Market orders are simple buy or sell orders that are to be executed immediately at current market prices. They provide immediate liquidity for someone willing to accept the prevailing market price.

A **limit order** is an order that sets the maximum or minimum at which you are willing to buy or sell a particular stock. For instance, if you want to buy stock ABC, which is trading at $12, you can set a limit order for $10. This guarantees that you will pay no more than $10 to buy this stock. Once the stock reaches $10 or less, you will automatically buy a predetermined amount of shares. On the other hand, if you own stock ABC and it is trading at $12, you could place a limit order to sell it at $15. This guarantees that the stock will be sold at $15 or more.

The primary advantage of a limit order is that it guarantees that the trade will be made at a particular price; however, it's possible that your order will not be executed at all if the limit price is not reached.

Traders choose order submission strategies on the basis of how quickly they want to trade, the prices they are willing to accept, and the consequences of failing to trade.

Validity Instructions

These indicate when the order may be filled.

A **day order** (the most common) is a market or limit order that is in force from the time the order is submitted to the end of the day's trading session.

A **good-till-canceled order** requires a specific canceling order. It can persist indefinitely (although brokers may set some limits, for example, 90 days).

An **immediate-or-cancel order** (IOC) will be immediately executed or canceled by the exchange. Unlike a **fill-or-kill** order, IOC orders allow for partial fills.

An order may be specified **on the close** or **on the open**, then it is entered in an auction but has no effect otherwise.

Different types of orders allow you to be more specific about how you'd like your broker to fulfill your trades. When you place a stop or limit order, you are telling your broker that you don't want the market price (the current price at which a stock is trading), but that you want the stock price to move in a certain direction before your order is executed.

With a **stop order**, your trade will be executed only when the security you want to buy or sell reaches a particular price (the **stop price**). Once the stock has reached this price, a stop order essentially becomes a market order and is filled. For instance, if you own stock ABC, which currently trades at $20, and you place a stop order to sell it at $15, your order will only be filled once stock ABC drops below $15. Also known as a "stop-loss order," this allows you to limit your losses. However, this type of order can also be used to guarantee profits. For example, assume that you bought stock XYZ at $10 per share and now the stock is trading at $20 per share. Placing a stop order at $15 will guarantee profits of approximately $5 per share, depending on how quickly the market order can be filled.

Stop orders are particularly advantageous to investors who are unable to monitor their stocks for a period of time, and brokerages may even set these stop orders for no charge.

One disadvantage of the stop order is that the order is not guaranteed to be filled at the preferred price the investor states. Once the stop order has been triggered, it turns into a market order, which is filled at the best possible price. This price may be lower than the price specified by the stop order. Moreover, investors must be conscientious about where they set a stop order. It may be unfavorable if it is activated by a short-term fluctuation in a stock's price. For example, if stock ABC is relatively volatile and fluctuates by 15% on a weekly basis, a stop loss set at 10% below the current price may result in the order being triggered at an inopportune or premature time.

Clearing Instructions

These indicate how to arrange the final settlement of the trade. For example, which entity is responsible for clearing and settling the trade? The broker, the custodian, or the clearing house?

6. Primary Security Markets

```
   The <b>primary markets</b> are those in which new issues of bonds, preferred stock, or
```

- New issue.
- Key factor: issuer receives the proceeds from the sale.

There are two important rules in the primary capital markets:

- **Rule 415** allows large firms to register security issues and sell them piecemeal over the following two years. Such issues are called **shelf-registration**. This rule allows a single registration document to be filed that permits the issuance of multiple securities.
- **Rule 144A** allows corporations (including non-U.S. firms) to place securities privately with large, sophisticated investors. The issuer of a **private placement** reduces issuing costs because it does not have

to complete the extensive registration documents. However, investors will require a higher return since no secondary market exists and thus the liquidity risk is high.

New stock issues are divided into two groups:

- **Initial public offerings** (IPOs). These are new shares that a firm offers to the public *for the first time.* They are typically underwritten by investment bankers through negotiated arrangements (the most common form), competitive bids, and best-effort arrangements (investment bankers act as brokers, not taking the price risk).
- **Seasoned equity issues**. These are new shares issued by firms that already have stocks outstanding.

A **rights issue** is an option that a company can opt for to raise capital under a secondary market offering or by using a seasoned equity offering of shares to raise money. It is a special form of shelf offering or shelf registration. With the issued rights, existing shareholders have the privilege of buying a specified number of new shares from the firm at a specified price within a specified time.

Government bond issues are sold at Federal Reserve auctions.

7. Secondary Security Market and Contract Market Structures

```
   The <b>secondary markets</b> permit trading in outstanding issues; that is, stocks or
```

- The existing owner sells to another party.
- The issuing firm does not receive proceeds and is not directly involved.

Secondary markets support primary markets.

- The secondary market provides *liquidity* to the individuals who acquired these securities, and the primary market benefits greatly from the liquidity provided by the secondary market because investors would hesitate to acquire securities in the primary market if they thought they could not subsequently sell them in the secondary market.

- Secondary markets are also important to issuers because the *prevailing market price* of the securities is determined by transactions in the secondary market. New issues of outstanding securities (seasoned securities) in the primary market are based on the prices in the secondary market. Forthcoming IPOs in the primary market are priced based on the prices of comparable stocks in the public secondary market.

Trading Sessions

Securities exchanges differ in when stocks are traded.

In a **call market**, trading for individual stocks takes place at *specified times*. The intent is to gather all the bids and asks for the stock and attempt to arrive at a *single price* where the quantity demanded is as close as possible to the quantity supplied.

- This trading arrangement is generally used during the early stages of development of an exchange when there are few stocks listed or a small number of active investors/traders.
- Call markets also are used at the opening for stocks on the NYSE if there is an overnight buildup of buy and sell orders, in which case the opening price can differ from the prior day's closing price.
- The concept is also used if trading is suspended during the day because of some significant new information. The mechanism is considered to contribute to a more orderly market and less volatility in such instances because it attempts to avoid major up-and-down price swings.

In a **continuous market**, trades occur *any time* the market is open. Stocks are priced either by auction or by dealers. In an auction market, there are sufficient willing buyers and sellers to keep the market continuous. In a dealer market, enough dealers are willing to buy or sell the stock.

Please note that dealers may exist in some auction markets. These dealers provide temporary liquidity and ensure market continuity if the market does not have enough activity.

Although many exchanges are considered continuous, they (e.g., NYSE) also employ a call-market mechanism on specific occasions.

8. Well-Functioning Financial Systems

```
   Well-functioning financial systems have the following characteristics:<p> </p><ul clas
```

- **Complete markets.** The instruments needed to solve investment and risk management problems are available to trade.

- **Liquidity.** As asset can be bought and sold quickly (that is, it has **marketability**, which means an asset's likelihood of being sold quickly) at a price close to the prices for previous transactions (**price continuity**), assuming no new information has been received. In turn, price continuity requires **depth**, which means that numerous potential buyers and sellers must be willing to trade at prices above and below the current market price.

- **Operational efficiency.** Low transaction costs (as a percentage of the value of the trade) include the cost of reaching the market, the actual brokerage costs, and the cost of transferring the asset. This attribute is often referred to as **internal efficiency**.

- **Informational (or external) efficiency.** Timely and accurate information is available on the price and volume of past transactions and the prevailing bid-price and ask-price. Prices rapidly adjust to new information; thus the prevailing price is fair because it reflects all available information regarding the asset. Prices will be most informative in liquid markets because information-motivated traders will not invest in information and research if establishing positions based on their analysis is too costly.

A well-functioning financial system promotes wealth by ensuring that capital allocation decisions are well-made. It also promotes wealth by allowing people to share risks associated with valuable products that would otherwise not be undertaken.

9. Market Regulation

```
   Regulators generally seek to promote fair and orderly markets in which traders can tra
```

The objectives of market regulation are to:

- *control fraud.* Customers may not know how to protect themselves, since the financial markets are quite complex.
- *control agency problems.* Financial agents often have different goals from their customers. How to effectively measure the services they provide?
- *promote fairness.* For example, insider trading is prohibited in most markets as it offends basic notions of fairness.
- *set mutually beneficial standards.* Common financial standards allow investors to compare companies easily.
- *prevent undercapitalized financial firms from exploiting their investors by making excessive risky investments.* Regulators generally require that financial firms maintain minimum levels of capital to reduce the probability that these firms will fail and hurt their customers.
- *ensure that long-term liabilities are funded.* Insurance companies and pension funds need to maintain adequate reserves to ensure they can pay their liabilities when due.

Security Market Indexes

1. Index Definition and Calculations of Value and Returns

```
   A <b>security market index</b> is a means to measure the value of a set of securities
```

There are usually two versions of the same index:

- The price return takes into account only the capital gain on an investment. A **price return index** reflects only the prices of the constituent securities. The income generated by the assets in the portfolio, in the form of interest and dividends, is ignored.

  The value of a price return index is calculated as:

  n_i: the number of units of security i in the index portfolio.
  P_i: the unit price of security i.
  D: the value of the divisor.

- The total return takes into account not only the capital appreciation on the portfolio, but also the income received. A **total return index** reflects the prices and the reinvestment of all income received since the inception of the index.

Single Period Returns

The single-period price return of an index is the weighted average of the price returns of the individual securities:

or

Since the total return of an index includes price appreciation and income, we need to add the weighted average of income to the above formula to calculate the single-period total return:

or

Multiple-Period Returns

The single-period returns should be linked geometrically.

Similarly, to calculate the total return over multiple periods:

2. Index Construction and Management

```
    The steps to construct and manage a security market index:<p> </p><ul class="notes">
```

- The first decision is to identify the target market. Which market should the index represent?

- The second decision is to select specific securities to include in the index. How many securities to include? Which ones? The following factors are important:

  - Size: the larger, the better - but eventually the costs of taking a larger sample will outweigh the benefits.
  - The breadth of the sample: the sample must represent the total population.
  - The source of the sample: samples must be taken from each different segment of the population.

- The third decision is to determine the weight to be allocated to each security in the index (discussed below).

- When should the index be rebalanced?

- When should the security selection and weighting decisions be re-examined?

Price Weighting

This is an *arithmetic average* of current prices. Index movements are influenced by the differential prices of the components.

The weight of each security is calculated using this formula:

The index itself is computed by:

- Adding up the market price of each stock in the index, then
- Dividing this total price by the number of stocks in the index: price-weighted series = sum of stock prices / number of stocks in the series.

Example

Shares of firm A sell for \$100 and shares of firm B sell for \$25. The initial price index is (100 + 25) / 2 = 62.5. The divisor is therefore 2.

- Normal situation. Suppose that A increases by 10% to \$110 and B increases by 20% to \$30; the price index would be (110 + 30) /2 = 70. The rate of return would be: (70 - 62.5) / 62.5 = 12%.
- Stock split. If A were to split two for one, and its share price were therefore to fall to \$50, we would not want the average to fall since that would incorrectly indicate a fall in the general level of market prices. Following a split, the divisor must be reduced to a value that leaves the average unaffected by the split. The new divisor is: (50 + 25) / 62.5 = 1.2, which will make the initial value of the average unaffected.

Price-weighting is simple, but a price-weighted index has a downward bias.

- High-priced stocks have a greater impact on the index than low-priced stocks, as the scheme assumes that an investor purchases an equal number of shares for each stock in the index.
- Large successful firms consistently lose weight within the index since high-growth companies tend to split their stocks more often. Over time, low-growth small firms with high prices will dominate the index.

Both the Dow Jones Industrial Average (DJIA) and the Nikkei-Dow Jones Average use this method to weight an index.

Equal Weighting

All stocks carry equal weight regardless of their price or market value. A \$1 stock is as important as a \$10 stock, and a firm with a \$200 million market value is the same as one with a \$200 billion value.

The actual movements in the index are typically based on the arithmetic average of the percent changes in price or value for the stocks in the index: each percent change has equal weight. Such an index can be used by individuals who randomly select stock for their portfolios and invest the same dollar amount in each stock.

The weight of each security is calculated using this formula:

It assumes that equal dollar amounts are invested in each stock in the index at the beginning of the period. It is typically generated by taking the arithmetic or geometric mean of the percentage changes in the value of the stocks in the index.

The primary advantage of equal weighting is its simplicity. However, since the prices of securities keep changing, the index needs to be rebalanced frequently to maintain equal weights.

Market-Capitalization Weighting

This measurement is generated by deriving the initial total market value of all stocks used in the series. The importance of individual stocks in the sample depends on the market value of the stocks. There is an automatic adjustment for stock splits and other capital changes in this series.

The weight of each security is calculated using this formula:

Q_i is the number of shares outstanding of security i.

A market-value-weighted series is generated by:

- Adding up the total market value of all stocks in the index: market value = number of shares outstanding x current market value.
- Dividing this total by the total market value for the base period.
- Multiplying this ratio by the beginning index value: new market value = (current market value / base value) x beginning index value.

Example

Shares of firm A sell for $100 with 1 million shares and shares of firm B sell for $25 with 20 million shares. Their market value is therefore $100 million and $500 million, respectively. If A increases by 10% to $110 and B increases by 20% to $30, their market value will be $110 million and $600 million, respectively. The rate of return would be: (710 - 600) / 600 = 18.3%.

As you can see, firms with large market value have greater impact on the index than firms with small market value. Thus, over time the large-market-value stocks will dominate changes in a market-value-weighted series.

A free-float adjustment factor is introduced in the **float-adjusted market-capitalization weighting**. It represents the proportion of shares that are free-floated as a percentage of issued shares. The index therefore does not include restricted stocks.

Fundamental Weighting

Fundamentally based indices are indices in which stocks are weighted by one of many economic fundamental factors, especially accounting figures, which are commonly used when performing corporate valuation, or by a composite of several fundamental factors.

Index Management: Rebalancing and Reconstitution

Rebalancing is adjusting the weights of the constituent securities in an index. This is done to maintain the weight of each security consistent with the index's weighting method. There is no need to rebalance price-weighted indices.

Companies may disappear through mergers or acquisitions, or they can become insolvent. A company may no longer satisfy the requirements for index inclusion. Changing the composition of an index is called **reconstitution**. Reconstitution is undertaken to ensure the index represents the desired target market.

Rebalancing and reconstitution create turnover in an index. Reconstitution can dramatically affect the prices of current and prospective constituents.

3. Uses of Market Indices

```
    Security market indices are used:<p> </p><ul class="notes">
```

- *For predicting future market movements by technicians.* Technicians believe past price changes can be used to predict future price movements. For example, to project future stock price movements, technicians would plot and analyze price and volume changes for a stock market series like the DJIA.

- *To measure market rates of return in economic studies.*

- *As a proxy for the market portfolio of risky assets.* When calculating the systematic risk of an asset, it is necessary to relate its returns to the returns of an aggregate market index that is used as a proxy for the market portfolio of risky assets.

- *As benchmarks to evaluate the performance of professional money managers.* A basic assumption when evaluating portfolio performance is that any investor should be able to experience a rate of return comparable to the market return by randomly selecting a large number of stocks from the total market. Therefore, a stock-market index can be used as a benchmark to judge the performance of professional money managers.

- *To create and monitor an index fund or an exchange-traded fund (ETF).* An index fund is created to track the performance of the specific market series (index) over time.

4. Different Types of Security Market Indices

```
    <b>Equity Indices</b><p> </p>
```

There are different types of equity indices.

A **broad market index** represents an entire given equity market. Examples include the Russell 3000, the Wilshire 5000 Total Market Index, etc.

Local indices of individual countries lack consistency in sample selection, weighting, or computational procedures. Global equity indexes are created to solve this comparability problem. A **multi-market index** represents multiple security markets. For example, the Dow Jones World Stock Index includes 2,200 companies in 33 countries.

A **sector index** measures the performance of a narrow market segment, such as the biotechnology sector. It can be used to determine if a portfolio manager is good at sector allocation or not. It can also be used to track the performance of sector-specific funds.

Style strategies focus on the underlying characteristics common to certain investments. Growth is a different style than value, and large capitalization investing is a different style than small stock investing. A growth strategy may focus on high price-to-earnings stocks and a value strategy on low price-to-earnings stocks. Style indices are created to represent such securities.

Fixed Income Indices

The creation and computation of bond-market indices is more difficult than that for a stock market series.

- The universe of bonds is much broader than that of stocks.
- The universe of bonds is changing constantly because of new issues, bond maturities, calls, and bond sinking funds.
- The volatility of prices for individual bonds and bond portfolios changes because bond price volatility is affected by duration, which is changing constantly.
- Pricing individual bonds is difficult compared to the current and continuous transactions prices available for most stocks used in stock indexes.

All bond indices indicate total rates of return for the portfolio of bonds, including price change, accrued interest, and coupon income reinvested. They are relatively new and not widely published. Most indices are market-value weighted.

Bond indices can be categorized based on their broad characteristics, such as type of issuer, currency, maturity, and credit rating. For example, there are different indices for government bonds, high-yield bonds, corporate bonds, and mortgage-backed securities.

Commodity Indices

There are five major commodity sectors: energy, grains, metals, food and fiber, and livestock.

A commodity price index is a fixed-weight index of selected commodity prices, which may be based on spot or futures prices. It is designed to be representative of the broad commodity asset class or a specific subset of commodities, such as energy or metals.

- Different commodity indices have different weighting methods, which result in different risk and return profiles.
- A commodity index may track commodities directly, or indirectly by tracking futures contracts for certain commodities. For example, commodity indices may track energy products or currencies, or may tracks futures contracts in either of those. For a commodity index that consists of futures contracts on the commodities, the index returns are affected by factors such as the prices of the underlying commodity, the risk-free interest rate, and the roll yield.

Real Estate Investment Trust Indices

Types of real estate indices include appraisal indices, repeat sales indices, and REIT indices which track the performance of publicly traded REITs.

Hedge Funds Indices

There are many indices that track the hedge fund industry. Since hedge funds are illiquid, heterogeneous, and ephemeral, it is really hard to construct a satisfactory index.

Funds' participation in an index is voluntary, leading to self-selection bias because those funds that choose to report may not be typical of funds as a whole.

The short lifetimes of many hedge funds means that there are many new entrants and many departures each year, which raises the problem of survivorship bias. If we examine only funds that have survived to the present, we will overestimate past returns because many of the worst-performing funds have not survived, and the observed association between fund youth and fund performance suggests that this bias may be substantial.

When a fund is added to a database for the first time, all or part of its historical data is recorded ex-post in the database. It is likely that funds only publish their results when they are favorable, so the average performances displayed by the funds during their incubation period are inflated. This is known as "instant history bias" or "backfill bias."

Market Efficiency

1. The Concept of Market Efficiency

```
An <b>efficient capital market</b> is one in which security prices adjust rapidly to t
```

Why should capital markets be efficient? Competition is the source of efficiency, and price changes should be independent and random.

- A large number of competing profit-maximizing participants analyze and value securities, each independently of the others.
- New information regarding securities comes to the market in a random fashion, and the timing of an announcement is generally independent of others.
- Competing investors attempt to adjust security prices rapidly to reflect the effect of new information. The price adjustment is unbiased: sometimes the market will over-adjust and other times it will under-adjust; you cannot predict its behavior.

In an efficient market, the expected returns implicit in the current price of the security should reflect its risk. Investors buying the security should receive a return that is consistent with the perceived risk of the security.

In an efficient capital market the majority of portfolio managers cannot beat a buy-and-hold policy on a risk-adjusted basis. An index fund which simply attempts to match the market at the lowest cost is preferable to an actively managed portfolio.

Market Value versus Intrinsic Value

- **Intrinsic value** is the true, actual value of an asset. It is what the asset is really worth.
- **Market value** is the price of an asset. It is what buyers are willing to pay for the asset.

In an efficient market, the two values should be very close or the same. In other words, in an efficient market at any point in time the actual price of a security will be a good estimate of its intrinsic value. Though market value and intrinsic value may differ over time, the discrepancy will get corrected as new information arrives.

In an inefficient market, the two values may differ significantly.

Factors Affecting a Market's Efficiency

Some factors contribute to and some impede the degree of efficiency in a financial market.

- *The number of market participants.* The more investors and analysts that follow a financial market, the more efficient it becomes.
- *Information availability and financial disclosure.* All investors should have access to the necessary information to value securities. This should promote market efficiency.
- *Limits to trading.* Some researchers argue that restrictions on short selling impede market efficiency.

Transaction costs and information-acquisition costs should also be considered when evaluating market efficiency.

2. Forms of Market Efficiency

```
   There are three versions of the Efficient Market Hypothesis (EMH); they differ in thei
```

- The **weak-form hypothesis** asserts that stock prices already reflect all the information that can be derived by examining market trading data, such as the history of past prices, trading volume, or short interest. This implies that trend analysis is fruitless: if such data ever conveyed reliable signals about future performance, all investors would have become familiar with such signals already.

- The **semi-strong-form hypothesis** states that all publicly available information regarding the prospects of a firm must be reflected already in the stock price. Such information includes (in addition to past prices) fundamental data on the firm's product line, quality of management, balance sheet composition, patents held, earning forecasts, and accounting practices. Obviously this version encompasses the weak-form EMH. This hypothesis implies that an investor cannot achieve risk-adjusted excess returns using important *public* information.

Event studies examine how fast stock prices adjust to specific significant economic events. The results for most of these studies have supported the semi-strong-form EMH. About the only mixed results come from exchange-listing studies.

- The **strong-form hypothesis** states that stock prices reflect all information (from *public and private* sources) relevant to the firm, including information available only to company insiders. This version of EMH encompasses both the weak-form and the semi-strong-form EMH. It is quite extreme. It implies that no investor has monopolistic access to information that influences prices. Thus, no investor can consistently derive risk-adjusted excess returns. In fact, the strong-form EMH assumes **perfect markets**, in which all information is cost-free and available to everyone at the same time. In contrast, in an efficient market prices adjust rapidly to new *public* information.

Implications of EMHs

Technical Analysis

The assumptions of technical analysis directly oppose the notion of efficient markets.

- The process of disseminating new information takes time.
- Stock prices move to new equilibriums in a *gradual* manner.
- Hence, stock prices move in trends that persist.

Therefore, technical analysts believe that good traders can detect the significant stock price changes before others do. However, as confirmed by most studies, the capital market is weak-form efficient as prices fully reflect all market information as soon as the information becomes public. Though prices may not be adjusted perfectly in an efficient market, it is unpredictable whether the market will over-adjust or under-adjust at any time. Therefore, technical analysts should not generate abnormal returns and no technical trading system should have any value.

Fundamental Analysis

Fundamental analysts believe that:

- At any time, there is a basic intrinsic value for the aggregate stock market, various industries, or individual securities;
- These values depend on underlying economic factors such as cash flows and risk variables;
- Though market price and the intrinsic value may differ over time, the discrepancy will get corrected as new information arrives.

Therefore, by accurately estimating the intrinsic value, a fundamental analyst can achieve abnormal returns by making superior market timing decisions or acquiring undervalued securities.

Fundamental analysis involves aggregate market analysis, industry analysis, company analysis, and portfolio management. However, using historical data to estimate the relevant variables is as much an art and a product of hard work as it is a science. A fundamental analyst must do a superior job to predict earnings surprises to beat the market.

- *Market analysis.* Analysis relying solely on historical data will not yield superior, risk-adjusted returns as the EMH asserts that the market adjusts rapidly to public information. The analyst must be good at estimating the relevant variables that cause long-run trends of market movements.

- *Industry and company analysis.* The EMH implies that to achieve abnormal returns, an analyst must correctly estimate future values for variables that influence rates of return and predict future earnings surprises. The estimates must differ from the consensus. There will be no superior return if the analyst predicts the consensus and the consensus is correct. Therefore, the analyst should pay more attention to areas where the market is inefficient, such as stocks that are neglected by other analysts, stocks with high book value/market value ratios, and stocks with small market capitalization.

Portfolio Management

Since the capital markets are primarily efficient, the majority of portfolio managers cannot beat a buy-and-hold policy on a risk-adjusted basis. However, on many occasions the market fails to adjust prices rapidly in response to public information, and superior investment performance is likely to be achieved through active security valuation and portfolio management. This achievement relies on superior analysts who can time major market trends or identify undervalued securities. Hence, the decision of how one manages a portfolio (actively or passively) should depend on whether the manager has access to superior analysts.

If a portfolio manager has access to superior analysts, he or she can manage a portfolio actively, looking for undervalued securities based upon superior fundamental analysis (including predicting earnings surprises) and attempting to time the market when asset allocation is shifted between aggressive and defensive positions. The portfolio manager should ensure that the risk preferences of the client are maintained.

If a portfolio manager does not have access to superior analysts, he or she should:

- Determine and quantify his risk preferences;
- Construct the appropriate risk-level portfolio by dividing the total portfolio between lending or borrowing risk-free assets and a portfolio of risky assets;
- Diversify completely on a global basis to eliminate all unsystematic risk;
- Maintain the specific risk level by rebalancing when necessary;
- Minimize taxes and total transaction costs (reduce trading turnover and trade relatively liquid stocks to minimize liquidity costs).

The Rationale of Index Funds

If companies don't have superior analysts who can beat the market, then the analysts should simply attempt to match the market at the lowest cost. To achieve a market rate of return, diversification in numerous amounts of stocks is required, which may not be an option for a smaller investor. *Index funds* (also referred to as *market funds*) are security portfolios designed to duplicate the composition and therefore the performance of a selected market index series.

3. Market Pricing Anomalies

```
   Are the hypotheses supported by the data? Are there market patterns that lead to abnor
```

A **market anomaly** is a security price distortion in the market that seems to contradict the efficient market hypothesis. There are different categories of market anomalies.

Time-Series Anomalies

Calendar anomalies raise the question of whether some regularities exist in the rates of return during the calendar year that would allow investors to predict returns on stocks.

The **January anomaly**, also called **small-firm-in-January effect**, says that many people sell stocks that have declined in price during the previous months to realize their capital losses before the end of the tax year. Such investors do not put the proceeds from these sales back into the stock market until after the turn of the year. At that point the rush of demand for stock places an upward pressure on prices that results in the January effect.

The effect is said to show up most dramatically for the smallest firms because the small-firm group includes stocks with the greatest variability of prices during the year (and the group therefore includes a relatively large number of firms that have declined sufficiently to induce tax-loss selling).

Another possible reason for the January effect on stock markets is strategic selling by institutional investors at the end of their reporting periods. Portfolio managers may be reluctant to report holdings of stocks in their annual reports that have performed poorly in the previous period. Therefore, the managers sell these stocks at the end of their accounting periods (usually the end of December). This so-called "window-dressing" was suggested as a source of the January effect by Haugen and Lakonishok (1988).

Despite numerous studies, the January anomaly poses as many questions as it answers.

Other calendar anomalies include the monthly effect, weekend or day-of-the-week effect, and intraday effect.

Momentum and Overreaction Anomalies. The debate surrounding investor overreaction and contrarian investing is one of the most extensive and controversial areas of research in finance. The overreaction anomaly, evidenced by long-term reversals in stock returns, was first identified by De

Bondt and Thaler (1985), who showed that stocks which have performed poorly in the past three to five years demonstrate superior performance over the next three to five years compared to stocks that have performed well in the past. The study provided evidence that abnormal excess returns could be gained by employing a strategy of buying past losers and selling short past winners, or the contrarian strategy.

Although the overreaction anomaly and market momentum do seem to exist, researchers have argued that the existence of momentum is rational, and the additional return (based on the contrarian investment strategy) would come simply at the expense of increased risk.

Cross-Sectional Anomalies

If the semi-strong EMH is true, all securities should have equal risk-adjusted returns because security prices should reflect all public information that would influence the security's risk. Using public information, is it possible to determine what stocks will enjoy above-average, risk-adjusted returns?

The **size effect** relates to the impact of size (measured by total market value) on risk-adjusted rates of return. Some researchers found that small firms outperformed large firms after considering risk and transaction costs.

Basu's study concluded that publicly available P/E ratios conveyed valuable information, and that the risk-adjusted returns for stocks in the lowest P/E ratio quintile were superior to those in the highest P/E ratio quintile. This is known as the **value effect**.

Fama and French found that both size and BV/MV ratio are significant when included together, and that their importance dominates that of other ratios. The dramatic dependence of returns on market-to-book ratio is independent of beta, suggesting either that low market-to-book-ratio firms are relatively underpriced or that the market-to-book ratio is serving as a proxy for a risk factor that affects equilibrium expected returns.

Other Anomalies

Closed-End Investment Fund Discounts. Closed-end funds usually trade at substantial discounts relative to their net asset values. There are several explanations:

- Agency costs. The existence of management fees (from 0.5% to 2%) implies that funds will sell at a discount. However, open-end funds also charge fees. Boudreaux (1973) suggested that since fund managers buy and sell securities, discounts might reflect their differential ability to perform this task. However, this explanation does not explain why funds trade, on average, at discounts.

- Taxes. When a fund realizes a capital gain it must report this, and the tax liability is borne by the existing shareholders at the time the gain is realized. So if you buy a fund today and it realizes a large capital gain tomorrow, you must pay a tax even if you have not made any money. This implies that a fund with large unrealized capital appreciation is worth less than its net asset value to both existing and potential shareholders, and should thus sell at a discount. This explanation, like the others, has some apparent merit but fails to explain all the facts.

- Liquidity. One way in which a portfolio might be misvalued is if the fund held large quantities of stocks that cannot be freely sold in the open market. Such stocks, some have argued, are valued too highly in the calculation of net asset value. However, most closed-end funds hold little or no restricted stock and yet still sell at discounts.

When closed-end funds are terminated, either through a merger, liquidation, or conversion to an open-end fund, prices converge to reported net asset value.

In summary, a number of reasons have been put forth to explain closed-end fund discounts in the context of the efficient market hypothesis and rational agents. Several of these factors do have some merits, but taken together, these factors explain only a small portion of the total variation in discounts.

Earnings surprises. Price changes tend to persist after initial announcements. Stocks with positive surprises tend to drift upward; those with negative surprises tend to drift downward. Some refer to the likelihood of positive earnings surprises to be followed by several more earnings surprises as the "cockroach" theory; when you find one, there are likely to be more in hiding.

Research shows that the post-earnings-announcement drift occurs mainly in highly illiquid stocks, which have high trading costs and market impact costs, supporting the argument that transaction costs could be the source of the drift.

Initial public offerings. Because of uncertainty about price and the risk involved in underwriting stocks of previously closely held companies, it has been hypothesized that underwriters tend to under-price these new issues. Although there is some under-pricing of IPOs (about 15%) when they are offered, the price adjustment takes place within one day after the offering. Investors who acquire the stock after the initial adjustment do not experience abnormal returns.

Predictability of returns based on prior information. Finding that stock returns are related to prior information, such as interest rates, inflation rates, and dividend yields, would not result in abnormal trading returns.

Summary

Most empirical evidence supports the semi-strong form EMH. The test results of the strong-form EMH are mixed.

4. Behavioral Finance

```
   Some investors behave highly irrationally and make predictable errors. <b>Behavior fir
```

Loss Aversion

This is a theory that people value gains and losses differently and, asa result, will base decisions on perceived losses rather than perceived gains. Thus, if people were given two equal choices, one expressed in terms of possible losses and the other in possible gains, they would choose the former.

Overconfidence

Most people consider themselves to be better than average in most things they do. For example, 80% of drivers contend that they are better than "average" drivers. Is that really possible? Studies show that money managers, advisors, and investors are consistently overconfident in their ability to outperform the market. Most fail to do so, however.

Other behavior theories include representativeness, gambler's fallacy, mental accounting, etc.

Information Cascades

Information cascading is defined as a situation in which an individual imitates the trades of other market participants and completely disregards his or her own private information. A related concept is **herding**, which is clustered trading that may or may not be based on information. Some researchers argue that institutional investors trade together because they receive correlated private information or infer private information from previous trades, and institutional herding helps prices more quickly reflect market information and improve market efficiency. The result is that trading does not incorporate information and prices can move away from fundamentals.

Some researchers argue that information cascades help promote market efficiency.

Equity Investments (2)

Overview of Equity Securities

1. Equity Securities in Global Financial Markets

```
   Equity securities play a fundamental role in investment analysis and portfolio managem
```

Global equity securities have offered an average annualized real return of 5% based on historical data, while the average annual real return is about 1% or 2% for government bills and bonds. However, equity securities are more volatile than government bills and bonds. They represent a key asset class for global investors because of their unique return and risk characteristics.

2. Types and Characteristics of Equity Securities

```
<b>Common Shares</b><p> </p>
```

Common shares represent ownership shares in a corporation.

The two most important characteristics of common shares are:

- **Residual claim** means the shareholders are the last in line of all those who have a claim on the assets or income of the corporation.
- **Limited liability** means that the greatest amount shareholders can lose in the event of the failure of the corporation is the original investment.

Each share of voting common stock entitles its owner to one vote on any matters of corporate governance that are put to a vote at the corporation's annual meeting. Shareholders who do not attend the annual meeting can **vote by proxy**, empowering another party to vote in their name.

Statutory voting, also known as **straight voting**, is a procedure of voting for a company's directors in which each shareholder is entitled to one vote per share. For example, if you owned 100 shares, you would have 100 votes.

Cumulative voting is another procedure of voting for a company's directors. Each shareholder is entitled to one vote per share times the number of directors to be elected. For example, if you owned 100 shares and there were three directors to be elected, you would have 300 votes. This is advantageous for individual investors because they can apply all of their votes toward one person.

Common shares can be callable or putable. **Callable common shares** give the issuer the right to buy back the shares from shareholders at a pre-determined price. **Putable common shares** give shareholders the right to sell the shares back to the issuer at a pre-determined price.

Preference Shares

A **preferred share**, also called a **preference share**, has features similar to both equities and bonds.

- Like a bond, it promises to pay to its holder fixed dividends each year. In this sense it is similar to an infinite-maturity bond, that is, a perpetuity. It also resembles a bond in that it does not convey voting power regarding the management of the firm.
- A preferred share is an equity investment in the sense that failure to pay the dividend does not precipitate corporate bankruptcy. It has priority over a common share in the payment of dividends and upon liquidation.

Preferred dividends can be **cumulative**; that is, unpaid dividends cumulate and must be paid in full before any dividends may be paid to common shareholders. All passed dividends on a cumulative stock are **dividends in arrears**. A stock that doesn't have this feature is known as a **noncumulative** or straight preferred stock and any dividends passed are lost forever if not declared. The implication is that the dividend payments are at the company's discretion and are thus similar to payments made to common shareholders.

Participating preferred shares offer the holders the opportunity to receive extra dividends if the company achieves some predetermined financial goals. The investors who purchased these shares receive their regular dividends regardless of how well or how poorly the company performs, assuming the company does well enough to make the annual dividend payments. If the company achieves predetermined sales, earnings, or profitability goals, the investors receive additional dividends. Most preferred shares are **non-participating**.

Convertible preferred shares give the assurance of a fixed rate of return plus the opportunity for capital appreciation. The fixed-income component offers a steady income stream and some protection of capital. The option to convert these preferred shares into common shares gives the investor the opportunity to gain from a rise in share price.

3. Private versus Public Equity Securities

```
   Private securities are not publicly traded. They don't have market-determined quoted p
```

The most common investment strategies are:

- **Venture capital** is financing for privately held companies, typically in the form of equity in less mature companies, for the launch, early development, or expansion of a business. A venture firm must provide returns to its investors and has a long horizon to do so. Therefore, it has to make a high multiple on its investment and must hold out for a nice acquisition or an IPO. It must build the business from scratch to be able to carry a very high enterprise value.

- A **leverage buyout** (**LBO**) is the acquisition of a company or division of a company with a substantial portion of borrowed funds. A buyout fund seeks companies that are undervalued with high predictable cash flow and operating inefficiencies. If it can improve the business, it can sell the company or its parts, or it can pay itself a nice dividend or pay down some company debt to deleverage.

- A **private investment** in public equity, often called a **PIPE deal**, involves the selling of publicly traded common shares to private investors. Generally, companies are forced to pursue PIPEs when capital markets are unwilling to provide financing and traditional equity market alternatives do not exist for that particular issuer.

4. Investing in Non-Domestic Equity Securities

```
   There are a variety of methods for investing in non-domestic equity securities.<p> </p
```

Direct Investing

Investors can buy and sell securities directly in foreign markets. However, they have to worry about currency conversions, unfamiliar market practices, and differences in accounting practices.

Depository Receipts

Depository Receipts (DRs) are domestically traded securities representing claims of shares of foreign stocks. Those shares are held in deposit in a local bank, which in turn issues DRs in the name of the foreign company. Investors buy and sell DRs in local currency and receive all dividends in local currency.

An **unsponsored DR** is issued by a broker/dealer or depository bank without the involvement of the company whose stock underlies the DR.

A **sponsored DR** is issued with the cooperation of the company whose stock underlies the DR. These shares carry all the rights of the common shares, such as voting rights.

A **global depository receipt** (GDR) is a DR issued outside the company's home country and outside the U.S. A GDR is very similar to an ADR. It is typically used to invest in companies from developing or emerging markets.

An **American depositary receipt** (ADR) is a U.S. dollar-denominated DR that trades on a U.S. exchange. Sponsored ADRs are classified at three levels:

- A **Level I ADR** is used when the issuer is not initially seeking to raise capital in the U.S. markets or does not wish to, or can't, list its ADRs on an exchange or on Nasdaq. A Level I ADR program offers an easy and relatively inexpensive way for an issuer to gauge interest in its securities and begin building a presence in the U.S. securities markets. Level I ADRs are traded in the over-the-counter (OTC) market.

- In a **Level II ADR** program, the ADRs are listed on the U.S. securities exchange or quoted on Nasdaq, thereby offering higher visibility in the U.S. market, more active trading, and greater liquidity. Level II ADR programs must comply with the full registration and reporting requirements of the SEC's Exchange Act.
- In the most high-profile form of sponsored ADR program, **Level III**, an issuer floats a public offering of ADRs in the U.S. and lists the ADRs on one of the U.S. exchanges or Nasdaq. The benefits of a Level III program are substantial: it allows the issuer to raise capital and leads to much greater visibility in the U.S. market.
- A **SEC Rule 144A ADR** allows foreign companies to raise capital by privately placing these DRs with qualified institutional investors. SEC registration is not required.

Other methods to invest in non-domestic equity securities include **global registered shares** and **baskets of listed depository receipts**.

5. Risk and Return Characteristics of Equity Securities

```
<b>Return Characteristics</b><p> </p>
```

There are two main sources of equity securities' total return:

- **Capital gains/losses** are the difference between the net sales price of a stock and its net cost.
- **Dividends** are the portion of the firm's earnings paid to common and preferred shareholders.

Investors who purchase non-domestic equities may incur **foreign exchange gains or losses**.

Reinvestment income of dividends is also a source of return.

Risk Characteristics

The risk of an equity security is the uncertainty of its expected total return. The measurement of the risk is typically the standard deviation of its expected total return over a number of periods.

Analysts use different methods to estimate an equity's expected return and risk.

Different types of shares have different risk characteristics. Common shares are more risky than preferred shares. Some shares (e.g., callable) are more risky than other shares (e.g., putable).

6. Equity Securities and Company Value

```
Companies issue equity securities to raise capital and increase liquidity. The book va
```

Accounting Return on Equity

ROE is net income (available to common shares) divided by the total book value of equity (common shares).

The book value can be the book value at the beginning of the period or the average book value.

Apparently management's accounting choices (e.g., FIFO versus LIFO) can have a big impact on computed ROEs.

An increasing ROE is not always good. Investors should examine the source of changes in the company's net income and shareholders' equity over time to determine why its ROE is increasing.

A company's price-to-book ratio can be used to indicate investors' expectations for the company's future cash flows generated by its positive net present investment opportunities. The ratio should be used to compare companies mainly in the same industry.

The Cost of Equity and Investor's Required Rate of Return

The cost of debt is simply the periodic coupon rate or interest rate. The cost of equity, which is usually used as a proxy for investors' minimum required rate of return, is difficult to estimate because there is no existing one. Two models can be used to estimate the cost of equity: DDM and CAPM.

Introduction to Industry and Company Analysis

1. Uses of Industry Analysis

```
    Company analysis and industry analysis are closely interrelated. Company and industry
```

Industry analysis is useful for:

- Understanding a company's business and business environment.
- Identifying active equity investment opportunities.
- Formulating an industry or sector rotation strategy.
- Portfolio performance attribution.

2. Approaches to Identifying Similar Companies

```
    There are three main approaches to classifying companies:<p> </p><ul class="notes">
```

- **Products and/or service supplied.** This is the main approach to industry classification. Companies are categorized based on the products and/or services they offer. The term "sector" is used to refer to a group of related industries.

- **Business-cycle sensitivities.** A cyclical industry is sensitive to business cycles. Its revenues are generally higher in periods of economic prosperity and lower in periods of economic downturn. The performance of a non-cyclical industry is independent of the business cycle.

Non-cyclical industries can be sorted into two categories:

- A **defensive** (or **stable**) industry demonstrates stable performance during both economic expansion and contraction.
- Companies in a growth industry achieve above-normal growth rates and profitability at any stage of the general business cycle.

However, there are limitations when using these industry descriptors. For example, some industries may include both growth companies and defensive companies.

Note two things:

- Business-cycle sensitivity is a continuous spectrum.
- A global company can experience economic expansion in one part of the world while experiencing recession in another part.

- **Statistical similarities.** Statistical cluster analysis is defined as the art of finding groups in data such that the degree of natural association is high among members within the same class (internal cohesion) and low between members of different categories (external isolation). This technique can be used to categorize companies into different industries.

3. Industry Classification Systems

```
   Commercial industry classification systems include:<p> </p><ul class="notes">
```

- The **Global Industry Classification Standard** (**GICS**) is an industry taxonomy for use by the global financial community. It is used as a basis for S&P and MSCI financial market indexes in which each company is assigned to a sub-industry and to a corresponding industry, industry group, and sector, according to the definition of its principal business activity.

- The **Russell Global Sectors** classification system uses a three-tier structure to classify global companies based on the products or services they offer.

- The **Industry Classification Benchmark** (**ICB**) categorizes individual companies into subsectors based primarily on their source of revenue (or the majority of revenue).

Various governmental agencies use a number of classification systems to facilitate the comparison of data over time and among countries that use the same system. These systems include:

- The International Standard Industrial Classification of All Economic Activities (ISIC) is used by the United Nations, its agencies and many countries in the world.
- The Statistical Classification of Economic Activities in the European Community (NACE) is the European version of ISIC.
- The Australian and New Zealand Standard Industrial Classification.
- The North American Industry Classification System.

The structures of these systems are very similar. The limitation of current classification systems is that the narrowest classification unit assigned to a company generally cannot be assumed to constitute its peer group for the purpose of detailed fundamental comparisons or valuation.

Peer Group Analysis

This is the practice of comparing a firm's results to those of similar firms.

Commercial industry classification systems often provide a starting point for constructing a peer group. Start with companies in the same industry, review the subject company and its competitors' annual reports, and confirm that each comparable company's primary business activity is similar to that of the subject company. Useful questions to ask are:

- What proportion of revenue and operating profit is derived from business activities similar to those of the subject company?
- Does a potential peer company face a demand environment similar to that of the subject company?
- Does a potential company have a finance subsidiary?

4. Principles of Strategic Analysis

```
   A business has to understand the dynamics of its industries and markets in order to cc
```

- **The threat of substitutes.** Substitutes not only limit profits in normal times, they also reduce the bonanza an industry can reap in good times. The threat of a substitute is high if it offers an attractive price-performance trade-off to the industry's product and/or the buyer's cost of switching to the substitute is low.

- **The bargaining power of customers.** How strong is the position of buyers? Can they work together in ordering large volumes? This force influences the prices that firms can charge. It can also influence cost and investment as powerful buyers demand costly services.

- **The bargaining power of suppliers.** How strong is the position of sellers? Suppliers, if powerful, can exert an influence on the producing industry, such as selling raw materials at a high price to capture some of the industry's profits. In some cases, a monopolist supplier can dictate its terms to entire industries. This force determines the cost of raw materials and other inputs.

- **The threat of new entrants.** How easy or difficult is it for new entrants to start competing? Barriers to entry are unique industry characteristics that define the industry. Barriers reduce the rate of entry of new firms, thus maintaining a level of profits for those already in the industry. From a strategic perspective, barriers can be created or exploited to enhance a firm's competitive advantage.

- **The intensity of rivalry.** Does strong competition between the existing players exist? Is one player very dominant or are all equal in strength and size?

The elements of a thorough industry analysis include the following:

Barriers to Entry

In theory, any firm should be able to enter and exit a market, and if free entry and exit exists, then profits always should be nominal. In reality, however, industries possess characteristics that protect the high profit levels of firms in the market and inhibit additional rivals from entering the market. These are **barriers to entry**. They are advantages that incumbents have relative to new entrants.

The threat of entry in an industry depends on the height of entry barriers that are present. If entry barriers are low, the threat of entry is high and industry profitability is moderated.

Generally, high barriers to entry can lead to better pricing and less competitive industry conditions. However, barriers to entry are not barriers to success, and high barriers to entry do not necessarily lead to good pricing power and attractive industry economics. Barriers to entry can also change over time.

Industry Concentration

Industry concentration is often, although not always, a sign that an industry may have pricing power and rational competition. Industry fragmentation is a much stronger signal, however, that the industry is competitive and pricing power is limited.

Certainly there are important exceptions. There are industries that are concentrated with weak pricing power and there are also industries that are fragmented with strong pricing power. The level of industry concentration is just a guideline.

Industry Capacity

Tight capacity -> more pricing power

Overcapacity -> price cutting

The analyst should think not only about current capacity conditions but also about future changes in capacity levels: how long does it take for supply and demand to reach equilibrium? Are the tight supply conditions sustainable?

In general it takes longer to shift physical capacity than to shift financial and human capital to new uses.

Market Share Stability

Stable market shares -> less competitive industries

Unstable market shares -> highly competitive industries and limited pricing power

Industry Life Cycle

The industry life cycle reflects the vitality of an industry over time. Each industry develops along a similar cyclical path that includes the following stages:

- **Embryonic**: new products, slow growth, high price, weak revenue, high-risk investments.
- **Growth**: growing sales, significant profitability, and lack of competition.

- **Shakeout**: slowing growth, intense competition, and declining profitability. Competitive strategy is very important at this stage, since above-average growth can be attained only by increasing the market share.
- **Mature**: little or no growth, industry consolidation, and relatively high barriers to entry.
- **Decline**: falling demand, sales, and negative growth. Some companies fail, others exit the industry to compete in other lines of business. Companies that have the strongest competitive advantages remain in the industry and fight for market share.

There are certainly limitations of industry life-cycle analysis. Demographics and changes in technology as well as political and regulatory environments all play a role in affecting the cash flow and risk prospects of different industries. Some stages may become longer or shorter than expected and some stages may even be skipped altogether. Another limitation is that not all companies in an industry have similar performances.

Price Competition

Price competition and thinking like a customer are important factors that are often overlooked when analyzing an industry. Whatever factors most influence customer purchasing decisions are also likely to be the focus of competitive rivalry in the industry. Broadly, industries for which price is a large factor in customer purchase decisions tend to be more competitive than industries in which customers value other attributes more highly.

5. External Influences on Industry Growth, Profitability, and Risk

```
These external influences include:<p> </p>
```

Macroeconomic Influences

GDP, interest rates, inflation, the availability of credit, etc.

Technological Influences

Established companies face the threat of technological obsolescence, while technological developments may also help established industries reinforce growth. Infant industries face the threat of a new product not being accepted by the marketplace.

Demographic Influences

Broad shifts in population distribution, age, and income can have very marked effects on different industries. For example, a greater role of sports in the lives of many Americans has increased demand for sports trauma orthopedics.

In most cases, demographic shifts are easy to identify, because they occur over a very long time period. However, it is much harder to quantify such trends and determine their influence on a particular industry.

Governmental Influences

Government regulations, laws, and tax policies can have a marked influence on many industries. They may potentially increase or decrease an industry's prospects.

In certain cases, government policies create new industries. For example, after the Firestone case, governments required the original auto manufacturers to submit all information about their cars, which created a new auto business intelligence software industry.

Trade barriers established by governments support demand for specific domestic industries by fending off foreign competition (an example would be the steel industry in the U.S.).

Social Influences

Fashion changes tend to be short-term and less predictable. For example, new products in the cosmetics or film industries may enjoy a brief spark in demand, which will dissipate shortly.

Lifestyle changes tend to be long-term and more predictable. For example, as a result of greater health consciousness, natural foods and nutritional products enjoyed a boom and hard liquor sales were depressed.

6. Company Analysis

```
   After an analyst has gained an understanding of a company's external environment, he/s
```

A firm can pursue one of the two basic types of **competitive strategies**: low cost or differentiation. To achieve abnormal profitability, a company should either incur low costs in its production process, or receive premium-to-average market price based on its products' differences preferential to customers. High profits will be possible only if the company with a cost advantage can sell its products at high-enough prices and if the company with a differentiation advantage can keep the costs of superior products sufficiently low.

A checklist for company analysis includes a through investigation of:

- Corporate profile;
- Industry characteristics;
- Demand for products/services;
- Supply of products/services;
- Pricing; and
- Financial ratio.

Equity Valuation: Concepts and Basic Tools

1. Estimated Value and Market Price

```
   Equity valuation models are used to estimate the intrinsic value of an equity security
```

There are two uncertainties in this process:

- Which valuation model should we use?
- Are the inputs to be used in the model appropriate?

Analysts often use more than one valuation model because of concerns about the applicability of any particular model and the variability in estimates that result from changes in inputs.

The model should be kept as simple as possible. The goal is to minimize the inaccuracy of the forecast.

There are three major categories of equity valuation models:

- **Present value models**. Both dividend discount models and free-cash-flow-to-equity models belong to this category.
- **Multiplier models**. These are relative valuation models.
- **Asset-based valuation models**. These are based on the book value of assets and liabilities.

2. Background for the Dividend Discount Model

```
   <p> </p>A <b>cash dividend</b> is a cash amount, usually paid on a per share basis. It
```

- **Regular dividends** are dividends distributed by companies on a regular recurring basis, usually quarterly, semi-annually or annually.
- An **extra dividend** is a non-recurring distribution of company assets, usually in the form of cash, to shareholders. Generally, special dividends are declared after exceptionally strong company earnings results as a way to distribute the profits directly to shareholders. Companies in more cyclical industries are also likely to use this form of dividend payment. Extra dividends can also occur when a company wishes to make changes to its financial structure or to spin off a subsidiary company to its shareholders.

There is a belief that there is an optimal price range for every share. This is the price for the share where the price/earnings ratio and hence the company''s value is maximized. Consider a share that has become so costly that investors cannot afford to buy the share in the required even lot of 100 shares. To correct this situation the company would split its stock.

A **stock split** divides each outstanding share into several shares. In a 2-for-1 stock split, the holder of 1 share will get additional 1 share. It increases the number of shares outstanding and is generally used after a sharp price run-up to produce a large price reduction. Firms generally split their stocks only if the price is quite high and management thinks the future is bright. Therefore, stock splits are often taken as positive signal.

A **stock dividend** is a dividend paid in additional shares of stock rather than in cash. Stock dividends are expressed in percentage. For example, on a 100% stock dividend, the holder of 1 share will get additional 1 share. Stock dividends used on a regular basis will keep the stock price more or less constrained.

Both stock splits and stock dividends are used to keep stock prices within an "optimal" trading range. They are just more pieces of paper: they both divide the pie into smaller slices without affecting the fundamental position of the current stockholders. As a result, each shareholder will own more shares, but his or her slice of the firm's "pie" remains the same and each share is worth less.

A **reverse stock split** reduces the number of shares and increases the share price proportionately. For example, if you own 10,000 shares of a company and it declares a one for ten reverse split, you will own a total of 1,000 shares after the split. A reverse stock split has no affect on the value of what shareholders own.

Companies often reverse split their stock when they believe the price of their stock is too low to attract investors to buy their stock. It's usually a bad sign if a company is forced to reverse split. Companies do it to make their stock "look" more valuable, but in reality nothing changes. A company may also do a reverse split to avoid being de-listed.

Under a **stock repurchase plan**, a firm buys back some of its outstanding stock, thereby decreasing the number of shares, which should increase both EPS and the stock price. Unlike stock dividends and stock splits, share repurchases use corporate cash. It is an alternative way of paying cash dividends.

There are different reasons for share repurchases:

- Repurchase announcements are viewed as positive signals by investors because the repurchase is often motivated by management's belief that the firm's shares are undervalued. There is no question that the company has more information about itself than does any other entity, and is therefore the ultimate insider.
- It can remove a large block of stock that is overhanging the market and keeping the price of per share down.
- If the excess cash is thought to be only temporary, management may prefer to make the distribution in the form of a share repurchase rather than to declare an increased cash dividend which cannot be maintained.
- Companies can use the residual model to set a target cash distribution level, then divide the distribution into a dividend component and a repurchase component. The company has more flexibility in adjusting the total distribution than it would if the entire distribution were in the form of cash dividends.
- Tax reason: In some countries tax rate on capital gains is lower than the tax rate on cash dividends.
- Repurchases can be used to produce large-scale changes in capital structures. For example, if a firm''s capital structure is too heavily weighted with equity, it can sell debt and use the proceeds to buy back stocks, thus increase debt ratio.

A **liquidating dividend** is a payment by a firm to shareholders from capital rather than from earnings. This isn't really a good thing. It usually occurs when a company dissolves its business or sells part of its business for cash, and distributes the proceeds to its shareholders. The distribution would be treated as a capital gain for tax purposes.

Dividend Payment Chronology

Declaration Date: The date on which a firm's directors issue a statement declaring a dividend. At the time of the declaration, the company will state the holder-of-record date and the payment date.

Ex-dividend Date: The date on which the right to the current dividend no longer accompanies a stock. This is the first date that a share trades without (i.e. "ex") the dividend.

Holder-of-Record Date: If the company lists the stockholder as an owner on this date, then the stockholder receives the dividend. On this day the company closes its stock transfer book. It is typically two business days after the ex-dividend date. Unlike the ex-dividend date, this is determined by the company.

Payment Date: The date on which a firm actually mails dividend checks. The date is determined when the dividend declaration is made.

Example:

3. Present Value Models: The Dividend Discount Model

```
   Under the DDM, the value of a common stock is the present value of all future dividenc
```

One-Year Holding Period

Assume an investor wants to buy a stock, hold it for one year, and then sell it. To determine the value of the stock using DDM, the investor must estimate the dividend to be received during the period (D_1), the expected sale price at the end of the holding period (P_1), and the required rate of return (r). Then:

Multiple-Year Holding Period

If the investor anticipates holding the stock for several years and then selling it, the valuation estimate is harder. You must forecast several future dividend payments and estimate the sale price of the stock several years in the future.

Infinite Period DDM (The Gordon Growth Model)

Assume the future dividend stream will grow at a constant rate, g, for an infinite period, that r is greater than g, and that D_1 is the dividend to be received at the end of period 1. Then:

From the formula we can see that the crucial relationship that determines the value of the stock is the spread between the required rate of return (r) and the expected growth rate of dividends (g). Anything that causes a decline in the spread will cause an increase in the computed value, whereas any increase in the spread will decrease the computed value.

The process of estimating the inputs to be used in the DDM:

- Estimate the required rate of return (r):
    - Estimate the real risk-free rate.
    - Estimate the expected rate of inflation.
    - Calculate the nominal risk-free rate: (1 + real risk-free rate) x (1 + expected rate of inflation) - 1.
    - Estimate the risk premium of the stock.
    - Calculate the required rate of return on the stock: nominal risk-free rate + risk premium.
- Estimate the dividend growth rate (g):
    - Estimate the firm's retention ratio.

  - Estimate the firm's expected return on equity (ROE).
  - Calculate the dividend growth rate: retention rate (b) x return on equity (ROE)

Multistage Dividend Discount Models

The infinite period DDM has four assumptions:

- The stock pays dividends.
- Dividends grow at a constant rate (g).
- The constant growth rate will continue *forever*.
- The required rate of return is *greater than* the growth rate; otherwise the model breaks down since the denominator is negative.

Growth companies are firms that have the opportunities and abilities to earn rates of return on investments that are consistently above their required rates of return. They may experience this high growth for some finite periods of time, and the infinite period DDM cannot be used to value these true growth firms because these high-growth conditions are temporary and therefore inconsistent with the assumptions of the DDM. The higher growth rate cannot be maintained forever, and the growth rate probably exceeds the required rate of return.

In analyzing the initial years of exceptional growth, analysts examine each year individually. Due to competition, the growth rate of a supernormal growth company is expected to decline and eventually stabilize at a constant level. When the firm's growth rate stabilizes at a rate below the required rate of return, analysts can compute the remaining value of the firm assuming constant growth using the DDM.

Note that there is no automatic relationship between growth and risk: a high-growth company is not necessarily a high-risk company.

Example

A stock paid a $10 dividend last year. Dividends are expected to grow 30% for years 1 and 2, 15% for years 3 to 5, and then 5% from year 6. The required rate of return on equity is 10%.

- The present value of the first stage of supernormal growth: $V_1 = \$10 \text{ x } (1 + 0.3)/(1 + 0.1)^1 + \$10 \text{ x } (1 + 0.3)^2 / (1 + 0.1)^2 = 25.8$
- The present value of the second stage of supernormal growth: $V_2 = \$10 (1 + 0.3)^2 \text{ x } (1 + 0.15)/(1 + 0.1)^3 + \$10 (1 + 0.3)^2 \text{ x } (1 + 0.15)^2/(1 + 0.1)^4 + \$10 (1 + 0.3)^2 \text{ x } (1 + 0.15)^3/(1 + 0.1)^5 = 45.90$
- The terminal value of constant growth at the end of the 5th year: terminal value = $\$10 \text{ x } (1 + 0.3)^2 \text{ x } (1 + 0.15)^3 \text{ x } (1 + 0.05) / (10\% - 5\%) = \$539.80$
- The present value of constant value: $V_{\text{constant growth}} = \$539.8/(1 + 0.1)^5 = 335.30$
- The total value of the stock: $25.8 + $45.9 + $335.3 = 407.00

Growth Rate of Dividends

The dividend growth rate is determined by:

- The proportion of earnings paid out in dividends (the **payout ratio**);
- The growth rate of the earnings.

Since the long-term payout ratio of a firm is pretty stable, its dividend growth rate primarily depends on the **earnings growth rate**. The earnings growth rate depends on:

- The proportion of earnings retained and reinvested (the **retention ratio**);
- The rate of return on new investments.

Assuming no external financing, the growth rate of earnings is as follows:

g = Retention Rate (b) x Return on Equity (ROE)

Note that ROE = Profit Margin (Net Income/Sales) x Total Asset Turnover (Sales/Total Assets) x Financial Leverage (Total Assets/Equity)

FCFE Valuation Model

This model is quite similar to the dividend discount model. The main difference is the definition of cash flows. **Free Cash Flow to Equity** (**FCFE**) is cash available to stockholders after payments to and inflows from bondholders. It is the cash flow from operations net of capital expenditures and debt payments (including both interest and repayment of principal).

Free cash flow always reflects the capital that can potentially be paid out to shareholders, notwithstanding the dividend policy. This model is preferable to DDM models when actual dividends differ significantly from FCFE.

However, many companies have negative free cash flow for years due to large capital demands. Since prediction of free cash flow far in the future would be imprecise, the FCF model cannot be used for growth companies.

4. Preferred Stock Valuation

```
   A preferred stock pays a fixed dividend for an infinite period. Thus, a preferred stoc
```

where r is the required rate of return on preferred stock, and the dividend is assumed to be perpetual.

The basic types of preferred stock include:

- **Cumulative**. The cumulative feature of a preferred stock means that if the company withholds any part of expected dividends, these payments are in arrears and must be settled before any other dividends. Most preferred stock carries this attribute.
- **Callable**. This feature gives the issuer the right to redeem the stock at a date and price outlined in the prospectus. Most preferred stock is callable. This feature essentially reduces the value of the preferred stock.

  A **retractable preferred share** allows an investor to redeem the share whenever the investor wants. As a result, the value of the preferred share is increased.
- **Convertible**. This is an option for the preferred stockholder to convert the shares into a fixed number of common shares at any point after a pre-determined date. While exchanges are initiated by the shareholder, there is sometimes a provision allowing the company to call for the conversion.
- **Participating**. This attribute offers the investor the opportunity to earn a dividend beyond the stated rate as outlined in the prospectus. Most preferred stock is non-participating.

5. Multiplier Models

```
   A <b>price multiple</b> is a ratio of a stock's price to a measure of value per share,
```

The **P/E ratio** compares stocks on the basis of how many dollars an investor is willing to pay for a dollar of expected earnings. It is also called the **earnings multiplier**.

The two alternative types of P/E are trailing P/E and leading P/E.

- A **trailing P/E** (also current P/E) is a price multiple comparing the stock's current market price to the company's earnings during the last four fiscal quarters. The EPS in such calculations are sometimes referred to as trailing twelve months (TTM) EPS. Trailing P/E is the price/earnings ratio published in stock listings of financial newspapers.

- A **leading P/E** (also forward P/E or prospective P/E) is a price multiple comparing the stock's current market valuation with the company's forecasted earnings for the next full fiscal year or for the next four quarters.

You can use the DDM to develop the earnings multiplier model:

- From DDM: $P = D_1/(k - g)$
- Divide both side of the above equation by E_1:

For example, a stock has an expected dividend payout ratio of 90%, a required rate of return of 15%, and expected growth rate of 10%. The P/E ratio is 0.9/(0.15 - 0.1) = 18. If an investor has projected next year's earnings to be $5, the current value of the stock is $5 x 18 = $90.

Thus the P/E ratio is determined by:

- The *expected* dividend payout ratio;
- The estimated required rate of return on the stock (k);
- The *expected* growth rate of dividends for the stock.

Since a firm's long-run target payout ratio is rather stable, the spread between k and g is the main determinant of the size of the P/E ratio.

A **price-to-book ratio** is a price multiple comparing a company's current market share price to its book value per share.

A **P/S multiple** is a price multiple that divides the price per share by the last 12 months' net sales per share.

A **price-to-cash-flow ratio** equals market price per share / cash flow per share.

The Method of Comparables

Price multiples are price-scaled by a measure of value, which provides the basis for the method of comparables. The method involves the comparison of a company's actual price multiple to some benchmark value to evaluate if an asset is relatively fairly priced, relatively undervalued, or relatively overvalued. The economic rationale is the law of one price - similar assets should sell at approximately equal prices. That is, if two companies are identical in all respects, their shares should be quoted at the same price in an efficient market.

Some of the most widely used benchmarks involve the multiple of a closely matched individual stock and the average or median value of the multiple for the stock's company or industry peer group.

This method is the most popular application of price multiples. It allows investors to determine a stock's relative valuation as compared to the benchmark. For example, many analysts point out that even after the technology market crash of 2000-2001, some technology stocks still remain overvalued on the basis of the P/E multiple compared to normal historical valuations.

By its nature, the method of comparables allows analysts to value investments only on a relative basis (i.e., only in comparison to the benchmark value of multiples). It cannot be used for absolute valuation of stocks, since the benchmark itself may well depart from its fair value. Although Sun Microsystems seems to be undervalued compared to other storage hardware companies, an analyst would need more information to recommend purchase of its stocks, since the entire storage industry may currently be overvalued and primed for decline.

6. Enterprise Value

```
   Since EBITDA are distributed between all types of investors in a company (common share
```

Enterprise value (**EV**) is total company value minus the value of cash and investments.

EV = MV of common stock + MV of preferred stock + MV of debt - cash and investments

EV/EBITDA is an indication of company value, not equity value.

Example

- Net income: 34.0
- Interest expense: 7.62
- Cash outflow for interest payments: 4.0
- Depreciation and amortization: 17.2
- Marginal tax rate: 25%
- Cash and marketable securities: 8.9
- Investments: 6.2
- Price per common share on the Paris Stock Exchange: 13.8
- Total number of shares issued: 20,000,000
- Total number of shares in treasury stock: 1,320,000

The company's financial statements show that the only interest-paying liability assumed by the company is a 5-year $200MM note maturing in 3 years' time and currently trading at 4.13%. The note is paying semi-annual coupons and all interest payments have been met so far.

The company also has preferred stock that is not trading on any exchange. The book value of the preferred stock is $45. No preferred dividends are currently in arrears.

Solution:

1. Calculation of EBITDA

EBITDA = Net Income + Interest Expense + Depreciation and Amortization + Tax Expense

Tax Expense = (Net Income / (1 - Tax Rate)) - Net Income = [34.0 / (1 - 0.25)] - 34.0 = 11.3

EBITDA = 34.0 + 7.62 + 17.2 + 11.3 = 70.12

1. Calculation of Enterprise Value (EV)

Total market value of common stock = price per share of common stock x number of shares outstanding = Price per share x (shares issued - treasury stock) = 13.8 x (20,000,000 - 1,320,000) = 257.7 MM

Since the company's preferred stock is not publicly traded, we will use its book value for calculation of EV.

Semi-annual coupon on the bond = Cash outflow for interest payments / 2 = 4 / 2 = 2

We know:

the bond's term to maturity = 3 years

Yield to maturity = 4.13%

Semi-annual coupon payments = 2MM

Face Value = 200MM

Therefore, we can calculate the bond's total current market value = $188.1.

EV = Total market value of common stock + Total market value of preferred stock + Total market value of debt - Cash balances - Investments = 257.7 + 45.0 + 188.1 - 8.9 - 6.2 = 475.7.

1. Calculation of EV/EBITDA ratio

EV/EBITDA = 475.7 / 70.12 = 6.78

Advantages:

- EBITDA is more often positive than net income.
- By adding back depreciation and amortization, EBITDA does not vary according to the depreciation method used. The EV/EBITDA ratio is often used for valuation of capital-intensive companies.
- It is more appropriate than P/E for comparing companies with different financial leverage, since EBITDA is not influenced by interest expenses.

Disadvantages:

- When capital expenditures do not equal depreciation, EBITDA is not a technically correct proxy to cash flow. This qualification to EBITDA comparisons can be meaningful for the capital-intensive businesses to which EV/EBITDA is often applied.
- EBITDA includes non-cash revenues due to the accrual accounting principle.

7. Asset-Based Valuation

```
The theory underlying the asset-based approach is that the value of a business is equa
```

Pursuant to accounting conventions, most assets are reported on the books of the subject company at their acquisition value, net of depreciation where applicable. These values must be adjusted to fair market value wherever possible.

The value of a company's intangible assets, such as goodwill, is generally impossible to determine apart from the company's overall enterprise value. For this reason, the asset-based approach is not the most probative method of determining the value of going business concerns. In these cases, the asset-based approach yields a result that is probably less than the fair market value of the business.

The result from an asset-based valuation model may be used as a "sanity check" when compared to other models of valuation.

Fixed Income

Fixed Income (1)

Fixed-Income Securities: Defining Elements

1. Basic Features of a Fixed-Income Security

```
A <b>fixed income security</b> is a financial obligation of an entity (the issuer) tha
```

Issuers of bonds include supranational organizations, sovereign governments, non-sovereign governments, quasi-government entities, and corporate issuers. The risk of the issuer failing to make full and timely payments of interest and/or repayment of principal is called **credit risk**. Credit risk is inherent in all debt investments.

The **maturity date** is the date when the bond issuer is obligated to pay the outstanding principal amount. It defines the remaining life of the bond.

- It defines the time period over which the bondholder can expect to receive interest payments and principal repayment.
- It affects the yield on a bond.

- It affects the price volatility of the bond resulting from changes in interest rates: the longer the maturity, the greater the price volatility.

The **par value** (**principal**, **face value**, **redemption value**, or **maturity value**) is the amount that the issuer agrees to repay the bondholder on the **maturity date**.

- Bonds can have any par value, though a par value of $1,000 is the most common.
- The price of a bond is typically quoted as a percentage of its par value. For example, a value of 90 means 90% of the par value.
- A bond may trade above (trading at a **premium**) or below (trading at a **discount**) its par value.

The interest rate that the issuer agrees to pay each year is called the **coupon rate** (or **nominal rate**). The coupon is the annual amount of the interest payment: par value x coupon rate.

In the U.S. most issuers pay the coupon semi-annually.

If you have a "6.5 of 12/1/2019 trading at 97," you have a bond that has a 6.5 coupon rate, matures on 12/1/2019 and is selling for 97% of its par value.

A **floating-rate security**'s coupon payments are reset periodically according to some reference rate. The typical coupon formula is:

coupon rate = reference rate + quoted margin.

- Examples of reference rates are LIBOR, U.S. Treasury yields.
- The quoted margin is the additional amount that the issuer agrees to pay above the reference rate. It is a constant value and can be positive or negative. It is often quoted in basis points.
- The coupon rate is determined at the coupon reset date but paid at the next coupon date.

A **zero-coupon bond** promises to pay a stipulated principal amount at a future maturity date, but it does not promise to make any interim interest payments. The value of a zero-coupon bond increases overtime, and approaches par value at maturity. The return on the bond is the difference between what the investor pays for the bond at the time of purchase and the principal payment at maturity. The implied interest rate is earned at maturity.

For example, if an investor purchases a zero-coupon bond for $60 with a par value of $100, the investor will earn $40 of interest over the life of the bond. The investor receives no payments until maturity of the bond when he or she will receive $100.

Bonds can be issued in any currency. If an issue has coupon payments in one currency and principal payments in another currency, it is called **dual-currency issue**. The holders of **currency option bonds** can choose the currency in which coupons and principals are paid.

2. Bond Indenture

```
An <b>indenture</b> is the contract between the issuer and the bondholder specifying t
```

Bondholders may have great difficulty in ascertaining whether the issuer has been fulfilling its obligations specified in the indenture. The indenture is thus made out to a third-party trustee as a representative of the interests of the bondholders; a trustee acts in a fiduciary capacity for bondholders.

Legal Identity of the Bond Issuer and its Legal Form

The issuer is identified in the indenture by its legal name. It is obligated to make timely payments of interest and repayment of principal. Bonds can be issued by a subsidiary of a parent legal entity. They can also be issued by a holding company. A special-purpose vehicle/entity (a separate legal entity) can issue bonds collateralized by assets transferred from its sponsor. If bankruptcy occurs, the sponsor's creditors cannot go after such assets; this is known as **bankruptcy remote**.

Source of Repayment Proceeds

The source of repayment proceeds varies, depending on the type of bond.

- Supranational bonds: repayment of previously loans, or the paid-in capital from members
- Sovereign bonds: taxing authority and money creation
- Non-sovereign government bonds: general taxing authority of the issuer, project cash flows, and special taxes
- Corporate bonds: the issuer's operating cash flows
- Securitized bonds: cash flows from the underlying financial assets

Asset or Collateral Backing

Collateral backing can increase a bond issuer's credit quality.

- Seniority ranking affect credit. In general, secured debt takes priority over unsecured debt if the issuer goes bankrupt. Within unsecured debt, senior debt ranks ahead of subordinated debt. **Debentures** can be either secured or unsecured.
- Types of collateral backing include collateral trust bonds, equipment trust certificates, mortgage-backed securities, and covered bonds.

An unsecured bond is not secured by collateral.

Covered bonds are debts issued by banks that are fully collateralized by residential or commercial mortgage loans or by loans to public sector institutions.

Credit Enhancement

Credit enhancement reduces credit risks.

Internal credit enhancement considerations include:

- Tranche structure. The senior tranches get paid first, and the subordinated tranches get paid only if there are enough funds left. The subordinated tranches absorb the credit risk, making the senior tranches less risky.
- Overcollateralization. The amount of overcollateralization can be used to absorb losses. If the liability of the structure is $100 million and the collateral's value is $105 million, then the first $5 million loss will not result in a loss to any of the tranches.
- Excess spread. Underlying assets support a higher level of payment than that promised to security holders.

External credit enhancements are financial guarantees from third parties. Examples include surety bonds, bank guarantees, and letters of credit. If the third-party defaults, the external credit enhancement will fail. A cash collateral account can mitigate this concern.

Bond Covenants

Affirmative covenants set forth certain actions that borrowers must take, such as:

- Paying interest and principal on a timely basis
- Paying taxes and other claims when due
- Keeping assets in good conditions and in working order
- Submitting periodic reports to a trustee so that the trustee can evaluate the issuer's compliance with the indenture

Negative covenants set forth certain limitations and restrictions on the borrower's activities, such as:

- Limitations on the borrower's ability to incur additional debt unless certain tests are met
- Limitations on dividend payments and stock repurchases
- Limitations on the sale of assets

3. Legal, Regulatory and Tax Considerations

An important consideration for investors is where the bonds are issued and traded; thi

The bond market can be classified into two markets: an internal market and an external market.

Internal Bond Market

The internal bond market is also called the national bond market. It is divided into two parts: the domestic bond market and the foreign bond market. The **domestic bond market** is where domestic issuers issue bonds and where these bonds are subsequently traded.

A **foreign bond** (called a **Yankee bond** in the U.S., a **Samurai bond** in Japan, and a **Bulldog bond** in the U.K.) is a bond issued in a country's national bond market by an issuer not domiciled in the country where those bonds are subsequently traded.

- Regulatory authorities in the country where the bond is issued impose rules governing the issuance of foreign bonds.
- Issuers of foreign bonds include national governments and their subdivisions, corporations, and **supranationals** (entities formed by two or more central governments through international treaties).
- They can be denominated in any currency.
- They can be publicly issued or privately placed.

External Bond Market

This market is also referred to as the **international bond market**, the **offshore bond market**, or the **Eurobond market**. The bonds in this market are:

- underwritten by an international syndicate.
- offered simultaneously to investors in a number of countries at issuance.
- issued outside the jurisdiction of any single country. Therefore, they are not registered through a regulatory agency.
- in unregistered form.

Eurobonds are subject to a lower level of listing, disclosure, and regulatory requirements than domestic or foreign bonds.

Eurobonds are classified according to the currency in which the issue is denominated. For example, if a Eurobond is denominated in U.S. dollars, it is called a Eurodollar bond. A USD bond issued by Ford and sold in Japan is thus called a Eurodollar bond, not a Euroyen bond.

A **global bond** is a debt obligation that is issued and traded in both the Eurobond market and at least one domestic market (for example, a USD bond issued by the Canadian government sold in the U.S. and Japan). Issuers of global bonds typically have high credit quality, and regularly have large fund needs. The first global bond was issued by the World Bank.

Tax Considerations

In general the income portion of a bond is taxed at the ordinary tax rate. Some countries implement a capital gains tax. Some countries even differentiate between long-term and short-term capital gains. There may be specific tax provisions for bonds issued at a discount or bought at a premium.

4. Structure of a Bond's Cash Flows

<b>Principal Repayment Structures</b><p> </p>

Bullet bond. The issuer pays the full principal amount at the maturity date.

Amortizing bond. Its payment schedule requires periodic payment of interest and repayment of principal. If the entire principal is not amortized over the life of the bond, a **balloon payment** is required at the end of the term.

A **sinking fund arrangement** allows a bond's principal outstanding amount to be repaid each year throughout the bond's life or after a specific date.

A **call provision** is the right of the issuer to retire the issue prior to the stated maturity date. When only part of an issue is called, the bond certificates to be called are selected randomly or on a pro rata basis.

Coupon Payment Structures

The coupon payments of a **floating rate security** are reset periodically (e.g., quarterly) according to a reference rate such as LIBOR.

Example

Suppose that the reference rate is the 1-month LIBOR and the quoted margin is 100 basis points: if the 1-month LIBOR on the coupon reset rate is 5%, the coupon rate is 5% + 100 basis points = 6%.

The quoted margin does not need to be a positive value. For example, it could be -90 basis points.

A floating-rate security may have upper and/or lower limits on the coupon rate. A **cap** is the maximum coupon rate of a floater. It is an attractive feature for the issuer since it limits the coupon rate. A **floor** is the minimum coupon rate, and it is an attractive feature for the investor. A **collar** is a floater with both a cap and floor.

For example, assume the reference rate is the 1-month LIBOR, the quoted margin is 100 basis points, and there is a cap of 7%. If the 1-month LIBOR at reset date was 6.5%, the coupon rate per the formula would be 7.5% (6.5% + 1%), but with the cap the coupon rate is restricted to 7%.

A typical floater's coupon rate increases when the reference rate increases and decreases when the reference rate decreases. However, an **inverse floater**'s (also called a **reverse floater**) coupon rate moves in the opposite direction from the change in the reference rate: **coupon rate = K - L x reference rate**, where K and L are constant values set forth in the prospectus for the issue. To prevent a negative coupon rate there is a floor imposed.

For example, an inverse floater's coupon rate = 12% - 2 x 3-month LIBOR. If the three-month LIBOR is 2%, then the coupon rate for the next interest payment period is: 12% - 2 x 2% = 8%.

Step-up coupon bonds have low initial and gradually increasing coupon rates; that is, their coupon rates "step up" over time.

Stepped spread floaters. The quoted margins for these coupons can step to either a higher or a lower level over the security's life. For example, a five-year floating-rate note's coupon rate may be six-month LIBOR + 1% for the first two years, and three-month LIBOR + 3% for the remaining years.

Credit-linked coupon bonds. These coupons change when the issuer's credit rating changes.

Payment-in-kind coupon bonds. These coupons allow the issuer to pay coupons with additional amounts of the bond issue rather than in cash.

The payment structures for **index-linked bonds** vary considerably among countries. An **inflation-linked bond** or **linker** links its coupon payments and/or principal repayments to a price index. For example, the Treasury Inflation Protection Securities' (TIPS) coupon is based on inflation rate: coupon rate = rate of inflation + quoted margin. The index can even be a stock market index.

5. Bonds with Contingency Provisions

```
   An <b>embedded option</b> is a provision in a bond indenture that gives the issuer anc
```

Embedded options may benefit either the issuer or the bondholder. An embedded option benefits the issuer if it gives the issuer a right or it puts an upper limit on the issuer's obligations. An embedded option benefits the bondholder if it gives the bondholder a right or it puts a lower limit on the bondholder's benefits.

Callable Bonds

A bond issue that permits the issuer to call or refund an issue prior to the stated maturity date is referred to as a **callable bond**.

- The price that the issuer must pay to retire the issue is the call price.
- Bonds can be called in whole or in part. Most of the time, an entire bond issue is called. When only part of an issue is called, the bond certificates to be called are selected randomly or on a pro rata basis. This means that each bondholder will have the same percentage of his or her holdings redeemed.
- Typically, call provisions have a deferment period; that is, the issuer may not call the bond for a number of years until a specified first call date is reached. This feature is called a **deferred call**.
- The issuer has no obligation for early retirement of bonds.

A call option becomes more valuable to the bond issuer when interest rates fall. If interest rates fall, the issuer can retire the bond paying a high coupon rate, and replace it with lower coupon bonds. However, call provisions are detrimental to bondholders, since proceeds can only be reinvested at a lower interest rate.

Callable bonds exercise styles:

- American call: any time starting on the first call date
- European call: once on the call date
- Bermuda-style call: on predetermined dates following the call protection period

Putable Bonds

A put option grants the bondholder the right to sell the issue back to the issuer at a specified price ("put price") on designated dates. The repurchase price is set at the time of issue, and is usually par value.

Bondholders have the option of putting bonds back to the issuer either once during the lifetime of the bond (a "one-time put bond"), or on a number of different dates. The special advantages of put bonds mean that putable bonds have lower yield than otherwise similar bonds.

The price behaviour of a putable bond is the opposite of that of a callable bond. The put option becomes more valuable when interest rates rise.

Convertible Bonds

A convertible bond is an issue that grants the bondholder the right to convert the bond for a specified number of shares of common stock. This feature allows the bondholder to take advantage of favorable movements in the price of the issuer's common stock without having to participate in losses.

Example

Suppose you can buy a 10%, 15-year, $100 par value bond today for $110 that can be converted into 10 shares at $10 per share. The market price of stock = $8; no dividends.

- The **conversion price** is the price per share at which a convertible bond can be converted into common stock. In the example above, the conversion price would be $10.
- The **conversion ratio** is the number of common shares each bond can be converted into (in this case, 10). It is the par value / conversion price. It is determined at the time the convertible bond is issued.
- The **conversion value**, also known as **parity value**, is the market price of stock x conversion ratio ($8 x 10 = $80).
- The conversion premium is the difference between the bond's price and its conversion value ($110 - $80 = $30). Conversion parity occurs when the conversion premium is zero (here, when the stock price is $11 ($11 x 10 = $110)).

Warrants are securities entitling the holder to buy a proportionate amount of stocks at some specified future date at a specified price. They are similar to call options.

Fixed-Income Markets: Issuance, Trading, and Funding

1. Classification of Fixed-Income Markets

Below are common criteria used to classify fixed-income markets.<p> </p>

Type of Issuer

Three major market sectors are the government and government-related sector, the corporate sector, and the structured finance sector. In most countries, the largest issuers of bonds are national and local governments as well as financial institutions.

Credit Quality

A bond can be considered investment-grade or high-yield based on the issuer's creditworthiness (as judged by credit ratings agencies).

Maturity

Money market bonds have original maturities ranging from overnight to one year. Capital market bonds have original maturities longer than one year.

Currency Denomination

The majority of bonds are denominated in either Euros or U.S. dollars.

Type of Coupon

Some bonds pay a fixed rate of interest while others pay a floating rate of interest. The coupon rate of a floater is expressed as a reference rate, such as LIBOR, plus a spread. Different reference rates are used, depending on where a bond is issued and its currency denomination.

Interbank offered rates are sets of rates that reflect the rates at which banks believe they could borrow unsecured funds from other banks in the interbank market for different currencies and maturities. These rates may be used as reference rates for floating-rate bonds, mortgages, and derivatives.

Geography

There are domestic, foreign and Eurobond markets. Investors also make a distinction between the developed and emerging bond markets. Emerging market bonds usually exhibit higher risk than developed markets bonds.

2. Primary and Secondary Bond Markets

In primary bond markets issuers first sell bonds to investors to raise capital. In sec

Primary Bond Markets

Examples

- The sale of new government bonds by the U.S. Treasury to finance a government deficit
- A $100 million bond issue by P&G to finance the construction of a new soap production plant

There are two mechanisms for issuing a bond in primary markets.

Public Offering

Any member of the public may buy the bonds in a public offering.

- **Underwritten offerings**. The function of buying the bonds from the issuer is called the underwriting. An investment bank (called the underwriter) takes the risk of buying the whole issue as **firm commitment underwriting**. It makes a profit by selling the bonds for more than what it paid for them.

  There are six phases: the determination of the funding needs, the selection of the underwriter, the structuring and announcement of the bond offering, pricing, issuance, and closing.

- **Best effort offerings**. The investment bank serves only as a broker to sell the bonds. It agrees to do its best, receives a commission for selling the bonds and incurs less risk associated with selling the bonds.

- **Shelf registrations**. An issuer files the bond registration with regulators before it makes an actual public offering of the issue. The issuer may be able to offer additional bonds to the general public without preparing a new and separate offering circular.

- **Auctions**. The issuer announces the terms of the issue and interested parties submit bids for it. Auctions often yield the most money for the issue. They allow the issuer to sell directly to the public, eliminating the underwriting fee. In major developed bond markets, newly-issued sovereign bonds are most often sold to the public via auction.

Private Placement

A private placement bond is a non-underwritten, unregistered corporate bond sold directly to a single investor or a small group of investors. Because the bonds are not registered, SEC regulations require firms to offer such bonds privately only to investors deemed sophisticated: insurance companies, pension funds, banks, and endowments.

Secondary Bond Markets

The secondary market arises after issue, when bonds are sold from one bondholder to another. Its purpose is to provide liquidity - ease or speed in trading a bond at price close to its fair market value. The buying and selling of existing bond issues is done primarily through a network of brokers and dealers who operate through organized exchanges and over-the-counter (OTC) markets. Most bonds are traded in OTC markets.

Corporate bonds typically settle on a T + 3 basis. Government and quasi-government bonds typically settle at T + 1.

3. Sovereign Bonds, Non-Sovereign Bonds, Quasi-Government Bonds, and Supranational Bonds

```
<b>Sovereign Bonds</b><p> </p>
```

Sovereign bonds are issued by a country's central government for fiscal reasons. They take different names and forms, depending on where they are issued, their maturities, and their coupon types. For example, U.S. government bonds with an original maturity shorter than one year are known as T-bills. The most recently issued U.S. Treasury bond of a particular maturity is known as a **on-the-run** issue.

Sovereign bonds are usually unsecured and are backed by the taxing authority of a national government. Credit rating agencies perform sovereign risk analysis in both local currency and foreign currency. The risk level of local and foreign currency is different. Generally, if an issuer is planning to default, it is more likely to do so with a foreign currency issue, as it has less control over foreign currency with respect to its exchange rate.

Sovereign bonds can be domestic bonds, foreign bonds, and Eurobonds. They can be fixed-rate, floating-rate or inflation-linked bonds. For example, Treasury Inflation Protection Securities (TIPS) are T-notes or T-bonds that are adjusted for inflation.

Non-Sovereign Bonds

Non-sovereign bonds are bonds issued by local governments. The sources of repayment proceeds are (the):

- General taxing authority of the issuer
- Project cash flows
- Special taxes

This type of bonds receives high credit ratings due to low default rates. They often trade at a higher yield than their sovereign counterparts.

Quasi-Government Bonds

Quasi-government bonds are issued by the government through various political subdivisions. Most of them are not secured by collateral and don't have government guarantees. Their credit ratings are very high due to extremely low historical default rates.

Supranational Bonds

Supranational bonds are bonds issued by supranational agencies such as the World Bank.

4. Corporate Debt

```
    Corporations issue different types of debt.<p> </p>
```

Bank Loans and Syndicated Loans

A bilateral loan is a loan from a single lender to a single borrower. A syndicated loan is a loan from a group of lenders to a single borrower. Most loans are floating-rate loans.

Commercial Paper

Commercial paper describes a short-term, unsecured promissory note that is used by companies as a source of short-term and bridge financing.

- Although defaults are rare, investors in this market are still exposed to credit risk.
- Many issuers roll over their paper on a regular basis. Issuers are required to secure backup lines of credit to minimize rollover risk.
- Due to higher credit risk and less liquidity, the yield from commercial paper is higher than that of short-term sovereign bonds.
- A U.S. commercial paper (USCP) is typically issued on a discount basis, while a Eurocommercial paper (ECP) is typically issued on an interest-bearing basis.

Corporate Notes and Bonds

Corporate bonds and notes take different forms, depending on the maturities, coupon payment and principal repayment structures, collateral backing and contingency provisions. These concepts are covered in the previous reading.

Medium-term notes (**MTN**) are corporate debt obligations offered to investors continually over a period of time by an agent of the issuer. The maturities vary from nine months to 30 years. Note that the term "medium-term" is not related to the term to maturity of the securities.

5. Structured Financial Instruments

```
    <p> </p>Structured financial instruments include asset-backed securities (ABS) and col
```

Capital Protected Instruments

They provide profit potential, and also protect your capital investment (fully or partially). A CPI product claiming 100% capital protection assures that investing $100 will at least allow you to recover the same amount of $100 after the investment period. If the product generates positive returns (say +15%), you yield a positive return of $115.

Say you have $10,000 to invest for one year, and you aim to protect your capital fully. Any positive return above that will be welcome.

U.S. Treasury bonds provide risk-free guaranteed returns. Assume a one-year Treasury bond offers a 5% return. If you invest $10,000 for one year, your maturity amount after one year will be 10,000 x 1.05 = $10,500.

Reverse engineering this simple calculation allows you to get the required protection for your capital. How much should you invest today to get $10,000 after one year? $10,000 / 1.05 = $9,523

You should invest $9,523 today in Treasury bonds to yield $10,000 at maturity. This secures your principal amount of $10,000.

You can use the remaining $477 to purchase a call option on some underling asset.

The T-bond + option can be prepackaged to form a capital protected instrument.

At maturity, you are guaranteed 100% of the capital invested even if the call option expires worthless. On the other hand, you have the call option which provides unlimited upside potential (if the call is in-the-money at expiry) while limiting the downside to the price (premium) paid. In our example, you would lose a maximum of $477 - the price paid for the option.

A CPI product is easy to design for the investor with a basic understanding of bonds and options. Depending on the investor's appetite for risk, capital protection can be at 100%, 90%, 80%, or less.

Yield Enhancement Instruments

Yield enhancement instruments offer the potential for a higher expected return, subject to increased risk. A **credit-linked note (CLN)** is a yield enhancement instrument that allows the issuer to transfer specific credit risks to credit investors.

Issuers of credit-linked notes use them to hedge against the risk of a specific credit event that could cause them to lose money, such as when a borrower defaults on a loan. Such instrument provide a function similar to insurance.

Investors who buy credit-linked notes generally earn a higher yield on the note in return for accepting exposure to specified credit risks.

Participation Instruments

A participation instrument allows investors to participate in the return of an underlying asset. A good example of a participating instrument is a **floating rate bond** whose coupon rate adjusts periodically according to a pre-specified formula - usually a reference rate plus a risk margin (spread).

Leveraged Instruments

They are created to magnify returns and offer the possibility of high payoffs from small investments. A good example is an **inverse floater**, which is a bond or other type of debt instrument that has a coupon rate that varies inversely with a benchmark interest rate. Investors who purchase inverse floaters will receive interest payments that are adjusted according to changes in the current interest rates. For an inverse floater, the interest rates the investor receives will adjust in the opposite direction of the prevailing rates; thus, when interest rates fall, the rate of the bond's payments increases.

Investors of inverse floaters face interest rate risk, which is the potential for investment losses due to changes in interest rates.

The general formula for the coupon rate of an inverse floater can be expressed as:

$$\text{Floating rate} = \text{Fixed rate} - (\text{Coupon leverage} \times \text{Reference rate})$$

As with all investments that employ leverage, inverse floaters introduce a significant amount of interest rate risk. When short-term interest rates fall, both the market price and the yield of the inverse floater increases, magnifying the fluctuation in the bond's price.

On the other hand, when short-term interest rates rise, the value of the bond can drop significantly, and holders of this type of instrument may end up with a security that pays little interest. Thus, interest rate risk is magnified and contains a high degree of volatility.

6. Short-Term Funding Alternatives Available to Banks

```
   Banks have different short-term funding sources. These include:<p> </p><ul class="note
```

- Retail deposits: checking accounts, savings accounts, money market accounts, etc.
- Central bank funds: funds available from the central bank, or from other banks in the central bank funds market
- Interbank funds: the market of loans and deposits between banks
- Large denomination negotiable certificates of deposit: non-negotiable or negotiable CDs, large-denomination CDs or small denomination CDs

Repurchase Agreements

A repurchase agreement is the sale of a security with a commitment by the seller to buy the same security back from the purchaser at a specified price at a designated future date. It is actually a collateralized loan. The difference between the purchase (repurchase) price and the sale price is the dollar interest cost of the loan. The implied interest rate is called the **repo rate**.

A loan for one day is called an overnight repo. A loan for more than one day is called a term repo. If a repo agreement lasts until the final maturity date, it is known as a "repo to maturity."

The repo rate is lower than the cost of bank financing. It is a function of a few factors, including the risk associated with the collateral, the term of the repo, the delivery requirement for the collateral, the supply and demand conditions of the collateral, and the interest rates of alternative financing in the money market. The more difficult it is to obtain the collateral, the lower the repo rate. Hot collateral or special collateral is collateral that is highly sought-after by dealers and can be financed at a lower repo rate than general collateral.

From a dealer's perspective, if it is lending cash, the repo is then referred to as a **reverse repurchase agreement**.

Credit risks are faced by both parties. The difference between the market value of the security used as collateral and the value of the loan is the **repo margin**. It is most likely to be lower when:

- The maturity of the repo is short.

- The quality of the collateral is high.
- The credit quality of the counterparty is high.
- The underlying collateral is in short supply or there is a high demand for it.

Introduction to Fixed-Income Valuation

1. Bond Prices and the Time Value of Money

```
The idea that the value of any financial asset equals the present value of its expecte
```

The cash flows of a bond have two components: periodic coupon payments and principal repayment at maturity, or when the bond is retired. Both the *amount* and the *timing* of the cash flows should be identified to value the bond.

For example a five-year Treasury coupon note with a par-value of $1,000 has a 10% coupon. Its cash flows are as follows:

For a given discount rate, the present value of a single cash flow to be received in the future is the amount of money that must be invested today to generate that future value. The **present value** of a cash flow will depend on when a cash flow will be received (i.e., the timing of a cash flow) and the **discount rate** (i.e., interest rate) used to calculate the present value.

i = the discount rate

t = the number of years to receive the cash flow

The sum of the present value for a security's expected cash flows is the value of the security:

N = the number of annual periods till maturity

The convention of the bond market is to quote annual interest rates that are double semi-annual rates. For most bonds coupon payments are semi-annual. Coupon payments are adjusted by dividing the annual coupon payment by two and adjusting the discount rate by dividing the annual discount rate by two:

The differences are:

A bond is priced at a **premium** above par value when the coupon rate is greater than the market discount rate. It is priced at a **discount** below par value when the coupon rate is less than the market discount rate. The amount of any premium or discount is the present value of the "excess" or "deficiency" in the coupon payments relative to the yield-to-maturity.

Yield-to-Maturity is the interest rate that will make the present value of the cash flows from a bond equal to its price. It is the *promised* rate of return on a bond if an investor buys and holds the bond to its maturity date.

Consider a 10%, two-year bond selling for $1,036.30 (selling at premium). The cash flows for this bond are (1) four payments of $50 every six months, and (2) a payment of $1,000 two years from now.

$1036.30 = $50/(1 + YTM/2)1 + $50/(1 + YTM/2)2 + $50/(1 + YTM/2)3 + $1050/(1 + YTM/2)4. Through trial and error or by using a financial calculator, YTM is found to be 8%.

2. Relationships between Bond Price and Bond Characteristics

```
<b>Price / Discount Rate Relationship</b><p> </p>
```

The value of a bond is equal to the present value of its coupon payments plus the present value of the maturity value.

The higher the discount rate, the lower a cash flow's present value and therefore since the value of a security is the present value of the cash flows, the higher the discount rate, the lower a security's value.

Example

A 1-year, semi-annual-pay bond has a $1,000 face value and a 10% coupon.

- At a discount rate of 8%, the bond value is $1,019 (premium).
- At a discount rate of 10%, the bond value is $1,000 (par).
- At a discount rate of 12%, the bond value is $982 (discount).

The degree of price change is not always the same for a particular bond. The price/yield relationship for an option-free bond is *convex*. In other words, this is not a straight-line relationship.

For a given change in yield, the price increases by more than it decreases. P_1 - P > P - P_2.

Option-free bonds exhibit **positive convexity**, which means that for a large change in interest rates, the amount of price appreciation is greater than the amount of price depreciation. This also means that the price change is greater when the level of required yield is low (and vice versa).

Coupon and Maturity Effects

All else being equal,

- Maturity effect: The longer the term to maturity, the greater the price volatility.
- Coupon effect: The lower the coupon rate, the greater the price volatility.

Constant-Yield Price Trajectory

As a bond moves closer to its maturity date, its value changes. More specifically, assuming that the discount rate does not change, a bond's value:

- decreases over time if the bond is selling at a premium
- increases over time if the bond is selling at a discount
- is unchanged if the bond is selling at par value

At the maturity date, the bond's value is equal to its par value ("pull to par value").

Pricing Bonds with Spot Rates

The valuation approach illustrated so far is the traditional approach, which uses a single interest rate to discount all of a bond's cash flows. It views all cash flows of a bond as the same, regardless of their timing. In reality, however, each individual cash flow of the bond is unique. Therefore, using a single discount rate in the bond valuation model may result in a mis-priced bond, thereby creating arbitrage opportunities.

The **arbitrage-free approach** values a bond as a package of cash flows, with each cash flow viewed as a zero-coupon bond and discounted at its own unique discount rate. These spot rates are used to discount cash flows to get the **arbitrage-free value** of a bond.

The arbitrage-free approach has three steps.

- Take each individual cash flow of a coupon as a stand-alone zero-coupon bond. Each cash flow is the

face value of the corresponding zero.
- Value each zero-coupon bond by discounting its cash flow at the corresponding spot rate.
- Add up the value of each zero to calculate the total value of the zero-coupon bond portfolio.

3. Flat Price, Accrued Interest, and the Full Price

```
   Coupon interest is paid not daily, but monthly, semi-annually or annually.  If an inve
```

Accrued interest is calculated as a proportional share of the next coupon payment using either the actual/actual or 30/360 method to count days.

The amount that the buyer pays the seller the agreed upon price for the bond plus accrued interest is called the **full price** (**dirty price**). The agreed-upon bond price without accrued interest is simply referred to as the **flat price** (**clean price**). Flat prices are quoted in order to not to misrepresent the daily increase in the full price as a result of interest accruals.

Here is how to calculate the full price:

Note that the next coupon payment is discounted for the remainder of the coupon period.

An easier formula is used to to get the present value of the bond at the last coupon payment date and find its (future) value on the settlement date.

4. Matrix Pricing

```
   Most bonds don't trade on a daily basis. Usually only the most recent large issues hav
```

Matrix pricing is the practice of interpolating among values for similar instruments arranged in a matrix format.

- It attempts to categorize bonds with similar features (e.g., type of issuer, credit rating, coupon, maturity, etc.) and apply a general yield level to the entire category of bonds. Typically a required yield over the benchmark rate is estimated.
- It then calculates the approximate price of a specific bond within a category using the derived yield level.
- It represents an educated guess and not an actual offer or trade price.

5. Yield Measures for Fixed-Rate Bonds

```
   Yield measures are used to evaluate the rate of return on bonds.<p> </p><ul class="not
```

- They are typically annualized.
- Money market rates are simple interest rates and non-money market rates are compounded.

The periodicity of an annual interest rate is the number of periods in the year.

Consider a two-year, zero-coupon bond priced now at 88 per 100 of par value.

Note:

- The effective annual rate is the same.
- The bond equivalent yield and the periodicity are inversely related.

- When comparing different bonds, it is essential to compare the yields for the same periodicity to make a statement about relative value.

To convert an annual yield from one periodicity to another:

Example

- A Eurobond pays coupons annually. It has an annual-pay YTM of 8%.
- A U.S. corporate bond pays coupons semi-annually. It has a bond equivalent YTM of 7.8%.
- Which bond is more attractive, if all else equal?

Solution 1

- Convert the U.S. corporate bond's bond equivalent yield to an annual-pay yield.
- Annual-pay yield = $[1 + 0.078/2]^2 - 1 = 7.95\% < 8\%$
- Therefore, the Eurobond is more attractive since it offers a higher annual-pay yield.

Solution 2

- Convert the Eurobond's annual-pay yield to a bond equivalent yield (BEY).
- BEY = $2 \times [(1 + 0.08)^{0.5} - 1] = 7.85\% > 7.8\%$
- Therefore, the Eurobond is more attractive since it offers a higher bond equivalent yield.

Street convention yields assume that payments are made on scheduled dates, excluding weekends and holidays. The **true yield** is calculated using a calendar including weekends and holidays. The **government equivalent yield** is based on actual/actual day count.

The **current yield** relates the annual dollar coupon interest to the market price. For example, the current yield for a 5%, two-year bond with a price of $978 is 5.11% (($1000 x 5%) / $978)). This is the simplest of all yield measures, and fails to recognize any capital gain or loss, reinvestment income or accrued interest.

The **simple yield** is similar to the current yield but includes the straight-line amortization of the discount or premium.

The standard YTM measure assumes that the bond will be held to maturity. It is not an appropriate yield measure for callable bonds, because they may be retired before maturity. For callable bonds a **yield to first call**, which assumes that the bond will be called on the first call date, is computed.

Callable bonds typically have multiple call dates, each with its own call price. The **yield to worst** is the lowest potential yield that can be received on a bond without the issuer actually defaulting. It illustrates the worst possible yield an investor may realize. The **option-adjusted-yield** is the yield-to-maturity after adding the theoretical value of the call option to the price.

6. Yield Measures for Floating-Rate Notes and Money Market Instruments

```
<b>Floating-Rate Notes</b><p> </p>
```

Interest rate volatility affects the price of a fixed-rate bonds. A floating-rate note (a floater, or an FRN) maintains a more stable price than a fixed-rate note because interest payments adjust for changes in market interest rates. With a floater, interest rate volatility affects future interest payments.

The **quoted margin** is typically the specified yield spread over or under the reference rate, which is often LIBOR. It is used for compensating the investor for the *difference* in the credit risk of the issuer and that implied by the reference rate.

The **discount margin**, also known as the required margin, is the spread required by investors and to which the quoted margin must be set in order for the FRN to trade at par value on a rate reset date. Changes in the discount margin usually come from changes in the issuer's credit risk.

Money Market Instruments

Unique characteristics:

- Yield measures are annualized but not compounded.
- Often quoted using non-standard interest rates (discount rate or add-on rate? 360-day year or 365-day year? redemption value amount (FV) or price at issuance (PV)?)
- Different periodicities

Money market instruments need to be converted to a common basis for analysis.

Money market yield (also known as **CD equivalent yield**) is the annualized HPY on the basis of a 360-day year using simple interest.

Discount rate:

Note that the denominator is FV, not PV. The rate of return is therefore understated.

Add-on rate:

Note the only difference: the denominator is PV, not FV.

Bond equivalent yield: money market rate stated on a 365-day add-on rate basis.

Example

90-day T-bill, face value 100, quoted discount rate: 2.5% for an assumed 360-day year.

PV = 100 x (1 - 90/360 x 0.025) = 99.375

To calculate the bond equivalent yield for a 365-day year:

AOR = (365/90) x (100 - 99.375)/99.375 = 2.55%

7. The Maturity Structure of Interest Rates

```
  A <b>yield curve</b> is typically constructed on the basis of observed yields and matu
```

The most common type is the upward-sloping yield curve. The longer maturity issues have higher yields than the shorter maturity issues.

A **spot rate** is the yield on a zero-coupon bond. A series of spot rates (**spot curve**) can be used to discount the cash flows of a bond.

Default-free spot rates can be derived from the Treasury par yield curve by a method called **bootstrapping**. The basic principle of bootstrapping is that the value of a Treasury coupon security should be equal to the value of the package of zero-coupon Treasury securities that duplicate the coupon bond's cash flows.

Example

Determine the spot rate for the fourth period cash flow. The coupon rate is 4.11%, paid semi-annually.

The coupon for each period should be discounted at the corresponding spot rate.

$100 = 2.055/(1+0.03/2)^1 + 2.055/(1+0.033/2)^2 + 2.055/(1+0.035053/2)^3 + 102.055/(1+i/2)^4 = 2.0246 + 1.9888 + 1.9506 + 102.055/(1+i/2)^4$

i = 2.0669% and the annualized spot rate is 4.1339%.

A **par curve** is a sequence of yields-to-maturity in which each bond is priced at par value. A par curve is obtained from a spot curve. All bonds used to derive the par curve are assumed to have the same credit risk, periodicity, currency, liquidity, tax status, and annual yields.

A **forward rate** refers to the interest rate on a loan beginning some time in the future. In contrast, a spot rate is the interest rate on a loan beginning immediately. For example, the two-year forward rate one year from now is 4%. This means that if you borrow a two-year loan one year from now, you will pay an interest of 4%.

A **forward curve** is a series of forward rates, each with the same time frame.

Forward rate calculations are usually based on a theoretical spot rate curve. They are sometimes referred to as **implicit forward rates**.

Given spot rates for maturities of j and k years, you can compute the forward rate ($f_{j,\ k-j}$) that applies for the period from year j to year k using the relationship:

Example

Compute the annualized six-month forward rate two years from now.

In order to compute the six-month forward rate two years from now, first determine the spot rate for the fifth period (0.087) and the spot rate for the fourth period (0.0700). Then complete the following calculation: $[(1 + 0.0875/2)^5/(1 + 0.0700/2)^4] - 1 = 7.95\%$.

Similarly, implied spot rates can be calculated as geometric average of forward rates.

8. Yield Spreads

```
    A bond's yield-to-maturity can be separated into a benchmark and a spread.<p> </p><ul
```

- Benchmark rates are usually yields-to-maturity on government bonds or fixed rates on interest rate swaps.

Changes in benchmark rates (risk-free rate of return) capture macroeconomic factors that affect all bonds in the market: inflation, economic growth, foreign exchange rates, and monetary and fiscal policy.

- Changes in spreads (risk premium component) typically capture microeconomic factors that affect a particular bond: credit risk, liquidity and tax effects.

Different spread measures:

- **G spread**: the spread over or under a government bond rate, also known as the nominal spread. For example, suppose a 10-year, 8%-coupon bond is selling at $104.19, yielding 7.40%. The 10-year Treasury bond (6% coupon rate) has a YTM of 6.00%. Therefore, the G spread is 7.40% - 6.00% = 1.40%, or 140 basis points.
- **I spread**: the yield spread over or under the standard swap rate in that currency of the same tenor.

- **Z spread** (**zero volatility spread**): the constant yield spread over the benchmark spot curve such that the present value of the cash flows matches the price of the bond.

- **OAS** (**option-adjusted spread**): Z spread - option value. It is used for bonds with embedded options.

  - For callable bonds and bonds with prepayment options (e.g., most mortgage-backed and asset-backed securities), option cost > 0 and thus OAS < Z-Spread. The option cost is positive since the options are a detrimental to bondholders.
  - For putable bonds, option cost < 0 and thus OAS > Z-Spread.

Example

Phil Deter was interested in purchasing a non-Treasury bond for 110.2950. Given the Treasury spot rate data below, and assuming that the non-Treasury bond had a coupon of 9.60%, what is the likely Z-spread that Phil will earn over the duration of his investment?

It is important to add all of the cash flows for each bond (discounted at the appropriate spot rate) and compare these to the purchase price by trial-and-error.

The bond with a spread of 143 Basis Points has a purchase price of 4.70 + 4.58 + 4.46 + 4.32 + 4.15 + 4.00 + 84.08 = 110.2950. Since this purchase price corresponds with the bond corresponding with Phil's interest, the appropriate spread must be 143 Basis Points.

Introduction to Asset-Backed Securities

1. Benefits of Securitization for Economies and Financial Markets

```
<strong>Securitization</strong> repackages relatively simple debt obligations, such as
```

Securitization has several benefits:

- **Investors** can have direct access to mortgages and portfolios of receivable that would be unattainable if all the financing were performed through banks.
- **Banks** can remove assets from their balance sheets, therefore increasing the pool of available capital that can be loaned out.
- **Borrowers** can pay lower costs when borrowing.

The end result is lower cost and risk, more liquidity, and improved economic efficiency.

2. The Securitization Process

```
The process:<p> </p><ul class="notes">
```

- A lender originates loans, such as to a homeowner or corporation.
- The lender sells certain assets (e.g., loans) to a special purpose vehicle (SPV). The structure is legally insulated from management.
- The SPV issues debt, dividing up the benefits and risks among investors.
- Payments from borrowers are deposited into the SPV, then transferred to investors.

The parties to a securitization transaction:

- Originator: the seller of the collateral
- The SPV: the issuer of the securities, also called the trust
- The third parties: the loan servicer, attorneys, trustees, underwriters, rating agencies, and guarantors

The SPV is a bankruptcy-remote vehicle that plays a pivotal role in the securitization process. It issues securities backed by the underlying assets. The underlying assets are used as collateral for the securities. Cash flows generated from the underlying assets are used to service the debt obligations on the securities.

The SPV separates the assets used as collateral from the corporation seeking financing.

- It makes it possible that the asset-backed securities have a higher credit rating than the parent company.
- If bankruptcy occurs, the SPV can shield assets from the parent company's creditors.

Prepayment tranching refers to dividing cash flows from securitized assets among different classes of securities so that some receive repayment of principal before others. It is used to reallocate the prepayment risk of the underlying loans among different classes of securities. In the simplest cases, a deal might offer several classes of serially maturing securities. Some investors might prefer the securities with shorter maturities while others might favor the ones with longer maturities. Collateralized mortgage obligations (CMOs) are the most ubiquitous examples of time tranching.

Credit tranching refers to the creation of a multi-layered capital structure that includes senior and subordinated tranches (classes). The structure is designed so that any losses caused by defaults will be passed on to the subordinated tranches first. Credit tranching is thus used to reallocate the credit risk associated with the collateral.

3. Residential Mortgage Loans

```
   A <b>mortgage</b> is a loan secured by the collateral of a specified real estate prope
```

The interest rate on a mortgage loan is called the **mortgage rate** or **contract rate**. The ratio of the property's purchase price to the amount of the mortgage is called the **loan-to-value ratio**.

The basic idea behind the design of the **fixed-rate, level-payment, fully amortized mortgage loan** is that the borrower pays interest and repays principal in equal installments over an agreed-upon period of time, called the maturity or term of the mortgage. Thus at the end of the term the loan has been fully amortized.

Each monthly payment for this mortgage design is due on the first of each month and consists of:

- Interest of 1/12 of the fixed annual interest rate multiplied by the amount of the outstanding mortgage balance at the beginning of the previous month
- A repayment of a portion of the outstanding mortgage balance (principal)

The difference between the monthly mortgage payment and the portion of the payment that represents interest equals the amount that is applied to reduce the outstanding mortgage balance. Early payments are mostly interest with a small amount of principal repayment, while later payments are mostly principal repayment with a small amount of interest.

The following table illustrates the annual breakdown of a 30-year (360 month) mortgage on a $100k loan that yields 9.5% interest.

As the example shows, the portion of the monthly mortgage payment applied to interest *declines* each month, and the portion applied to principal repayment *increases*.

The various mortgage designs throughout the world specify:

- The maturity of the loan
- How the interest rate is determined (fixed rate, variable rate or hybrid rate)
- How the principal is repaid (i.e., whether the loan is amortizing or not, whether it is fully amortizing or partially amortizing with a balloon payment)
- Whether the borrower has the prepayment option and whether there are any prepayment penalties
- The rights of the lender in a foreclosure (whether the loan is a recourse or non-recourse loan)

Prepayments and Cash Flow Uncertainty

Mortgage borrowers may have an embedded option that allows them to prepay all or part of their loan at anytime during the life of the mortgage. Prepayments occur for one of several reasons:

- Homeowners prepay the entire mortgage when selling their home.
- Refinancing when market rates fall below the contract rate.
- In the case of homeowners who cannot meet their mortgage obligations, the property is repossessed and sold, and the proceeds are used to pay off the mortgage.
- If property is destroyed by fire or another insured catastrophe occurs, insurance proceeds are used to pay off the mortgage.

The prepayment could be the entire outstanding balance or a partial paydown of the mortgage balance. When a prepayment does not cover the entire outstanding balance it is called a **curtailment**.

Thus, lenders do not know with certainty what their cash flows (both the amount and timing) will be for any given period. This risk is referred to as **prepayment risk**. Usually, mortgages are prepaid when interest rates fall. Hence, lenders receive repayment at a time when they cannot reinvest the cash at the same high rate at which they had originally invested.

Mortgages with prepayment penalties originated in 1996 to deter prepayment. In this structure, there is a period of time during which, if the loan is prepaid in full or in excess of a certain amount of the outstanding balance, there is a prepayment penalty. This period is referred to as the **lockout period** or penalty period.

4. Mortgage Pass-Through Securities

```
   A <b>mortgage pass-through security</b> is created when one or more mortgage holders f
```

Below is an illustration of how the pass-through process channels mortgage payments from homeowners to coupon payments to mortgage bond investors.

In the U.S., there are two sectors for securities backed by residential mortgages:

- **Agency pass-throughs** have been pooled and securitized by one of the quasi-government mortgage agencies: Ginnie Mae, Fannie Mae and Freddie Mac. Note that only the Ginnie Mae securities are backed by the full faith and credit of the U.S. government.

- **Non-agency**, or **private-label mortgage-backed securities** have been pooled and securitized by private banks or other corporations. Unlike agency, non-agency MBSs must rely on various types of credit enhancements to compensate for the lack of a government credit guarantee.

Measures of Prepayment Rate

The cash flow of a mortgage pass-through security depends on the cash flow of the underlying pool of mortgages and consist of monthly mortgage payments representing interest, the scheduled repayment of principal, and any prepayments, net of servicing and other fees.

Estimating the cash flow from a pass-through requires forecasting prepayments. One way of forecasting is to assume that some fraction of the remaining principal in the pool is prepaid each month for the remaining term of the mortgages. The prepayment rate assumed for a pool, called the **conditional prepayment rate** (**CPR**), is based on a pool's characteristics (including its historical prepayment experience) and the current and expected economic environment. The CPR is an annual prepayment rate. For example, if the CPR is 8%, we would expect that 8% of the remaining principal will be prepaid each year.

The **single-monthly mortality rate** (**SMM**) is the percentage of a pool's remaining principal that is expected to be prepaid each month. SMM is the CPR converted from an annual term to a monthly rate.

Example: if the CPR is 10%, the SMM = $1 - (1 - 0.1)^{1/12} = 0.87\%$

Now suppose that the principal balance in a pool is $10 million and $200k is scheduled to be repaid in a given month. The SMM is 0.87%. The forecasted prepayment amount for the month is 0.0087 x (10,000,000 - 200,000) = $85,260.

Note that scheduled principal repayments are neither included in the starting principal nor in the prepayment forecast. Prepayment does not include scheduled repayment.

The **Public Securities Association Prepayment Benchmark** is a schedule of prepayment speeds deemed to be "usual." Actual prepayment speeds are quoted relative to the PSA benchmark. If prepayments are following the PSA schedule, the speed is said to be 100% PSA. If prepayments are twice as fast as the PSA benchmark, the speed is said to be 200% PSA.

Example: if a seven-year old mortgage pool is experiencing a CPR of 9%, Its relative to PSA speed is quoted as 150 PSA. The CPR for a seven-year old pool should be 6% according to PSA; 9% is 50% greater than the 6% PSA. Thus, the current speed is 150 PSA.

The PSA benchmark is generally not a good forecast. It is simply a way of communicating how quickly principal on an underlying pool of mortgages is being prepaid. Further, since each measure can be translated into the other, there is no real advantage or disadvantage in using anyone of these measures on the exam or in practice. All three are communicating the same thing in different ways.

Weighted Average Life

A **weighted average coupon** (**WAC**) is the weighted average of the mortgage rates of the mortgages in a pool, where each mortgage's weight is proportional to that mortgage's outstanding principal balance relative to the total of all the principal balances.

Example: suppose there are three mortgages with respective balances of $100k, $200k, and $300k. The mortgage rates are 8%, 9%, and 10% respectively. The weights are 100/600, 200/600, and 300/600. The WAC is (1/6 x 8%) + (2/6 x 9%) + (3/6 x 10%) = 9 1/3%.

A **weighted average maturity** (**WAM**) is a weighted average of the remaining maturities of the mortgages in the pool, where each mortgage's weight is proportional to that mortgage's outstanding principal balance relative to the total of all the principal balances.

Example: suppose there are three mortgages with respective balances $100k, $200k, and $300k. The remaining maturities are 200 months, 250 months, and 300 months. The weights are 1/6, 2/6, and 3/6, respectively. The WAM is thus (1/6 x 200) + (2/6 x 250) + (3/6 x 300) = 266 2/3 months.

Prepayment Risk

Prepayment risk encompasses contraction risk and extension risk.

Contraction risk is risk when interest rates decline and prepayments speed up. The timing of a pass-through security's cash flows is shortened.

Contraction risk has two components:

- Like a callable bond, a pass-through security has negative convexity, due to the borrower's embedded option to prepay. The upside potential is limited for investors.
- Cash flow must be reinvested at a lower rate than the original contract rate. The reinvestment rate risk is therefore high for pass-through securities.

Note that during contraction, duration drops when rates drop. That is, when interest rates drop, the sensitivity of the bond price to interest rates is dampened.

Extension risk is the risk that when interest rates rise, prepayments will slow. The timing of a pass-through security's cash flows is lengthened.

Extension risk also has two components:

- Anticipated funds are tied up for a longer period of time than originally expected and therefore cannot be reinvested at the new high rates.
- As rates rise, the pass-through security declines in value more than a non-callable bond would. This is because the delayed cash flows make the duration (interest rate sensitivity) rise.

Note that during extension, duration rises just when rates rise.

5. Collateralized Mortgage Obligations

```
<b>Collateralized mortgage obligations</b> are securities issued against a pool of mor
```

As previously mentioned, some institutional investors are concerned with extension risk and other with contraction risk. The mere creation of a CMO cannot eliminate prepayment risk; it can only distribute the various forms of this risk among different classes of bondholders. The technique of redistributing the coupon interest and principal from the underlying collateral to different classes (so that a CMO results in instruments that have varying convexity characteristics more suitable to the needs and expectations of different investors) broadens the appeal of mortgage-backed products to various traditional fixed-income investors.

A **tranche** is a slice of the cash flows generated by a mortgage pool. The claim of each tranche is governed by a specific formula. A CMO distributes prepayment risk among tranches so as to create products that provide better matching of assets and liabilities for institutional investors.

There are many types of CMO structures; three are discussed here.

Sequential-Pay Tranches

The first generation of CMOs was structured so that each tranche would be retired sequentially; such structures are referred to as **sequential-pay tranches**.

In a "plain vanilla" CMO structure, there may be four tranches: A, B, C and Z. The first three tranches, with tranche A representing the shortest-maturity bond, receive periodic interest payments from the underlying collateral. Tranche Z is an **accrual bond** that receives no periodic interest until the other three tranches are retired.

- When principal payments (both scheduled payments and prepayments) are received by the trustee for the CMO, they are applied toward retiring the tranche A bonds.
- After all the tranche A bonds have been retired, all principal payments received are applied toward retiring tranche B bonds.
- Once all the tranche B bonds have been retired, tranche C bonds are paid off in the same fashion.
- Finally, after the first three tranches of bonds have been retired, the cash flow payments from the remaining underlying collateral are used to satisfy the obligations on the **Z bonds** (original principal plus accrued interest). It is also called **accrual tranche**.

There is some protection provided against prepayment risk for each tranche. For example, prioritizing the distribution of principal effectively protects tranche A against extension risk (the protection coming from tranches B, C and D). Similarly, tranches C and D are protected against contraction risk.

Note that tranche Z (the accrual tranche) appeals to investors who are concerned with reinvest risk. Since there are no coupon payments to reinvest, reinvestment risk is eliminated until all the other tranches are paid off.

Planned Amortization Class Tranches

A **Planned Amortization Class** (**PAC**) bond is a CMO product that was created to have a similar cash flow structure to a sinking fund corporate bond within a specified range of prepayment rates (i.e., the cash flow pattern to the bond holder is known). The cash flow for PAC bonds is more predictable because there is a **principal repayment schedule** that must be satisfied. PAC bondholders, therefore, have priority over all other classes in the CMO issue in receiving principal payments from the underlying collateral in order to satisfy the repayment schedule.

The greater certainty regarding the cash flow for PAC bonds comes at the expense, of course, of the non-PAC classes, called the **companion** or **support classes**.

- If the actual prepayment speed is faster than the upper limit of the PAC range, the companion bonds receive the excess. This means that the companion bonds absorb the contraction risk.

- If the actual prepayment speed is slower than the lower limit of the PAC range, then in subsequent periods the PAC bondholders have priority for principal payments (both scheduled payments and prepayments). This reduces extension risk, which is absorbed by the companion bondholders.

The upper and lower PSA levels used to construct the principal payment schedule are called the **initial PAC collar**. A key consideration is that prepayment protection is ensured as long as companion bonds are not fully paid off. Consequently, the degree of prepayment protection changes over time as actually prepayments occur. For example, if prepayments over the first few years are at the lower end of the initial PAC collar, there will be more companion bonds remaining, which will result in greater prepayment protection for the PAC bonds. A new collar can be calculated, which will allow PAC bondholders to realize their original principal payment schedule as long as prepayments are within the collar. This new collar is called the **effective collar**. The effective collar is a wider range of prepayment speeds over which the life and cash flows of a PAC are predictable. An effective collar is necessary because the capacity of the support tranche to absorb prepayments is gradually diminished over the life of the security.

Support Tranches

A **companion bond** or a **support bond** absorbs the surplus or shortfall cash flows from a pool, allowing PAC bond to have a much more predictable series of cash flows. The support tranche is exposed to both contraction and extension risks. By definition, it is exposed to the greatest amount of prepayment risk.

- If too much principal is repaid, the overage goes first to the support tranche; the protected tranche receives only the scheduled amount.
- If not enough principal is prepaid, the support tranche gets none; the protected tranche gets all or nearly all of the promised amount.

Credit Enhancements

All non-agency asset-backed securities are credit-enhanced.

External credit enhancements are financial guarantees from third parties. The most common forms are:

- Monoline insurance companies. They guarantee the timely repayment of bond principal and interest when an issuer defaults. They are so named because they provide services to only one industry.
- Letter of credit from a bank
- Guarantee by the seller of the assets.

A guarantee does not completely remove the risk of default. Rather, it partially isolates it. Many factors can force an insured bond to default.

An **internal credit enhancement** is a tranche design or reserve structure that protects one or all investors against losses from default.

- A **senior-subordinated structure** is a two-tranche structure. The senior tranche gets paid first and the subordinated tranche gets paid only if there are enough funds left after the senior is paid. The subordinated tranche absorbs the credit risk, making the senior tranche less risky than the subordinated tranche.

The level of protection provided by the subordinated tranche changes over time due to prepayments. Prepayments change how much of the remaining pool is allocated to each of the two tranches. If the subordinated tranche gets prepaid early because of fast prepayment, the level of protection for the senior tranche declines. To guard against this problem, prepayments are allocated between the two tranches so that the percentage of the mortgage balance of the subordinated tranche to that of the mortgage balance for the entire deal, known as the **level of subordination**, is maintained at an acceptable level.

A commonly used shifting interest percentage schedule is as follows:

- If prepayments in month 30 are $2 million, the entire amount is paid to the senior tranche.
- If prepayments in month 100 (year 9) are $2 million, the senior tranche gets $400,000 (20%) and the subordinated tranche gets $1.6 million.

A structure can have more than one subordinated tranche.

- A **reserve account** is used by the sponsor to channel some of the pool's cash flows into a reserve to be paid out in the case of default or missed payments. Think of a reserve account as a rainy-day fund. There are two forms:
    - Cash Reserve Funds: The SPV sets aside a cash reserve, independent of underlying assets, to cover expected losses. The higher the reserve, the higher the credit rating.
    - Excess Spread Accounts: Underlying assets support a higher level of payment than that promised to security holders. For example, if gross WAC (weighted average coupon) is 8.00%, and after service fee is 7.75% (25bp for servicing the security) then the SPV may set up the ABS to pay 7.25%, using the 50bp spread to fund a reserve for expected losses.
- The amount of collateral backing the structure must be at least equal to the amount of the liability. If the amount of the collateral exceeds the amount of the liability of the structure, the deal is said to be **over-collateralized**. The amount of over-collateralization can be used to absorb losses. If the liability of the structure is $100 million and the collateral's value is $105 million, the first $5 million loss will not result in a loss to any of the tranches.

6. Commercial Mortgage-Backed Securities

Commercial mortgage-backed securities are securitizations of mortgage loans backed by

Credit Risk

The loans that serve as CMBS collateral are commonly secured by commercial real estate such as apartment buildings, shopping malls, warehouse facilities, etc. Unlike most residential mortgage loans, these loans often do not provide recourse to the borrower or any form of guarantee. Lenders and investors look to the collateral, not the borrower, for ultimate repayment. Thus, analysis of the cash flows generated by the underlying properties as well as their value is critical. Credit analysis should be performed on a loan-by-loan basis because of the unique economic characteristics of each income-producing property in a pool.

There are two relevant measures:

- The **debt-to-service coverage ratio** (DSC) is net operating income / debt service. Loans with a debt service coverage ratio above 1.00 have a lower likelihood of default because they have a built-in excess cash flow buffer available; this would have to erode before the borrower would experience losses and consider defaulting.
- The **loan-to-value ratio** (LTV) is the ratio of loan amount to the value of the collateral property. A lower LTV loan is considered more creditworthy due to its better default protection.

Basic CMBS Structure

The major structural component of a CMBS deal is *credit tranching*. To have AAA rated tranches, there must be enough credit support from tranches that absorb any losses on the underlying collateral first. The tranches with less underlying credit support have lower credit ratings and investors are rewarded with commensurately higher yields. When delinquencies and defaults occur, cash flows otherwise due to the subordinated class are diverted to the senior classes to the extent required to meet scheduled principal and interest payments. Thus, subordination allows issuers to create highly-rated securities from collateral of various levels of quality.

Call Protection

Call protection comes in two forms: at the structure level and at the loan level. The creation of sequential-pay tranches is an example of call protection at the structure level.

Call protection at the loan level comes in several forms, including **prepayment lockout** (usually two - five years), and stiff **prepayment penalties** that serve as a deterrent to the borrowers. **Treasury make-whole** (**yield maintenance**) is a common form of prepayment penalty that requires the borrower to accompany any prepayment with a premium which, when reinvested by the loan owner in Treasuries for the remaining term of the loan (had it not been prepaid), would exactly recreate the lost yield on the prepaid loan. **Fixed-percentage prepayment penalties** are also common, and require the prepaying borrower to pay a premium equal to a set of percentages of the balance being prepaid.

An innovation in CMBS call protection is **Treasury defeasance**. This concept is similar to Treasury yield maintenance in that, instead of prepaying the loan, the borrower substitutes Treasury securities to replicate the cash flows of the mortgage.

Balloon Maturity Provisions

CMBS investors face two credit issues with respect to the CMBS mortgage pool: **operational defaults** (the risk that the property will not generate sufficient cash flow to make the monthly payments on the mortgage loan), and **refinance risk** (the risk that, at maturity, the property will lack sufficient value to be sold or refinanced in an amount sufficient to make the balloon payment.) Commercial mortgage loans usually balloon after 10 or 15 years and must be paid in full. Note that balloon loans have short maturities but longer amortizing terms, resulting in a lump sum payment due at maturity, which makes default highly likely if the borrower cannot refinance.

7. Non-Mortgage Asset-Backed Securities

```
   Asset-backed securities are backed by a wide range of asset types.<p> </p>
```

Auto Loan Receivable-Backed Securities

- Underlying collaterals: amortizing auto loans and lease receivable
- Prepayment risk due to repossession, early payoff, insurance settlement from wreck, sale of vehicle, or refinancing (rare)
- Credit enhancement: senior/subordinated structure, reserve account, overcollateralization, excess interest on receivables

Credit Card Receivable-Backed Securities

- Collateral is non-amortizing loans
- Fixed or floating interest rates
- The lockout period is the amount of time that principal repayments are reinvested in new receivables rather than returned to security holders. During this period the cash flow paid out to security holders is based only on finance charges collected and fees. The principal amortizing period is from the end of the lockout period through the end.
- Early amortization is a type of credit enhancement. It is usually triggered when there is a sudden increase in delinquencies in the underlying loans or when excess spread, the issuer's net profit after deducting servicing fees, charge-offs and other costs, falls below an acceptable level.

8. Collateralized Debt Obligations

Collateralized debt obligations (CDOs) are a type of structured asset-backed security

CDO Structure

Collateralized loan obligations (CLOs) are CDOs backed primarily by leveraged bank loans. Collateralized bond obligations (CBOs) are CDOs backed primarily by leveraged fixed-income securities.

CDOs are assigned different risk classes, or tranches.

- *A senior tranche:* between 70% and 80% of the deal and receives a floating-rate payment
- *Subordinated or mezzanine tranches:* receive a fixed coupon rate
- *Equity tranche:* provides equity protection to other tranches. It receives any remaining interest that is received from the collateral but not paid to the senior and mezzanine tranches.

Problem: The majority of investors are paid a floating rate, whereas the underlying pool bonds pay a fixed rate. If rates rise, the **collateral manager** can get burned. This is because the manager could find itself having to pay out an increasingly higher rate to tranche holders while its source of funds - interest on the underlying bonds - is fixed.

Solution: The manager must protect against rising rates. Entering a swap as the fixed payer (variable receiver) solves the problem, as this position provides positive cash flow when interest rates rise.

Example

Consider the following CDO transaction:

- The CDO is a \$100 million structure.
- The collateral consists of bonds that all mature in five years. The coupon rate of these bonds is the five-year T-bond rate plus 500 basis points.
- The senior tranche is 75% of the deal. It pays a floating rate of LIBOR + 50 basis points.
- There is only one mezzanine tranche of \$10 million with a coupon rate of the five-year T-bond rate + 400 basis points.
- The asset manager enters into a swap in which it pays a fixed rate equal to the five-year T-bond rate + 150 basis points and receives LIBOR. The notional amount of the swap is \$75 million.
- Assume that there is no default or asset management fee. All payments are made annually each year for simplicity.

Analysis

The equity tranche is \$100 - \$100 x 0.75 - 10 = \$15 million.

Each year:

- The collateral will pay interest of \$100 x (T-rate + 5%) million to the CDO.
- The CDO pays \$75 x (T-rate + 1.5%) to the counterparty of the swap and receives \$75 x LIBOR.
- Interest to senior tranche: \$75 x (LIBOR + 0.5%)
- Interest to mezzanine tranche: \$10 x (T-rate + 4%)

Netting the interest payments paid and received:

\$100 x (T-rate + 5%) - \$75 x (T-rate + 1.5%) + \$75 x LIBOR - \$75 x (LIBOR + 0.5%) - \$10 x (T-rate + 4%) = (\$15 x T-rate + \$3.1) million

If the five-year T-rate at the time the CDO is issued is 5%, the amount available each year for the equity tranche is \$15 x 0.05 + 3.1 = \$3.85 million.

Cash CDOs

Cash CDOs involve a portfolio of cash assets, such as loans, corporate bonds, asset-backed securities or mortgage-backed securities. Ownership of the assets is transferred to the legal entity (a SPV) issuing the CDOs' tranches. The risk of loss on the assets is divided among tranches in reverse order of seniority.

Motivation - Arbitrage vs. Balance Sheet

- Arbitrage transactions attempt to capture for equity investors the spread between the relatively high yielding assets and the lower yielding liabilities represented by the rated bonds. The majority of CDOs are arbitrage-motivated.
- Balance sheet transactions, by contrast, are primarily motivated by the issuing institutions' desire to remove loans and other assets from their balance sheets, to reduce their regulatory capital requirements and improve their return on risk capital. A bank may wish to offload credit risk in order to reduce its balance sheet's credit risk.

9. Covered Bonds

```
<p> </p><b>Covered bonds</b> are securities issued by a bank or mortgage institution a
```

The institutions may replace prepaid, non-performing or defaulted loans with performing loans to minimize the risk of the underlying assets.

There is usually on bond class per cover pool.

Covered bonds typically have AAA ratings. They are safer and offer lower yields than otherwise similar ABS.

Fixed Income (2)

Understanding Fixed-Income Risk and Return

1. Sources of Return

```
There are three sources of return on a fixed-rate bond:<p> </p><ul class="notes">
```

- Receipts of the promised coupon and principal payments on the scheduled dates.
- Reinvestment income - interest income generated by reinvesting coupon interest payments. Interest income, which is the sum of coupon payments and reinvestment income, is the return associated with the passage of time.
- Any capital gains or losses if the bond is sold prior to maturity. These are caused by a change in the yield-to-maturity.

Example

Harshal Shahe purchases a bond, but isn't sure how much of his total return will come from reinvested interest compared to coupon interest and capital gain. What is the total reinvested interest (i.e., interest on interest) that Herschel earns, assuming a price of $941.12, coupon of 15.00%, 18.5 year maturity and a market interest rate of 16.00%?

First calculate total future cash flows: PV x $(1+r)^N$ = Price of bond of $941.12 x (1 + Market Interest rate 16.00% divided by $2)^{37}$ = $16,230.27

less initial purchase price = $941.12

less coupon interest payments = 15.00% * $1000 * 18.50 = $2,775.00

less capital gain (add capital loss) = $1000 - $941.12 = $58.88

= $12,455.27

The yield-to-maturity measures an investor's return from the bond correctly only if these assumptions are true:

- The bond is held to maturity.
- The coupon reinvestment rate is the same as the YTM.
- The issuer does not default.

The total return is the sum of:

- the future value of reinvested coupon payments.
- the sale price, or redemption of principal if the bond is held to maturity.

There are two types of interest rate related risks:

- **Reinvestment risk**. Future interest rates may be less than the YTM. Two factors can affect the degree of reinvestment risk:

  - **Maturity**. The longer the maturity, the higher the reinvestment risk. This implies that the yield to maturity measure for long-term coupon bonds tells little about the potential return that an investor may realize if the bond is held to maturity. For long-term bonds, in high-interest rate environments the reinvestment income component may be as high as 70% of the bond's potential total dollar return.
  - **Coupon rate**. The higher the coupon rate, the higher the reinvestment risk. This implies that a bond selling at a premium will be more dependent on reinvestment income than a bond selling at par. Zero-coupon bonds have no reinvestment risk if held to maturity because there is no periodic cash flow to be reinvested.

- **Market price risk**. If the bond has to be sold prior to maturity, its sale price will be lower if rates are higher.

These risks offset each other to a certain extent. Which one dominates depends on the bondholder's investment horizon. The shorter the investment horizon, the smaller the coupon reinvestment risk, but the bigger the market price risk.

2. Macaulay, Modified and Effective Durations

```
   Bond duration measures the sensitivity of the full price change to a change in interes
```

- **Yield duration** statistics measure the sensitivity of a bond's full price to the bond's own yield-to-maturity. They include the Macaulay duration, modified duration, money duration, and price value of a basis point.
- **Curve duration** statistics measure the sensitivity of a bond's full price to the benchmark yield curve, e.g., effective duration.

Duration is the weighted average time to receive the present value of each of the bond's coupon and principal payments. For example, a bond with a duration of three means that, on average, it takes three years to receive the present value of the bond's cash flows.

Macaulay Duration

Frederick Macaulay developed the concept of duration approximately 80 years ago. He demonstrated that a bond's duration was a more appropriate measure of time characteristics than the term to maturity of the bond, because duration incorporates both the repayment of capital at maturity, the size of the coupon and timing of the payments.

Macaulay duration is defined as the weighted average time to full recovery of principal and interest payments. The weights are the shares of the full price corresponding to each coupon and principal payment.

Alternatively, Macaulay duration can be calculated using a closed-form formula.

Modified Duration

Modified duration shows how bond prices move proportionally with small changes in yields. Specifically, modified duration estimates the percentage change in bond price with a change in yield.

$-D_{mod}$ = the modified duration for the bond

Di = yield change in basis points divided by 100

P = beginning price for the bond

Modified duration assumes that the price/yield relationship is a straight line. However, the price/yield relationship is convex, not linear. Suppose that the bond has an initial yield of Y_0. A tangent line can be drawn to the price/yield relationship at Y_0. The slope of the tangent line is related to the duration of the bond. If the yield falls to Y_1, the price will rise to P_1. Due to the linear assumption, the price change measured by duration is P_2 - P_0.

To approximate modified duration:

V_- = the price if yields declines

V_+ = the price if yield increases

V_0 = the initial price

For example, consider a 9% coupon 20-year option-free bond selling at 134.6722 to yield 6%. If the yield is decreased by 20 basis points from 6.0% to 5.8%, the price would increase to 137.5888. If the yield increases by 20 basis points, the price would decrease to 131.8439. Thus: ApproxModDur = (137.5888 - 131.8439)/(2 x 134.6722 x 0.002) = 10.66. This tells you that for a 1% change in the required yield, the bond price will change by approximately 10.66%.

Macaulay duration is mathematically related to modified duration.

A bond with a Macaulay duration of 10 years, a yield to maturity of 8% and semi-annual payments will have a modified duration of:

Dmod = 10/(1 + 0.08/2) = 9.62 years

Effective Duration

Effective duration measures interest rate risk in terms of a change in the *benchmark yield curve*. It is very similar to approximate modified duration.

A pricing model can be used to estimate the price change resulting from a change in the benchmark yield curve instead of the bond's own yield-to-maturity. V- and V+ are adjusted to reflect any changes in the cash flows (due to embedded options) that result from the change in benchmark yield curve.

Effective duration should be used for bonds with embedded options.

Key-Rate Durations

It is important to distinguish interest rate risk from yield curve risk.

- The interest rate risk is the sensitivity of a bond to parallel shifts of the yield curve.
- The yield curve risk is a bond's sensitivity to changes in the shape of the yield curve.

Parallel shifts in the yield curve rarely occur. An analyst may want to measure the change in the bond's price by changing the spot rate for a particular key maturity and holding the spot rate for the other key maturities constant. The **key rate duration** is the sensitivity of the value of a bond to changes in a single spot rate, holding all other spot rates constant. There is a key rate duration for every point on the spot rate curve so there is a vector of durations representing each maturity on the spot rate curve.

3. Properties of Bond Duration

```
    Bond duration is affected by many variables.<p> </p><ul class="notes">
```

- The fraction of the period that has gone by (t/T). A plot of Macaulay duration (or modified duration) against time for a single bond with constant yield will show a *saw-tooth pattern*, with Macaulay duration declining steadily until a coupon payment results in an upwards jump.

- The Macaulay duration of a *zero-coupon bond* is its time-to-maturity.

- The Macaulay duration of a *perpetual bond* (perpetuity) is (1 + r) / r.

- *Coupon rate* is inversely related to Macaulay duration and modified duration.

- *Yield-to-maturity* is also inversely related to Macaulay duration and modified duration.

- *Time-to-maturity* and Macaulay and modified duration are usually positively related.

    - They are *always* positively related on bonds priced at par or at a premium above par value.
    - They are *usually* positively related on bonds priced at par or at a discount below par value. The exception is long-term, low coupon bonds, on which it is possible to have a lower duration than on an otherwise comparable shorter-term bond.

Callable Bonds

A callable bond exhibits positive convexity at high yield levels and negative convexity at low yield levels. Negative convexity means that for a large change in interest rates, the amount of the price appreciation is less than the amount of the price depreciation.

- When the required yield for the callable bond is higher than its coupon rate, the bond is unlikely to be called. Therefore, the callable bond will have a similar price/yield relationship (positive convexity) as a comparable option-free bond.
- When the required yield becomes lower than the coupon rate, the value of the call option increases because it is getting more and more likely that the bond may be retired at the call price. The call price will set an upper limit on the price of the callable bond. In contrast, for an option-free bond, the bond price will rise unabated as the yield falls. If the required yield rises (but not higher than the coupon rate), the price of the non-callable bond falls and the price of the call option falls. As the price of a callable bond is the difference between the price of the non-callable bond and the price of the embedded option, the price of a callable bond will not fall as much as a non-callable bond. Therefore, a callable bond exhibits negative convexity at low yield levels.

Putable Bonds

The difference between the value of a putable bond and the value of an otherwise comparable option-free bond is the value of the embedded put option.

- When the required yield for the putable bond is low relative to the issuer's coupon rate, the price of the putable bond is basically the same as the price of the option-free bond because the value of the put option is small. An investor will not sell the bond to the issuer at the put price. Therefore, the putable bond will have a similar price/yield relationship to a comparable option-free bond.
- As rates rise, the price of the putable bond declines, but the price decline is less than that for an option-free bond. The value of the put option increases because it's getting more and more likely that the investor will sell the bond to the issuer at the put price. Therefore, the put price sets a lower limit on the price of the putable bond.

4. Bond Portfolio Duration

```
    There are two ways to calculate the duration of a bond portfolio:<p> </p><ul class="nc
```

- *The weighted average of the time to receipt of aggregate cash flows.* This method is based on the **cash flow yield**, which is the internal rate of return on the aggregate cash flows.

Limitations: This method cannot be used for bonds with embedded options or for floating-rate notes due to uncertain future cash flows. The cash flow yield is not commonly calculated. The change in cash flow yield is not necessarily the same as the change in the yields-to-maturity on the individual bonds. Interest rate risk is not usually expressed as a change in the cash flow yield.

- *The weighted average of the durations of individual bonds that compose the portfolio.* The weight is the proportion of the portfolio that a bond comprises.

$$\text{Portfolio Duration} = w_1 D_1 + w_2 D_2 + w_3 D_3 + \ldots + w_k D_k$$

w_i = the market value of bond i / market value of the portfolio
D_i = the duration of bond i
k = the number of bonds in the portfolio

This method is simpler to use and quite accurate when the yield curve is flat. Its main limitation is that it assumes a parallel shift in the yield curve.

To illustrate this calculation, consider the following three-bond portfolio in which all three bonds are option-free:

- 10% 5-year 100.0000 10 $4 million $4,000,000 3.861
- 8% 15-year 84.6275 10 $5 million $4,231,375 8.047
- 14% 30-year 137.8586 10 $1 million $1,378,586 9.168

In this illustration, it is assumed that the next coupon payment for each bond is exactly six months from now (i.e., there is no accrued interest). The market value for the portfolio is $9,609,961. Since each bond is option-free, modified duration can be used.

- w1 = $4,000,000/$9,609,961 = 0.416, D1 = 3.861
- w2 = $4,231,375/$9,609,961 = 0.440, D2 = 8.047
- w3 = $1,378,586/$9,609,961 = 0.144, D3 = 9.168

The portfolio's duration is: 0.416 (3.861) + 0.440 (8.047) + 0.144 (9.168) = 6.47.

A portfolio duration of 6.47 means that for a 100 basis point change in the yield for each of the three bonds, the market value of the portfolio will change by approximately 6.47%. Keep in mind that the yield for each of the three bonds must change by 100 basis points for the duration measure to be useful. This is a critical assumption and its importance cannot be overemphasized.

5. Money Duration of a Bond and the Price Value of a Basis Point

```
   Modified duration measures the <i>percentage price change</i> of a bond to a change ir
```

Money Duration = Dirty Price x Modified Duration

To calculate absolute price change:

Î”Dirty Price = - Money Duration x Î”Yield

In the U.S., money duration is called **dollar duration**. It is the approximate dollar change in a bond's price for a 100 basis point change in yield.

The **price value of a basis point** (PVBP) is the absolute change in the price of a bond for a one basis point change in yield. It is simply the money duration of a bond for a one basis point change in yield.

Example

Scott Marsh from Mass Avenue Research Management purchased a bond for a price of 93.555. This bond has a coupon of 14.70% and a modified duration of 3.00. Given market interest rates of 10.00% and a change in market rates of -66 basis points, what is the price value of a basis point?

The answer is the modified duration x 1 basis point x bond price or 3.00 x .0001 x 93.555 = $0.0281

Note: The other information has been placed into this question as a distraction. Don't be fooled by extraneous data!

To calculate PVBP:

P_- is the full price calculated by lowering the yield-to-maturity by one basis point.

P_+ is the full price calculated by raising the yield-to-maturity by one basis point.

6. Bond Convexity

```
   Duration is a first approximation of a bond's price or a portfolio's value to rate cha
```

Duration always gives a lower than actual price, the reason being convexity. Thus, a convexity adjustment would take into account the curvature of the price/yield relationship in order to give a more accurate estimated price.

To improve the estimate provided by duration, particularly for a large change in yield, a **convexity** measure can be used.

For a hypothetical 9%, 20-year bond selling to yield 6%, for a 20 basis point change in yield, P_0 = 134.6722, P_- = 137.5888, and P_+ = 131.8439 ==> convexity measure = (131.8439 + 137.5888 - 2 x 134.6722)/(134.6722 x 0.002^2) = 163.92.

Convexity indicates that as yield increases, the price of a bond declines at a declining rate. Given the convexity measure, the convexity adjustment to the duration estimate can be computed; the convexity adjustment is the amount that should be added to the duration estimate for the percentage price change.

Convexity Adjustment

Consider a situation where you are using duration to compute the effect of a 250 basis point change in yield, where duration is 6.655 and the convexity adjustment is 1.8271. Using an estimate of modified duration, you determine that the percentage change in price of this bond resulting from a 250 basis point increase in yield should be 2.5 x -6.655 = -16.6375%. Adding the convexity adjustment, the percentage price should change by -16.6375 + 1.8271 = -14.8104%. Summarizing the price change estimates in response to the 250 basis point increase in yield, you have:

Duration estimate = -16.63750%

Convexity adjustment + 1.8271%

Total: - 14.8104%

The actual decrease is 14.95%, so the convexity adjustment does improve the estimate.

If you estimate the change resulting from a 250 basis point decrease in yield, the results can be summarized as:

Duration estimate: 16.6375%

Convexity adjustment: + 1.8271%

Total: +18.4646%

The actual percentage increase in price is 18.62%. The convexity adjustment brings the modified duration estimate closer to the actual percentage change.

7. Interest Rate Risk and the Investment Horizon

```
<b>Yield Volatility</b><p> </p>
```

Prices of fixed income securities are affected by both the level of interest rates and the volatility of interest rates.

The risk of a default-free bond stems from two sources - interest rate shifts and risk of changes in the volatility of interest rates. The first type of risk is well-known. Managing interest rate risk requires measuring it first. Duration analysis has become an important tool, allowing portfolio managers to measure the sensitivity of their portfolios to changes in the level of interest rates.

The second type of risk is less familiar, although it can represent a major component of the total risk of a fixed-income portfolio. The greater the expected yield volatility, the greater the interest rate risk for a given duration and current value of a position.

Investment Horizon, Macaulay Duration, and Interest Rate Risk

The investment horizon is essential in measuring the interest rate risk of a fixed-rate bond.

When there is a parallel shift to the yield curve, the yield-to-maturity and coupon reinvestment rates are assumed to change by the same amount in the same direction. The Macaulay duration statistic identifies investment horizon so that the losses (or gains) from coupon reinvestment offset the gains (or losses) from market price changes.

The **duration gap** is the difference between the Macaulay duration and the investment horizon.

- When the investment horizon is *greater than* the Macaulay duration of the bond, coupon reinvestment risk dominates price risk. The investor's risk is to lower interest rates. The duration gap is *negative*.

- When the investment horizon is *equal to* the Macaulay duration of the bond, coupon reinvestment risk offsets price risk. The investor is hedged against interest rate risk. The duration gap is *zero*.
- When the investment horizon is *less than* the Macaulay duration of the bond, price risk dominates coupon reinvestment risk. The investor's risk is to *higher* interest rates. The duration gap is *positive*.

8. Credit and Liquidity Risk

```
A change in yield-to-maturity will cause a change in bond price. What is the source of
```

The yield-to-maturity on a corporate bond has two components:

- Government benchmark yield. A change in the yield can come from a change in either of these two components:
  - Expected inflation rate.
  - Expected real rate of interest.
- A spread over government benchmark. A change in the spread can come from a change in either of these two components:
  - Credit risk of the issuer. This involves the probability of default and degree of recovery if default occurs.
  - Liquidity of the bond. This refers to the transaction costs associated with selling a bond.

Regardless of the source of the yield-to-maturity change, the bond price change caused by a change in the yield-to-maturity will be the same.

In practice, there is often interaction between changes in benchmark yields and in the spread over the benchmark.

9. Empirical Duration

```
<p> </p>The purpose of calculating duration is to estimate a bondâ€™s interest-rate ri
```

Empirical duration is the calculation of a bond's duration based on historical data rather than a preset formula, like effective duration does. Simply put, empirical duration is more practical in that it uses historical data to model a bond price's sensitivity to different interest rate scenarios. When the historical yields rise or fall, the historical bond prices will fall or rise accordingly, and this data forms the basis for empirical duration.

The advantages of empirical duration include that the estimate does not rely on theoretical formulas and analytic assumptions; the investor only needs a reliable series of bond prices and a reliable series of Treasury yields. It is a better measure especially under stressed market conditions, for a portfolio consisting of a variety of different bonds from different issuers.

Disadvantages include that a reliable series of a bond's price may not be available, and the series of prices that is available might not be market based, but rather modelled or matrix priced (the price is based on similar security).

Fundamentals of Credit Analysis

1. Credit Risk

```
<b>Credit risk</b> is the risk of loss of interest and/or principal stemming from a bc
```

Credit risk has two components:

- **Default probability** addresses the likelihood that a borrower will default on its debt obligations,

without reference to estimated loss.
- **Loss severity**, also known as **Loss Given Default** (LGD), measures the portion of value an investor loses. If a bond defaults, investors can still expect to recover a certain percentage of the bond; that percentage is called the recovery rate. Loss severity = 1 - **recovery rate**

Expected loss = Default probability x Loss severity

The spread refers to the difference between the yield on a specific bond and a comparable maturity (or duration) Treasury. The part of the risk premium representing the default risk is known as the **credit spread**. If the perception of risk increases for the issuer or for the industry category representing the issuer, the spread may increase or widen. This risk associated with an increasing credit spread is known as the **credit spread risk**. If there are more concerns about economic security, the spread will widen (implying that the premium for risk increases).

Credit risk could be on account of:

- **Downgrade risk**: the risk that the issuer will be downgraded, resulting in an increase in the credit spread demanded by the market. The market tends to respond very quickly to news regarding a bond rating decline.

- **Market liquidity risk**: the widening of the bid-ask spread on an issuer's bonds. The size and the credit quality of the issuer affects market liquidity risk.

2. Seniority Ranking and Priority of Claims

```
    In finance, <b>seniority</b> refers to the order of repayment in the event of a sale c
```

- Secured debt holders get paid first.
- Unsecured debt holders get paid before equity owners.
- Senior creditors take priority over junior (subordinated) creditors.

The priority of claims is not always absolute. It can be influenced by several factors, such as government involvement, leeway accorded to bankruptcy judges, and the bias toward reorganization instead of liquidation.

3. Credit Ratings

```
    The rating agencies (Moody's, S&P, and Fitch) rate both issuers and issues. Issuer rat
```

Notching

A company's credit rating corresponds to its senior unsecured obligations. A rating agency may notch up secure debt from the company credit rating and notch down subordinated debt. A credit rating agency's notching policy primarily intends to reflect the relative recovery prospects of different instruments issued by the same issuer.

Risks in Relying on Agency Ratings

There are risks in relying too much on credit agency ratings.

Because creditworthiness is dynamic, initial/current ratings do not necessarily reflect the evolution of credit quality over an investor's holding period. Importantly, bond ratings do not always capture price risk because valuations often adjust before ratings change and the notching process may not adequately reflect the price decline of a bond that is lower ranked in the capital structure. Similarly, because ratings primarily reflect the probability of default but not necessarily the severity of loss given default, bonds with the same rating may have significantly different expected losses. And like analysts, credit rating agencies may have difficulty forecasting certain credit-negative outcomes, such as adverse litigation, leveraging corporate transactions, and such low likelihood/high severity events as earthquakes and hurricanes.

4. Credit Analysis

Credit analysts want to assess a company's ability to make timely payments of interest

The "four Cs" of credit analysis provide a useful framework for evaluating credit risk.

Capacity

The capacity, or ability to pay, reflects the funds flow from the organization and the generation of cash sufficient to meet the interest and principal repayments.

Credit analysis starts with industry analysis followed by company analysis to assess the cash flows or the ability of the issuer to repay its financial obligation.

Industry structure. Michael Porter's framework, which is covered in Reading 37 [Introduction to Industry and Company Analysis], can be used to analyze industry structure.

Industry fundamentals. These include the industry's sensitivity to macroeconomic factors, its growth prospects, its profitability and its business needs.

Company fundamentals. These include the company's competitive position, track record, management's strategy and execution, and ratio analysis. The ratios can be categorized into three groups: profitability and cash flow ratios, leverage ratios, and coverage ratios.

How does an analyst who has calculated a ratio know whether it represents good, bad, or indifferent credit quality? The analyst must relate the ratio to the likelihood that the borrower will satisfy all scheduled interest and principal payments in full and on time. In practice, this is accomplished by testing financial ratios as predictors of the borrower's propensity not to pay (to default). For example, a company with high financial leverage is statistically more likely to default than one with low leverage, all other things being equal. Similarly, high fixed-charge coverage implies less default risk than low coverage. After identifying the factors that create high default risk, the analyst can use ratios to rank all borrowers on a relative scale of propensity to default.

An issuer's ability to access liquidity is also an important consideration in credit analysis.

Collateral

Collateral analysis involves not only the traditional pledging of assets to secure the debt, but also the quality and value of those un-pledged assets controlled by the issue. Note that the value and quality of a company's assets may be difficult to observe directly. The key point is to assess the value of the assets relative to the issuer's debt level.

Covenants

Covenants deal with limitations and restrictions on the borrower's activities. They are important because they impose restrictions on how management operates the company and conducts its financial assets. This term covers both affirmative (obligated to do) and negative (limited in doing) covenants.

Character

Character relates to the ethical reputation as well as the business qualifications and operating record of the board of directors, management, and executives responsible for the use of the borrowed funds and its repayment. It covers many aspects, such as strategic direction, financial philosophy, conservatism, track record, succession planning, control systems, etc.

5. Credit Risk vs. Return: Yields and Spreads

The higher the credit risk, the greater the required yield and potential return demanc

Yield spread is the difference in yield between two securities.

- The yield of a corporate bond = yield on a risk-free bond + yield spread
- The yield spread here is composed of the liquidity premium and the credit spread: yield spread = liquidity premium + credit spread

Yield spreads, especially credit spreads, become wider during economic contractions. In times of credit improvement or stability, however, credit spreads can narrow sharply as well. This is known as *"flight to quality"*.

Factors that affect yield spreads include: the credit cycle, economic conditions, financial market performance, market making capacity, and supply/demand conditions.

How do spread changes affect the price of and return on these bonds? The impact depends on two factors:

- The basis point spread change
- The sensitivity of price to yield as reflected by modified duration and convexity

A **credit curve** is essentially the spread over treasuries of various maturities for a single bond issuer. It is typically upward-sloping, meaning the longer the bond maturity, the wider the spread.

6. Special Considerations of High-Yield, Sovereign, and Municipal Analysis

```
<b>High-Yield Bonds</b><p> </p>
```

High-yield bonds are issued by organizations that do not qualify for investment-grade ratings. These issuers must pay a higher interest rate to compensate investors for the increased risks. In analyzing the creditworthiness of high-yield corporate bonds, an analyst should pay close attention to the following:

- Liquidity: how liquid is the issuer? What is its ability to generate cash as needed?
- Projections of future earnings and cash flow
- Debt structure analysis (debt seniority)
- Corporate structure: what are the relationships among all of the subsidiaries? Are the subsidiaries potentially contributing to or draining resources from the creditors?
- Covenants: change of control put, payment restrictions, limitations on liens and additional indebtedness, restricted versus unrestricted subsidiaries, etc.
- Equity-like approach to high yield analysis

Equity-Like Approach

Traditionally, high-yield bonds have provided a greater return than high-grade bonds, but lower than equities. Similarly, high-yield bond risk has been higher than that of investment grade bonds, but less than equities. High-yield bonds have historically been more highly correlated with equity securities than with investment-grade bonds. Thus, some analysts believe that an equity analysis approach will provide a better framework for high-yield bond analysis than a traditional credit approach.

An equity-like approach to high-yield analysis can be helpful. Calculating and comparing enterprise value with EBITDA and debt/EBITDA can show a level of equity "cushion" or support beneath an issuer's debt.

Sovereign Debt

Two key issues for sovereign analysis:

- A government's ability to pay.
- A government's willingness to pay.

Both quantitative and qualitative analyses are employed in assessing sovereign risk with ratings performed in both local currency and foreign currency. It is important to evaluate the ratings in both currencies since historically the default rate on foreign currency debt has been greater than the default rate on local (or domestic) currency debt; there is different risk in the two ratings. Generally, if an issuer is planning to default, it is more likely to do so with a foreign currency issue. Thus, the ratings need to be performed for both types of issues.

A framework is presented in the reading. It highlights five broad areas:

- Institutional effectiveness and political risks
- Economic structure and growth prospects
- External liquidity and international investment position
- Fiscal performance, flexibility, and debt burden
- Monetary flexibility

Municipal Debt

There are two basic types of municipal bonds:

General obligation (GO) bonds depend on the general creditworthiness of a municipality to repay the debt. The credit analysis has some similarities to sovereign analysis. In general, a municipal analyst should look at employment, industry, and real estate valuation trends needed to generate taxes and fees.

Revenue bonds support specific projects. The credit analysis is identical to that of a corporate bond analysis. The focus is to assess whether or not the underlying cash flows from the project will be sufficient to meet the obligations.

Derivatives

Derivatives

Derivative Markets and Instruments

1. Introduction

```
   <p> </p>A <b>derivative</b> is a financial instrument that offers a return based on th
```

Derivatives usually transform the performance of the underlying asset. They are similar to insurance in that they allow for the transfer of risk from one party to another. The risk itself does not change, but the party bearing the risk does. The underlying asset is the source of the risk, referred to as the **underlying** - which does not always have to be an asset. The underlying could also include interest rates, credit, energy, weather, etc.

Derivatives are created in the form of legal contracts involving two parties, the buyer and the seller. The seller is sometimes known as the writer or the **short** party in the contract. The buyer, who purchases the derivative, is referred to as the **long** or the holder. The derivative contract always defines the rights and obligations of each party, and a legal system recognizes these.

There are two classes of derivatives - forward commitments and contingent claims. **Forward commitments** force the two parties to transact in the future at a previously agreed-on price. A **contingent claim** provides the right but not the obligation to buy or sell the underlying at a pre-determined price.

Benefits of Derivatives

Derivatives can be used to implement strategies that cannot be achieved with their underlying's alone. Derivatives have as an inherent feature a high degree of leverage. This means that investors typically only commit small amounts of money to a derivative position relative to the equivalent position in the underlying asset. Small movements in the underlying can lead to large movements in the derivative - both positive and negative. This has the effect of attracting lots of speculators in the derivative market looking for large gains. Furthermore, derivatives generally trade at low transaction costs in liquid markets.

There are numerous applications in **risk management** practice where the use of derivatives provides a useful tool for managing exposure to particular risks. For example, many financial institutions act as **hedgers**, meaning they use derivatives to reduce or eliminate certain forms of risk.

In addition, **arbitrageurs** use the derivative market to simultaneously buy and sell similar assets in different markets, creating a riskless profit while at the same time improving market efficiency.

2. Exchange-Traded versus Over-the-Counter Derivatives

```
    <p> </p>Based on the markets where they are created and traded, derivatives can be cla
```

Exchange-Traded Derivatives Markets

Exchange-traded derivatives are created, authorized, and traded on a derivatives exchange, an organized facility for trading derivatives.

- They are standardized instruments with respect to certain terms and conditions of contracts. This specification applies to features like the schedule of expiry dates and contract magnitude. The market participants in the exchange-traded derivatives markets are the market-makers (dealers) and speculators who are typically exchange members. The interplay between market makers and speculators creates a more liquid and more orderly market.

- The standardization also ensures **clearing** (verification of transaction and identities) and **settlement** (transfer of money) of derivatives contracts happens efficiently and allows for the provision of a credit guarantee by the clearinghouse. The clearinghouse can provide this guarantee through the requirement of a cash deposit called a **margin bond** or **performance bond**.

- Exchange-traded markets have **transparency** as full information on the transactions is disclosed to the exchange and regulatory bodies. This does mean a loss of privacy and, coupled with the standardization, a loss of flexibility. As an alternative to standardization, OTC markets provide a substitute for firms wishing to trade non-standardized products.

Over-the-Counter Derivatives Markets

Over-the-counter derivatives are transactions created by any two parties off a derivatives exchange.

- They don't have standardized terms and features. The parties set all of their own terms and conditions, and each party assumes the credit risk of the other party. It can be difficult for a dealer to find a contract that is a perfect match to hedge a position, and they usually have to rely on similar transactions in which they can lay off their risk. The ability to customize OTC contracts does not necessarily make the market less liquid than the standardized exchange-traded contracts. As many of the OTC instruments can be easily created, an offsetting instrument can be created, oftentimes between the same two transacting parties, to terminate the position.

- OTC markets do have a lower level of regulation than exchange-traded markets. However, post the 2007 financial crisis, regulatory oversight has been increasing. On full implementation of new rules, many OTC transactions will have to be cleared through central clearing agencies with information reported to the regulatory authorities.

3. Forward Commitments

```
<p> </p>Based on the rights and obligations of the parties that enter into the contrac
```

A **forward commitment** is an agreement between two parties in which one party agrees to buy and the other agrees to sell an asset at a future date at a price agreed on today. In essence, a forward commitment represents a **commitment** to buy or sell.

There are three types of forward commitments.

Forward Contract

A **forward contract** is an agreement to buy or sell an asset at a specified time in the future for a specified price.

- A forward contract is a forward commitment created in the **over-the-counter** market. It is **not** conditional; both the buyer and the seller are obliged to perform the contract as agreed.
- It is negotiated in the present and will be settled in the future. By contrast, a spot contract is settled immediately.
- The parties to the transaction specify the forward contract's terms and conditions, such as when and where delivery will take place and the precise identity of the underlying. In this sense the contract is said to be customized.
- Each party is subject to the possibility that the other party will default.
- In the financial world, the underlying asset of a forward contract can be a security (e.g., a stock or bond), a foreign currency, a commodity, an interest rate, or combinations thereof.
- The forward market is a private and largely unregulated market.

Futures Contract

A **futures contract** is created and traded on a futures exchange. It is a variation of a forward contract that has essentially the same basic definition but some additional features. Futures and forwards are essentially similar contracts; the principles for pricing and the applications of futures and forwards are almost identical. They differ only in the institutional settings in which they trade.

- Futures contracts always trade on an organized exchange.
- Futures contracts are always highly standardized with specified underlying goods, quantity (contract size), delivery date, trading hours and trading area. Some exchanges may specify that the contract can be only traded in a designated trading area on the floor (called a **pit**). For example, the Chicago Board of Trade (CBOT) establishes the following terms for the U.S. Treasury bond futures contract:
  - The contract is based on a U.S. Treasury bond with a maturity of at least 15 years.
  - The contract covers $100,000 par value of U.S. Treasury bonds.
  - The expiration months are March, June, September, and December.
  - Prices of the contract are quoted in points and 32nds of part of 100. That is, a price of 103 18/32 equals 103.5625. With a contract size of $100,000, the actual price is $103,562.50.
  - The minimum price fluctuation, or **tick** size, is 1/32. With a contract size of $100,000, the actual minimum size is $31.25.

  Anyone who wishes to trade a U.S. Treasury bond futures contract on the CBOT must accept these terms. If a customized contract is desired, a forward contract is the only alternative.

  Standardization of futures contracts promotes liquidity. Since futures contracts are standardized with generally accepted terms, they have an active secondary market where previously created futures contracts are bought and sold.

  - With standardized contracts, all market participants know exactly what is being offered for sale and what the transaction terms are.
  - Thus, people can quickly transact without wasting time examining contracts.
  - In addition, standardization makes it much easier for traders to find buyers and sellers.

  In contrast, forward contracts are customized and therefore usually do not trade after being created.

- Performance on futures contract is guaranteed by a **clearinghouse** : a financial institution associated with the futures exchange that guarantees the financial integrity of the market to all traders.

  - The clearinghouse acts as the intermediary counterparty to the buyer and seller in each trade.
  - The clearinghouse adopts the position of buyer to every seller and seller to every buyer.
  - Every trader in the futures markets has obligations only to the clearinghouse, and the clearinghouse guarantees the fulfillment of the contract of the trading parties.
  - Since the clearinghouse is well-capitalized, its default risk is very small.

  In contrast, there is no clearinghouse in a forward market. Traders in the forward market have direct obligations to each other. However, traders often do not know each other and cannot evaluate each others' credit risks, so they are concerned with the reliability of their counterparties.

- All futures contracts require that traders post margins in order to trade. The futures exchange requires traders to settle the gains/losses of their accounts on a daily basis.

  - This is called **daily settlement** or **marking to market**.
  - Every day, the gain and loss incurred by each trader is computed based on the market price of the futures contracts.
  - After the contracts are marked-to-market, funds are transferred from the traders who have sustained losses to traders who have incurred gains.
  - The practice of daily settlement is equivalent to terminating a futures contract at the end of each day and reopening it the next day at the settlement price.

  Forward contracts, on the other hand, are typically settled at expiration. Until then, no money changes hands between the counterparties.

- Futures markets are regulated by an identifiable government agency, while forward contracts generally trade in an unregulated market. Futures contracts are public transactions. A futures transaction must be reported to:

  - the futures exchanges
  - the clearinghouse
  - at least one regulatory agency

  The price of futures transactions is available to the public through price-reporting services.

- Forward contracts are generally designed to be held until expiration, but a futures market provides sufficient liquidity to permit parties to enter the market and offset transactions previously created.

Swap

A **swap** is an agreement between two parties to exchange a series of future cash flows.

- It is a variation of a forward contract that is essentially equivalent to a series of forward contracts.
- It is custom-tailored to meet the specific needs of counterparties, so counterparties can choose the exact dollar amount and/or maturity that they need.
- Swaps are subject to default, but as the notional principal is not exchanged, the credit risk of a swap is much lower than that of a loan.

- Swaps are private transactions and thus are not traded on exchanges and can avoid regulation to a considerable degree.

The most commonly used swap is a **fixed-for-floating interest rate swap**, also referred to as a â€œplain vanilla swap.â€ The **notional principal** is the loan balance on which the interest rate payments are determined. As with futures and forwards, no money changes hands at the start of the contract, and the swap value is zero. However, as market conditions change, the value of the swap will change, being positive for one party and negative for the other.

4. Contingent Claims: Options

```
    <p> </p>A <b>contingent claim</b> is a derivative contract with a payoff dependent on
```

The primary types of contingent claims are **options**. The payoff of an option is contingent on the occurrence of an event.

Every option is either a **call option** or a **put option**. In essence, options represent the **right**, not commitment, to buy or sell. They are created only by selling and buying. The seller receives payment (the **premium**) for an option from the buyer, and confers rights to the option buyer.

- The premium (**value**) is paid when the option contract is initiated.
- The **price** at which the option holder can buy or sell the underlying is called the **exercise price** or **strike price**.

There are two fundamental kinds of options:

- **American option**. It permits the owner to exercise at any time before or at expiration.
- **European option**. The owner can exercise the option only at expiration.

The American option cannot be worth less than the European option, because the owner of the American option also has the right to exercise the option before expiration if he desires. Put it in another way, you can do with an American option anything you can do with a European option, plus you can exercise early. Thus, the American option gives the owner more flexibility.

Note: The terms "European" and "American" are not associated with geographical locations.

Moneyness refers to the potential profit or loss from the immediate exercise of an option. An option may be:

- **In-the-money** if its exercise would be profitable for its holder. A call (put) option is in-the-money if the stock price exceeds (is below) the exercise price;
- **Out-of-money** if its exercise would be unprofitable for its holder. A call (put) option is out-of-money if the stock price is less (higher) than the exercise price;
- **At-the-money** if the value of the underlying is equal to the exercise price. A call or put option is at-the-money if the stock price equals the exercise price.

Intrinsic value is the value of the option if it is exercised immediately.

Option Payoffs

The easiest time to determine an option's value is at expiration. At that point there is no future; only the present matters. An option's value at expiration is called its payoff.

For a European option at expiration:

- $c_T = Max(0, S_T - X)$
- $p_T = Max(0, X - S_T)$

Long call strategy. The worst that can happen is losing the entire premium (value) of the option. Potential profits are theoretically unlimited.

Short call strategy. The best thing that can happen to the seller of a call is never to hear any more about the transaction after collecting the initial premium. Potential losses from selling a call are theoretically unlimited.

Long put strategy. The smaller the stock price (ST), the greater the put option value.

Short put strategy.

5. Other Derivatives

```
    <h4>Credit Derivatives</h4>
```

A **credit derivative** provides credit protection for the buyer in the event of loss from a credit event.

In a **total return swap**, the underlying is typically a loan or a bond. The credit protection buyer pays the credit protection seller the total return on the bond (interest plus capital) in return for a fixed or floating rate of interest. If the bond defaults, the credit protection seller must continue to pay the interest while receiving no (or very little) return from the buyer.

In a **credit spread option**, the underlying is the yield spread between the yield on a bond and the yield of a benchmark default-free bond. This yield spread, or credit spread, is a reflection of investorsâ€™ perception of credit risk. The credit protection buyer selects the strike spread and pays an option premium to the seller. At expiration, the spread is compared with the strike spread, and if the option is in-the-money, the seller pays the buyer the determined payoff.

In a **credit-linked note**, the credit protection buyer usually holds a bond that may be subject to default and, to offset that risk, issues a credit-linked note with the condition that if the bond defaults, the principal payoff is reduced accordingly. Thus, the buyer of the credit-linked note takes on the credit risk of the underlying bond.

In a credit default swap, the credit protection buyer makes a series of regularly scheduled payments to the credit protection seller. The seller makes no payments until a credit event occurs. A credit event could be a declaration of bankruptcy, a failure to make a scheduled payment, or a restructuring. The CDS contract will explicitly define what constitutes a credit event. A CDS is essentially a form of insurance and provides loss coverage in return for the premium paid by the buyer to the seller.

Asset-backed Securities

Asset-backed securities are securities that are collateralized by a pool of securities such as mortgages, loans or bonds. Typically borrowers of mortgages, loans or bonds have the prepayment option to pay off their debts early.

When a mortgage asset portfolio is assembled into an ABS, the resulting instrument is called a collateralized mortgage obligation (CMO). When homeowners pay off their mortgages early (prepayment), the mortgage holders suffer, and an expected stream of returns has been terminated early. The funds now have to be reinvested at a typically lower rate. CMOs typically partition the mortgages into A, B, and C classes, and class C will bear the first wave of prepayment risk, followed by class B and class A. As the risk on the tranches is not equal (class C bears the most prepayment risk), the expected returns on the classes vary to compensate investors for the varying risk.

A **collateralized loan obligations (CDO)** does not have much prepayment risk but does have credit risk. A CDO allocates this risk to different tranches, senior, mezzanine, or junior tranches. When a default occurs, the junior tranche bears the risk first, followed by the mezzanine and then the senior tranche. Therefore, senior tranches have the lowest risk but also the lowest expected return.

6. Purposes and Criticisms of Derivative Markets

```
<p> </p>Some of the main benefits that financial derivatives bring to the market are:
```

- **Risk allocation, transfer, and management.** This refers to the process of identifying the desired level of risk, measuring the actual level of risk, and taking actions to bring the actual level of risk to the desired level of risk. Financial derivatives provide a powerful tool for limiting risks that individuals and firms face in the ordinary conduct of their business. For speculators risks associated with financial derivatives are not necessarily evil because they provide very powerful instruments for knowledgeable traders to expose themselves to calculated and well-understood risks in pursuit of profit.

- **Price discovery.** Futures, forwards and swaps provide valuable information about the prices of the underlying assets. Options provide information on the price volatility of the underlying assets.

- **Trading efficiency.** As the derivative markets are highly liquid, financial derivatives can be bought or sold with less transaction costs than directly trading the underlying assets. In addition, derivatives are designed to facilitate risk management, and serve as a form of insurance. The cost of insurance must be low relative to the value of the insured assets. Otherwise insurance would not exist.

- **Market efficiency.** Derivatives provide an alternative for investing in the underlying assets. If the prices of the underlying assets are too high, investors will invest in derivatives, thereby reducing the demand for the underlying assets. As a result, the derivatives market will force the prices of the underlying assets back to their appropriate levels.

- **Market completeness.** A **complete market** is a market in which any and all identifiable payoffs can be obtained by trading the securities available in the market. The financial derivatives help traders to more exactly shape the risk and return characteristics of their portfolios, thereby increasing the welfare of traders and the economy as a whole.

The complexity of derivatives means that sometimes the parties that use them don't understand them well. As a result, they are often used improperly, leading to potentially large losses. This can explain why unknowledgeable investors tend to consider derivatives excessively dangerous. Derivatives are also mistakenly characterized as a form of legalized gambling. This view tends to overlook the benefits of derivatives (e.g., risk management). In fact, derivatives make financial market work better, not worse.

7. Arbitrage

```
<b>Arbitrage</b> is a process through which an investor can buy an asset or combinatic
```

In a well-functioning market, arbitrage opportunities should not exist. If they do exist, arbitrage activities would quickly eliminate the price differential. The **no-arbitrage principle** states that any rational price for a financial instrument must exclude arbitrage opportunities. This is the minimal requirement for a feasible or rational price for any financial instrument. There is no free money.

The role of arbitrage:

- <u>It facilitates the determination of prices.</u> The combined actions of many investors engaging in arbitrage result in rapid price adjustments that eliminate any arbitrage opportunities, thereby bringing prices back.

- <u>It promotes market efficiency.</u> Efficient markets are those in which it is impossible to earn abnormal returns, which are returns that are in excess of the return required for the risk assumed. Arbitrage activities will quickly eliminate arbitrage opportunities available in the market, thereby promoting market efficiency.

Hedgers vs. Speculators: two parties involved in the risk management process

Depending on their prior risk exposures, participants in the derivatives market can be classified into hedgers and speculators.

- A **hedger** trades futures to *reduce* some pre-existing risk exposure.
    - Prior to the transaction, the hedger *does* have risk exposure.
    - After the transaction, the hedger reduces risk exposure.
    - At the time of entering into hedging transactions, the hedger knows the benefit (reduced risk).
    - Hedgers are often producers or users of a given commodity.
- A **speculator** takes a view of the market, and *accepts* the market's risk in pursuit of profit.
    - Prior to the transaction, the speculator has no risk exposure.
    - After the transaction, the speculator has increased risk exposure.
    - The profits/losses of a speculative transaction are not known immediately.

Basics of Derivative Pricing and Valuation

1. The Principle of Arbitrage

```
   Arbitrage means taking advantage of price differences in different markets. In well-fu
```

Arbitrage and Derivatives

Assume the risk-free rate is 5%. The current price of gold is $300 per ounce and the forward price of gold is $330 in one year's time. Is there an arbitrage opportunity?

Here is what you can do:

- Borrow $300 at 5% today.
- Buy one ounce of gold (price $300).
- Enter into a short forward to sell one ounce of gold for $330 in one year's time.
- After one year you sell the gold for $330, and repay the bank $300 plus $15 interest.

Hence, a profit of $15 can be made without any risk!

In fact, any delivery price above $315 will result in a risk-free profit using this strategy.

What if the delivery price is $310?

- Sell one ounce of gold for $300.
- Deposit the $300 in the bank at 5% interest.
- Enter into a forward to buy one ounce of gold in one year's time for the delivery price ($310).
- After one year, buy one ounce of gold for $310 and keep the $5 profit.

Again, a profit of $5 can be made without any risk.

Investors in the gold market will take advantage of any forward price that is not equal to $315, eventually bring the price to $315, which is known as the **arbitrage-free price**.

The arbitrage principle is the essence of derivative pricing models.

Arbitrage and Replication

A portfolio composed of the underlying asset and the riskless asset could be constructed to have exactly the same cash flows as a derivative. This portfolio is called the replicating portfolio. Since they have the same cash flows, they would have to sell at the same price (the law of one price).

Assume the forward price of gold is $315 in one year's time, and the spot price is $300. You have $300.

- You can deposit $300 in the bank at 5% interest. One year later you will get $315.
- You can also buy one ounce of gold, and a forward contract to sell it in one year for $315. One year later you will also get $315.

Why replicate?

- To explore pricing differentials
- Lower transaction costs

Replication is the essence of arbitrage.

Risk Aversion, Risk Neutrality, and Arbitrage-Free Pricing

Risk-seeking investors give away a risk premium because they enjoy taking risk. Risk-averse investors expect a risk premium to compensate for the risk. Risk-neutral investors neither give nor receive a risk premium because they have no feelings about risk.

Risk-neutral pricing: Suppose you want to price a derivative. The payoff of this derivate can be replicated using the underlying asset and risk-free rate. The market price of this derivative and the replicating strategy must be exactly the same under the principle of no arbitrage, regardless of risk preferences.

To obtain the derivative price we should assume the investor is risk-neutral, because an investor's risk aversion is not a factor in determining the derivative price. Risk can be eliminated by dynamic hedging in a situation where there is no arbitrage possible. Once risk is eliminated in this way the expected return becomes equal to the risk-free rate for all investors. Assets can be assumed to grow at the risk-free rate and also discounted at the risk-free rate.

2. The Concept of Pricing vs. Valuation

```
   The value of a forward, futures and swap contract is zero at initiation date. Its pric
```

Example

Two parties agree to a forward contract to deliver a zero-coupon bond at a price of $97 per $100 par in 3 month.

At initiation date:

- Value: 0
- Price: $97

At the contract's expiration, suppose the underlying zero-coupon bond is selling at a price of $97.25. The long is due to receive from the short an asset worth $97.25, for which a payment to the short of $97 is required.

- Value: $0.25
- Price: $97

3. Pricing and Valuation of Forward Contracts

```
   <b>Pricing and Valuation at Expiration</b><p> </p>
```

At expiration T, the value of a forward contract to the long position is:

$$V_T(T) = S_T - F_0(T)$$

where S_T is the spot price of the underlying at T and $F_0(T)$ is the forward price.

The forward price is the price that a long will pay the short at expiration and expect the short to deliver the asset.

Pricing and Valuation at Initiation Date

There is no cash exchange at the beginning of the contract and hence the value of the contract at initiation is zero.

$$V_0(T) = 0$$

The forward price at initiation is:

$$F_0(T) = S_0(1 + r)^T$$

Example

Consider a forward contract on a non-dividend paying stock that matures in 6 months. The current stock price is $50 and the 6-month interest rate is 4% per annum. Compute the forward price, F.

Solution: Assuming semi-annual compounding, F = 50 x 1.02 = 51.0.

If we add benefits Ê‡ (dividends, interest, and convenience yield), and costs Î¸ the forward price of an asset at initiation becomes

$$F_0(T) = S_0(1 + r)^T - (\text{Ê‡} - \text{Î¸}) (1 + r)^T$$

Consider a forward contract on a 4-year bond with 1 year maturity. The current value of the bond is $1018.86. It has a face value of $1000 and a coupon rate of 10% per annum. A coupon has just been paid on the bond and further coupons will be paid after 6 months and after 1 year, just prior to delivery. Interest rates for 1 year out are flat at 8%. Compute the forward price of the bond.

$F = 1018.86 \text{ x } 1.04^2 - 50 \text{ x } 1.04 - 50 = \$1,000$

Pricing and Valuation during the Life of the Contract

The value of a forward contract after initiation and during the term of the contract change as the price of the underlying asset (S) changes. The value (profit/loss) of a forward contract between initiation and expiration is the current price of the asset less the present value of the forward price (at expiration).

4. Forward Rate Agreements

```
  A <b>forward rate agreement</b> (<b>FRA</b>) is a forward contract in which one party,
```

- The long pays fixed rate and receives floating rate. If Libor rises the long will gain.
- The short pays floating rate and receives fixed rate. If Libor falls the short will gain.

The fixed rate is also called the **forward contract rate**. The interest rate to be determined at expiration is also called the **underlying rate**.

The buyer effectively has agreed to borrow an amount of money in the future at the stated forward (contract) rate. The seller has effectively locked in a lending rate. The buyer of a FRA profits from an increase in interest rates. The seller of a FRA profits from a decline in rates.

Example

Shell and Barclays enters into the following FRA:

- Shell, the end user, takes a long position in a FRA that expires in 30 days and is based on 60-day LIBOR.
- Barclays, a dealer, quotes a rate of 5.65% for this FRA.
- The notional principal of this FRA is $1,000,000.

By convention, this FRA is also referred to as a 1 x 3. At the expiration of the FRA in 30 days:

- Shell pays a fixed rate of 5.65% immediately.
- Barclays promises to pay a rate of 60-day LIBOR determined at expiration. Suppose that the 60-day LIBOR at expiration is 6%. Barclays will pay 6% of interest to Shell 60 days after the contract expiration date. In effect, the 6% interest is paid 90 days (30 + 60) from the contract initiation date.

Note the market convention quotes the time periods as months, but the calculations use days based on the assumptions of 30 days in a months. For example, a "1 x 3 FRA" expires in 30 days, and the payoff of the FRA is determined by 60-day Libor when the FRA expires in 30 days.

5. Why do Forward and Futures Prices Differ?

```
   In assigning a forward price, we set the price such that the value of the contract is
```

Unlike forward contract prices, however, futures prices fluctuate in an open and competitive market. The marking-to-market process results in each futures contract being terminated every day and reinitiated.

If we ignore the credit risk issue (futures contracts are essentially free of default risk as they are settled daily but forward contracts are subject to default risk), we should conclude that:

- The price of a futures contract will equal the price of an otherwise equivalent forward contract *if interest rates are known or constant*. Under this condition, any effect of the addition or subtraction of funds from the marking-to-market process can be shown to be neutral.
- The price of a futures contract will equal the price of an otherwise equivalent forward contract *if interest rates are uncorrelated with future prices.*
- *If interest rates are positively correlated with future prices,* futures will carry higher prices than forwards.
    - Traders with long positions will prefer futures over forwards, because futures will generate gains when interest rates are going up (and thus future prices are going up as they are positively correlated), and traders can invest these gains for higher returns.
    - Traders will incur losses when interest rates are going down and can borrow to cover those losses at lower rates.
    - Gold futures are good examples in this case, as gold futures prices and interest rates would tend to be positively correlated.
- If futures prices are negatively correlated with interest rates, traders will prefer not to mark to market, so forward contracts will carry higher prices. Interest rate futures are good examples in this case: interest rate and fixed-income security price move in opposite directions.

6. Pricing and Valuation of Swap Contracts

```
   <b>Swaps</b> are derivative securities in the form of agreements between two counterpa
```

Example

Party A agrees to pay a fixed rate of interest on $10 million each year for 3 years to Party B. In return, Party B agrees to pay a floating rate of interest on $10 million each year for 3 years to Party A.

A swap involves a series of payments over its tenor, and *can be considered a series of forward contracts.* In contrast, forwards, futures and options only involve a single payment or two payments (i.e., when the option is purchased and when it is exercised).

In general, neither party pays any money to the other at the initiation of a swap. *A swap has zero value at the start.*

A swap can be viewed as combining a series of forward contracts into a single transaction. However, there are some small differences. For example, swaps are a series of equal fixed payments, whereas the component contracts of a series of forward contracts would almost always be priced at different fixed rates. In this context we often refer to a swap as a series of **off-market forward contracts**, reflecting the fact that the implicit forward contracts that make up the swap are all priced at the swap fixed rate and not at the rate at which they would normally be priced in the market.

7. The Value of a European Option at Expiration

```
   Almost anything with a random outcome can have an option on it. The underlying instrum
```

Every option is either a **call option** or a **put option**. Options are created only by selling and buying. Therefore, for every owner (buyer) of an option, there is a seller (writer).

- The premium (**value**) is paid when the option contract is initiated.
- The **price** at which the option holder can buy or sell the underlying is called the exercise price or strike price.

There are two fundamental kinds of options:

- **American option**. It permits the owner to exercise at any time before or at expiration.
- **European option**. The owner can exercise the option only at expiration.

The American option cannot be worth less than the European option, because the owner of the American option also has the right to exercise the option before expiration if he desires. Put it in another way, you can do with an American option anything you can do with a European option, plus you can exercise early. Thus, the American option gives the owner more flexibility.

Note: The terms "European" and "American" are not associated with geographical locations.

Moneyness refers to the potential profit or loss from the immediate exercise of an option. An option may be:

- **In-the-money** if its exercise would be profitable for its holder. A call (put) option is in-the-money if the stock price exceeds (is below) the exercise price;
- **Out-of-money** if its exercise would be unprofitable for its holder. A call (put) option is out-of-money if the stock price is less (higher) than the exercise price;
- **At-the-money** if the value of the underlying is equal to the exercise price. A call or put option is at-the-money if the stock price equals the exercise price.

Intrinsic value is the value of the option if it is exercised immediately.

Option Payoffs

The easiest time to determine an option's value is at expiration. At that point there is no future; only the present matters. An option's value at expiration is called its payoff.

For a European option at expiration:

- $c_T = \text{Max}(0, S_T - X)$
- $p_T = \text{Max}(0, X - S_T)$

Long call strategy. The worst that can happen is losing the entire premium (value) of the option. Potential profits are theoretically unlimited.

Short call strategy. The best thing that can happen to the seller of a call is never to hear any more about the transaction after collecting the initial premium. Potential losses from selling a call are theoretically unlimited.

Long put strategy. The smaller the stock price (ST), the greater the put option value.

Short put strategy.

8. Factors that Affect the Value of an Option

```
   The previous discussion tells us that the price is somewhere between zero and maximum,
```

For American options, which are exercisable immediately:

C_0 >= Max (0, S_0 - X)

P_0 >= Max (0, X - S_0)

If the option is in-the-money and is selling for less than its intrinsic value, it can be bought and exercised to net an immediate risk-free profit.

However, European options cannot be exercised early; thus, there is no way for market participants to exercise an option selling for too little with respect to its intrinsic value. Investors have to determine the lower bound of a European call by constructing a portfolio consisting of a long call and risk-free bond and a short position in the underlying asset.

First the investor needs the ability to buy and sell a risk-free bond with a face value equal to the exercise price and current value equal to the present value of the exercise price. The investor buys the European call and the risk-free bond and sells short (borrows the asset and sells it) the underlying asset. At expiration the investor shall buy back the asset.

This combination produces a non-negative value at expiration, so its current value must be non-negative. For this situation to occur, the call price has to be worth at least the underlying price minus the present value of the exercise price:

The lower bound of a European put is established by constructing a portfolio consisting of a long put, a long position in the underlying, and the issuance of a zero-coupon bond. This combination produces a non-negative value at expiration so its current value must be non-negative. For this situation to occur, the put price has to be at least as much as the present value of the exercise price minus the underlying price.

For both calls and puts, if this lower bound is negative, we invoke the rule that an option price can be no lower than zero.

Example

- All options expire in 60 days, have the same exercise price (X) of $60 and the same underlying asset.
- The current price of the underlying (S_0) is $50.
- The risk-free rate (r) is 5%.
- Find the lower bounds of American and European calls and puts.

Solution

- Time to expiration (T) = 60/365 = 0.1644
- European Call (c_0): MAX[0, 50 - 60/(1 + 5%)$^{0.1644}$] = MAX[0, -9.52] = 0
- American Call (C_0): MAX[0, 50 - 60/(1 + 5%)$^{0.1644}$] = MAX[0, -9.52] = 0
- European Put (p_0): MAX[0, 60/(1 + 5%)$^{0.1644}$ - 50] = MAX[0, 9.52) = 9.52
- American Put (P_0): MAX[0, 60 - 50) = 10
- Note that the lower bound of the American put is above the lower bound of the European put.

9. Put-Call Parity

```
First, consider an option strategy referred to as a <b>fiduciary call</b>, which consi
```

- If the price of the underlying is below X at expiration, the call expires worthless and the bond is worth X.
- If the price of the underlying is above X at expiration, the call expires and is worth S_T (the underlying price) - X.

At expiration the fiduciary call will end with X or S_T, whichever is greater. This combination allows protection against downside losses and is thus faithful to the notion of preserving capital.

Then, consider an option strategy known as a **protective put**, which consists of a European put and the underlying asset.

- If the price of the underlying is below X at expiration, the put expires and is worth X - S_T and the underlying is worth S_T.
- If the price of the underlying is above X at expiration, the put expires with no value and the underlying is worth S_T.

So, at expiration, the protective put is worth X or S_T, whichever is greater.

Thus, the fiduciary call and protective put end up with the same value. They are therefore identical combinations. To avoid arbitrage, their values today must be the same.

This equation is called **put-call parity**. It does not say that the puts and calls are equivalent, but it does show an equivalence (parity) of a call/bond portfolio and a put/underlying portfolio. Note that the put and call must have the same underlying, exercise price and expiration date.

By re-arranging the above equation:

Because the right side of this equation is equivalent to a call, it is often referred to as a **synthetic call**. It consists of a long put, a long position in the underlying, and a short position in the risk-free bond.

There are numerous other combinations that can be constructed. For example, the put can be isolated as:

The right side is a **synthetic put**, which consists of a long call, a short position in the underlying, and a long position in the risk-free bond.

Another example is

Synthetic positions enable investors to price options, because they produce the same results as options and have known prices. Consider the following example: a European call with an exercise price of $30 expires in 90 days. A European put with the same exercise price, expiration date and underlying is selling for $6. The underlying is selling for $40, and the risk-free rate is 10%. Based on the information, you can compute the value of the synthetic call: $c_0 = p_0 + S_0 - X/(1 + r)^T$ = 6 + 40 - 30/(1 + 10%)$^{90/365}$ = $16.7 (Note the time to expiration T = 90/365 = 0.2466). Since the synthetic call and the actual call have the same payoff, they must have the same price as well. Therefore, the price of the call should be $16.70.

Synthetic positions also tell how to exploit mispricing of options relative to their underlying assets. Continue with the above example. If the market price of the call is $10, the call is then underpriced. You can make a risk-free profit of $6.70 by selling the synthetic call for $16.7.

You cannot only synthesize a call or a put, but also synthesize the underlying or the bond.

Violations of put-call parity for European options:

Recall that the basic put-call parity equation is: $c_0 + X/(1 + r)^T$ (fiduciary call)= $p_0 + S_0$ (protective put). Violations of put-call parity occur when one side of the equation is not equal to the other.

- An arbitrageur can buy the lower-priced side and simultaneously sell the higher-priced side, thereby making a profit on the price difference. Since the fiduciary call and protective put have the same payoff, the arbitrageur's positions will perfectly offset at expiration.
- As more and more arbitrageurs perform these transactions, the price of the lower-priced portfolio will increase and the price of the higher-priced portfolio will decrease, until put-call parity is restored.

Example

Consider the following example involving call options with an exercise price of $100 expiring in half a year (T = 0.5). The risk-free rate is 10%. The call is priced at $7.5, and the put is priced at $4.25. The underlying price is $99.

- *Analysis:* The left side of the put-call parity equation is $c_0 + X / (1 + r)^T = 7.5 + 100/(1.10)^{0.5} = 102.85$. The right side is $p_0 + S_0 = 4.25 + 99 = 103.25$. This means the protective put is overpriced.

- *Our strategy:* You sell the protective put. This means you sell the put and sell short the underlying. Doing so will generate a cash inflow of $103.25. You buy fiduciary call, paying out $102,85, netting a cash inflow of $0.4. At expiration, if the price of the underlying is above 100:

    - The bond matures, paying $100.
    - Use the $100 to exercise call, receiving the underlying.
    - Deliver the underlying to cover the short sale.
    - The put expires with no value.
    - Net effect: no money in or out.

- What would you do if the price of the underlying is below 100 at expiration?

- So, $0.4 is received up front and nothing has to be paid out. The position is perfectly hedged and represents an arbitrage profit. The combined effects of other investors performing this transaction will result in the value of the protective put going down and/or the value of the covered call going up until the two strategies are equivalent in value.

10. Put-Call-Forward Parity

```
Assume that:<p> </p><ul class="notes">
```

- F(0, T): the price established today for a forward contract expiring at time T
- c_0: the call option price today
- p_0: the put option price today
- Both options expire when the forward contract expires: the time until expiration is also T.
- The exercise price of both options is X.

Consider two portfolios. Portfolio A consists of a long call and a long position in a zero-coupon bond with face value of X - F(0, T). Portfolio B consists of a long put and a long forward.

As the two portfolios have exactly the same payoff, their initial investments should be the same as well. That is:

This equation is **put-call parity for options on forward contracts**.

As F(0, T) = $S_0(1 + r)^T$, we rearrange the equation as the follows:

Consider the following example:

T = 90 days, r = 5%, X = \$95, S_0 = \$100, and the call price is \$10. The put price should be $c_0 + X/(1 + r)^T - S_0 = 10 + 95/(1 + 0.05)^{(90/365)} - 100 = \$3.86$.

Similarly, we can compute the call price given the price of the put.

Consider another example. The options and a forward contract expire in 50 days. The risk-free rate is 6%, and the exercise price is 90. The forward price is 92, and the call price is 5.5.

$p_0 = c_0 + [X - F(0, T)]/(1 + r)^T = 5.5 + (90 - 92)/1.06^{(50/365)} = 3.52$

Note that in this case X < F(0, T), which means that we short the bond instead of buying the bond as in portfolio A above.

Continue with those assumptions at the beginning of this subject. Consider a portfolio consisting of a long call, short put and a long position in a zero-coupon bond with face value of X - F(0, T). At expiration the value of the portfolio is:

- 0 (value of long call) + [-(X - S_T)] (value of short put) + [X - F(0, T)] (value of long bond) = S_T - F(0, T), if $S_T <= X$.
- [S_T - X] (value of long call) + 0 (value of short put) + [X - F(0, T)] (value of long bond) = S_T - F(0, T), if $S_T > X$.

As a forward contract's payoff at expiration is also S_T - F(0, T), the portfolio's initial value must be equal to the initial value of the forward contract (which is 0).

Solving for F(0, T), we obtain the equation for the forward price in terms of the call, put, and bond. Therefore, a synthetic forward contract is a combination of a long call, a short put and a zero-coupon bond with face value (X - F(0, T)). Note that we may either long or short this bond, depending on whether the exercise price of these options is lower or higher than the forward price.

11. Binomial Valuation of Options

```
    In finance, the binomial options model provides a generalisable numerical method for t
```

The binomial pricing model uses a "**discrete-time framework**" to trace the evolution of the option's key underlying variable via a binomial lattice (tree), for a given number of time steps between valuation date and option expiration. Each node in the lattice represents a *possible* price of the underlying at a *particular* point in time. This price evolution forms the basis for the option valuation. The valuation process is iterative, starting at each final node, and then working backwards through the tree to the first node (valuation date), where the calculated result is the value of the option.

Option valuation using this method is, as described, a three step process:

- price tree generation
- calculation of option value at each final node
- progressive calculation of option value at each earlier node; the value at the first node is the value of the option

We start off by having one binomial period for a European call option.

Define:

- r = risk-free rate
- c^+ = Max (0, S^+ - X) (call price if the stock price goes up: "up state")
- c^- = Max (0, S^- - X) (call price if the stock price goes down: "down-state")
- u = (S^+ / S_0) ("up state" price relative)
- d = (S^- / S_0) ("down-state" price relative)
- S_T = Stock price at time (T)
- Ï€ = Greek small letter pi.

Formulas:

- Ï€ = (1 + r - d) / (u - d) (risk-neutral "up" probability)
- c_0 = [Ï€ c^+ + (1 - Ï€) c^-] / (1 + r) (the price of the call option)
- n = (c^+ - c^-) / (S^+ - S^-) (the hedge ratio: the number of shares of stock per option to hedge)

We assume that the stock price will only take two possible values at the expiration date of the option. In our example:

- Current stock price = S_0 = \$80
- Stock price at expiration = \$90 (S^+) or \$75 (S^-)
- Exercise price of call option = \$85 (X)
- Time to expiration = T = 1/2 year (6 months)
- Risk-free rate of return = r = 6% (discrete and annual)

Step 1: Diagram Stock Price Dynamics and Option Values on Trees

Based on this information, tree diagrams for the stock value and call option payoffs (state dependent) would be drawn as follows:

Step 2: Compute Risk Neutral Probabilities of Up and Down States

Î¼ = (90/80) = 1.125

d = (75/80) = 0.9375

Ï€ = [$(1.06)^{0.5}$ - 0.9375] / (1.125 - 0.9375) = 0.4912

Step 3: Compute Expected Value of Call Option

c_0 = [0.4912 \$5 + (1 - 0.4912) \$0] / $(1.06)^{0.5}$ = \$2.385

Therefore, today's value of the 1-period option is \$2.385.

Alternatively, we can use combination of Stocks and Calls to create state-independent payoffs and then determine the no-arbitrage value of the option rights.

Step 1: Calculate the hedge ratio (shares per call) n = (\$5 - \$0) / (\$90 - \$75) = 0.3333

<u>Step 2: Use the hedge ratio to construct a portfolio of stocks and calls in which terminal payoff is state-independent.</u> Let TV denote the terminal value of the portfolio at expiration.

- If stock price = S^-, $TV^- = n\ S^- - c^-$
- If stock price = S^+, $TV^+ = n\ S^+ - c^+$
- $TV^- = TV^+$

Regardless of which way the underlying moves, the portfolio value should be the same (*perfectly hedged*).

Continuing from the previous example:

To form a perfectly hedged portfolio, an investor needs to buy 1/3 of a share of stock for each call that is written (sold), or buy 1 share of stock and sell (write) 3 calls. To see that this is true, consider the position of the portfolio at the expiration of the call if the investor writes 1 call.

Note that each call is worth \$5 to the buyer or holder, so the seller of the call is in a negative position.

<u>Step 3: Value the Option</u>

Guaranteed outcome is \$25 for this portfolio, regardless of the value of the stock at expiration. How much should you pay for this risk-free position?

The present value of the guaranteed \$25 to be received in six months is: \$25 / $1.06^{0.5}$ = \$24.28. To guarantee the \$25 outcome, the investor would have to buy 1/3 share of the stock and sell 1 call option.

\$24.28 = n S_0 - c_0 = initial investment â‡’ c_0 = \$80/3 - \$24.28 = \$2.38

This is the same value we ended up with using the direct approach.

Suppose the option of the previous example is selling for \$3 - a clear case of price not equaling value. Investors would exploit this opportunity by selling the option and buying the underlying. The number of units of the underlying purchased for each option sold would be the hedge ratio: $n = (c^+ - c^-) / (S^+ - S^-) = 0.3333$. Suppose we sell 300 calls and buy 100 shares. The initial outlay would be 100 x \$80 - 300 x \$3 = \$7100. Six months later, the portfolio value will be:

- 100 x \$75 - 0 = \$7500, if stock price = \$75.
- 100 x \$90 - 300 x \$5 = \$7500, if stock price = \$90. Note that the 300 here is the number of options bought, not the hedge ratio.

Our six-month return is 7500/7100 - 1 = 5.63%, and the annualized return is $(1.0563)^2 - 1$ = 11.58%. This risk-free return is much higher than the actual risk-free return of 6%.

If the option sells for less than \$2.38, an investor would buy the option and sell short the underlying, which would generate cash upfront. At expiration, the investor would have to pay back an amount less than 7%. All investors would perform this transaction, generating a demand for the option that would push its price back to \$2.38.

Therefore, when the option is trading at the price given by the model, a hedge portfolio would earn the risk-free rate.

If the option is a put, please note the following differences:

- Hedge ratio $n = (p^- - p^+) / (S^+ - S^-)$
- A risk-free hedge has the same positions in the two instruments (underlying and the put).

12. American Option Pricing

```
   American options can be exercised early. Their prices must always be <i>no less than</
```

American call options:

$$C_0 = Max[0, S_0\text{-}X/(1+r)T]$$

American call options, however, are never exercised early *unless* there is a cash flow on the underlying. The extra value of an American option, if any, comes from the fact that it can be exercised immediately.

American put options:

$$P_0 = max[0, X - S_0]$$

American put options nearly always have a possibility of early exercise, so they ordinarily sell for more than their European counterparts.

Alternative Investments

Alternative Investments

Introduction to Alternative Investments

1. Introduction

```
   Stocks, bonds and cash are the most commonly known traditional investments. Alternativ
```

Some of the distinctive characteristics of alternative investments compared with traditional investments:

- **Lower liquidity** due to their lack of standard markets and limited activities on both sides of the deal.
- **Less regulation** but rather unique legal and tax considerations.
- **Lower transparency** - certain alternative investments lack an efficient market mechanism and may subject their valuation to speculations, creating uncertainties. Risks of alternative investments increase due to absent of ready valuation information.
- **Higher fees** - costs of purchase and sale may be relatively high.
- **Limited and potentially problematic historical risk and return data.**Investors must be careful in evaluating the historical record of alternative investments as the higher than normal returns may be subject to a variety of biases, and the volatility of returns tend to be underestimated.

There are two basic investment strategies.

Passive managers "buy-and-hold." There are very limited ongoing buying and selling actions. Their portfolios are expected to generate *Beta* return.

Most alternative investment managers use **active**, *alpha*-seeking strategies. The assumption is inefficiencies exist that can be exploited to earn positive return after adjusting for beta risk. These active strategies include absolute return, market segmentation and concentrated portfolios.

Sharpe ratios and many downside risk measures are commonly used to measure risk and return of alternative investments.

Despite unique risks and considerations, alternative investments can be useful tools to improve the risk-return characteristics of an investment portfolio. They can increase diversification and reduce volatility given low correlations to more traditional investments.

Many alternative investments use a partnership structure.

- The general partner (the fund) manages the business, assumes unlimited liability, and receives a management fee and an incentive fee.
- Limited partners own fractional interest in the partnership.

2. Investment Methods

```
<p> </p>Investors can access alternative investments in three ways.
```

Fund Investing

The investor contributes capital to a fund which makes investments on the investor's behalf.

Advantages:

- access to fund manager's services and expertise;
- passive management;
- diversification;
- less capital required.

Disadvantages:

- need to pay for funds' services (fees);
- too many funds to select from: due diligence required.

Co-Investing

The investor invests in asset indirectly through the fund, but also possess rights to invest directly in the same assets.

Advantages:

- can learn from fund's process for better direct investing;
- reduced fees;
- more active management of the portfolio and deeper relationship with the manager.

Disadvantages: a bit of in between fund investing and direct investing.

Direct Investing

The investor makes a direct investment in an asset without using an intermediary.

Advantages:

- cost efficient (no management fees to pay);
- great flexibility;
- highest level of control over how the asset is managed.

Disadvantages:

- requires management expertise;
- lack of diversification;
- less access to a fund manager's sourcing network;
- requires greater levels of due diligence due to the absence of a fund manager;
- higher capital requirements.

Investors conduct due diligence prior to investing in alternative investments. The due diligence approach depends on the investment method.

3. Investment and Compensation Structures

```
<h4>Partnership Structures</h4>
```

A **limited partnership (LP)** is a partnership made up of two or more partners. It is often used as an investment vehicle in alternative investments such as hedge funds and real estates.

- The **general partner (GL)** oversees and runs the business. It has unlimited liability and has full management control of the business.
- The **limited partners (LPs)** do not partake in managing the business. They have limited liability up to the amount of their investment. This is the key advantage to LPs - their personal liability is limited.

A LP offers the ability to raise capital without giving up control. Investments in LPs are less regulated (e.g. less reporting, less formal structure) than offerings to the general public.

Compensation Structure

The general partner is typically compensated with an asset management fee and an incentive fee.

The **asset management fee** is usually stated as a percent per annum of assets under management. For private equity funds this is typically as a percent of *committed* capital, or *invested* capital.

The **incentive fee** is typically stated as a percent of profits. Profits are normally defined as the gross profits of the investments after returning partnership expenses including the asset management fee.

Usually, a **hurdle rate** of return must be achieved before sharing begins. So, a typical deal might be stated as "20% carry over an 8% pref with a catchup". This means that the partnership has to earn at least 8% return before the GP earns any carry. Above an 8% return, the GP gets the profit (i.e. the **catchup**) until the ratio of profit split is 20% to GP. Thereafter, the profits are split 80% to the LPs and 20% to the GP. This is the nearly universal structure of private equity sponsor compensation.

Common Investment Clauses, Provisions, and Contingencies

Catch-Up: The LPs receives 100% of the profits until their preferred return hurdle is reached. Above the hurdle, the GP receives 100% of the profits until he is "caught up" to his performance fee.

High-water mark is the highest level of value reached by a fund. It is often used as a threshold to determine whether a fund manager can gain a performance fee. Investors benefit from a high-water mark by avoiding paying performance-based bonuses for poor performance or for the same performance twice.

An investor who buys into a fund at a net asset value (NAV) below the high-water mark will enjoy the upside from the subscription NAV to the high-water mark without paying a fee. This situation is known as a "free ride." It allows new investors to benefit from buying into an under-performing fund without penalizing existing investors.

High-water mark vs. Hurdle rate:

- Under the high-water mark clause, the performance fee of the current term can be impacted by the previous performance of the fund.
- Under the hurdle rate, the current performance bonus is independent of the fund's historical return.

How does a partnership allocate investment returns? A **waterfall** defines the pecking order in which distributions are allocated to limited and general partners.

Usually, the GP receives a disproportionately larger share of the total profits relative to GP's initial investment once the allocation process is complete. This is done to incentivize the fund's GP to maximize profitability for its investors.

There are two common types of waterfall structures:

- **deal-by-deal (American)** favors the GP. GP gets paid before LPs get all their invested capital and preferred return.
- **whole-of-fund (European)** is more LPs friendly. The distribution is applied at the fund level. LPs get their capital and preferred return before the GP gets any profits.

In private equity, **clawback** refers to the LPs' right to reclaim part of the GP's carried interest, in cases where subsequent losses mean the GP received excess compensation. Clawbacks are typically calculated when a fund is liquidated.

4. Hedge funds

```
<p> </p>To "hedge", according to Webster's dictionary, is "a means of protection or de
```

A **hedge fund** is a private "pool" of capital for accredited investors only and organized using the limited partnership legal structure. The general partner is usually the money manager and is likely to have a very high percentage of his/her own net worth invested in the fund.

Characteristics of Hedge Funds

The fund has an offering memorandum, which is intended to provide much of the necessary information to support an investor's due diligence. Among several topics, the offering memorandum will specify the trading style, hedging strategies, and instruments to be employed by the fund at the discretion of the general partner (e.g., being long and /or short stock; use of puts, calls, and futures; use of OTC derivatives).

A **fund of funds** invests in a portfolio of hedge funds to provide access, diversification, risk management and due diligence benefits to investors. Such funds of funds generally charge a fee for their services. Recently funds of funds have been criticized for the significant incremental costs they impose.

Although some hedge funds don't use leverage at all, most of them do. Leverage in hedge funds often runs from 2:1 to 10:1, depending on the type of assets held and strategies used. High leverage is often part of the trading strategy and is an essential part of some strategies in which the arbitrage return is so small that leverage is needed to amplify the profit. As in any other investments, however, leverage also amplifies losses when the market direction turns out to be unfavorable.

Investor redemptions can also magnify losses for hedge funds.

Hedge Fund Strategies

Hedge funds utilize alternative investment strategies for the purpose of achieving superior returns relative to risk (i.e., return vs. standard deviation). Performance objectives range from conservative to aggressive. The degree of hedging varies. In fact, some do not hedge at all while others simply use S&P put options and futures in lieu of shorting equities. Consequently, there is a broad spectrum of expected risk and return within the hedge fund universe.

Hedge funds can be classified in a variety of ways. Here is one way of classification (by investment strategy):

- **Equity hedge strategies** take long and short positions in equity and equity derivative securities. For example, the key feature of *market neutral funds* is the low correlation between their returns and the general market's movements. Other examples include fundamental growth, fundamental value, quantitative directional, short bias and sector specific strategies.

- **Event-driven investing** is an investing strategy that seeks to exploit pricing inefficiencies that may occur before or after a corporate event, such as a bankruptcy, merger, acquisition or spinoff.

    - *Merger arbitrage.* Before the effective date of a merger, the stock of the acquired firm typically sells at a discount to its announced acquisition value. A risk arbitrage involves buying stocks of the acquired firm and simultaneously selling the stocks of the acquirer. However, there is the risk that the merger may fall though.
    - *Distressed debt investing.* The securities of companies having financial problems usually sell at deeply discounted prices. Distressed securities funds take bets on the debt and/or equity securities of such companies. For example, if a fund manager believes such a company will successfully return to profitability, he or she will buy its securities. If the manager believes the company's situation will deteriorate, he or she will take a short position in its securities.
    - *Activist.* A fund takes large positions in companies and uses the ownership to participate in the management.
    - *Special situations*, such as corporate spin-offs.

- A **relative-value arbitrage strategy** seeks to take advantage of price differentials between related financial instruments, such as stocks and bonds, by simultaneously buying and selling the different securities - thereby allowing investors to potentially profit from the "relative value" of the two securities. Examples include fixed income convertible arbitrage, fixed income asset backed, fixed income general, volatility, and multi-strategy.

- **Macro funds** take bets on the direction of a market, a currency, an interest rate, a commodity, or any macroeconomic variable. For example, George Soros of the Quantum fund took a billion dollar profit from his historical bet against Sterling and the Bank of England in September 1992.

In terms of performance, hedge funds are generally viewed as having:

- Net returns higher than those available for equity or bond investments.
- A low correlation with conventional investments.

However, the performance data from hedge fund databases and indices suffer from serious biases such as self-selection bias, instant history bias and survivorship bias.

Hedge Fund Valuation Issues

Questions to ask:

- Which price to use? Bid price, ask price, average price or estimated value?
- Any liquidity discounts? The lack of liquidity under extreme market conditions can cause irreversible damage to hedge funds whose strategies rely on the presence of liquidity in specific markets.

Due Diligence

Generally, due diligence refers to the care a reasonable person should take before entering in an agreement or transaction with another party. The due diligence that has to be performed by an institutional investor when selecting a hedge fund is highly specialized and time consuming, given the secretive nature of hedge funds and their complex investment strategies.

The key factors to consider include investment strategy, investment process, competitive advantage, track record, size and longevity, management style, key-person risk, reputation, investor relations, plans for growth, and systems risk management.

5. Private Capital

```
<p> </p>Private equity firms generally buy companies, repair them, enhance them, and s
```

Private Equity Structure and Fees

A private equity firm is typically made up of limited partners (LPs) and one general partner (GP). The LPs are the outside investors who provide the capital. They are called limited partners in the sense that their liability extends only to the capital they contribute.

GPs are the professional investors who manage the private equity firm and deploy the pool of capital. They are responsible for all parts of the investment cycle including deal sourcing and origination, investment decision-making and transaction structuring, portfolio management (the act of overseeing the investments that they have made) and exit strategies.

The GP charges a **management fee** based on the LPs' **committed capital**. Generally, after the LPs have recovered 100% of their invested capital, the remaining proceeds are split between the LPs and the GP with 80% going to LPs and 20% to the GP.

The **clawback provision** gives the LPs the right to reclaim a portion of the GP's carried interest in the event that losses from later investments cause the GP to withhold too much carried interest.

Private Equity Strategies

Private equity investors have four main investment strategies.

The **leveraged buyout (LBO)** is a strategy of equity investment whereby a company is acquired from the current shareholders, typically with the use of financial leverage.

A buyout fund seeks companies that are undervalued with high predictable cash flow, low leverage and operating inefficiencies. If it can improve the business, it can sell the company or its parts, or it can pay itself a nice dividend or pay down some company debt to deleverage.

In a **management buyout (MBO)**, the current management team is involved in the acquisition. Not only is a far larger share of executive pay tied to the performance of the business, but top managers may also be required to put a major chunk of their own money into the deal and have an ownership mentality rather than a corporate mentality. Management can focus on getting the company right without having to worry about shareholders.

Venture capital is financing for privately held companies, typically in the form of equity and/or long-term debt. It becomes available when financing from banks and public debt or equity markets is either unavailable or inappropriate.

Venture capital investing is done in many stages from seed through mezzanine. These stages can be characterized by where they occur in the development of the venture itself.

- Formative-stage financing includes angel investing, seed-stage investing and early stage financing. The capital is used from the idea stage, to product development, and pre-commercial production stage.
- Later-stage financing is capital provided after commercial manufacturing and sales have begun but before any initial public offering.
- Mezzanine (bridge) financing is capital provided to prepare for the step of going public and represents the bridge between the expanding company and the IPO.

Growth capital (minority equity investments) earns profits from funding business growth or restructuring.

Distressed investing. A distressed opportunity typically arises when a company, unable to meet all its debts, files for Chapter 11 (reorganization) or Chapter 7 (liquidation) bankruptcy. Investors who understand the true risks and values involved can scoop up these securities or claims at discounted prices, seeing the glow beneath the tarnish.

Exit Strategies

Every private equity investment starts with the end in mind: no fund will support a private equity company without a clear exit plan. Common exit strategies include trade sale, IPO, recapitalization, secondary sales, and write off or liquidation.

Other Considerations

Risk and return:

- Higher long-term return opportunities than traditional investments.
- Performance maybe overstated due to various biases.
- Riskier than common stocks.
- Can add diversity to a portfolio of traditional investments.

Valuation approaches: market or comparable, discounted cash flow (DCF), and asset based.

Private Debt

Private debt refers to various forms of debt provided by investors to private entities. There are four categories.

Direct lending provides funds to borrowers that lack favorable alternatives to traditional bank lenders. The rates are normally higher, and the number of borrowers is usually very small. A **leveraged loan** is a type of loan that is extended to companies or individuals that already have considerable amounts of debt or poor credit history.

Mezzanine debt is the middle layer of capital that falls between secured senior debt and equity.

- junior ranking;
- unsecured;
- equity participation.

Mezzanine debt can be used as a financing source for corporate expansion projects, acquisitions, recapitalizations, management buy-outs (MBO) and leveraged buy-outs (LBO).

Venture debt is a type of loan designed specifically for early-stage, high-growth companies with venture capital backing. Instead of collateral, the lenders are compensated with the company's warrants on common equity for the high-risk nature of the debt instruments. The vast majority of venture-backed companies raise venture debt at some point in their lives from specialized banks such as Silicon Valley Bank.

Distressed debt refers to the securities of an issuer which has either defaulted, is under bankruptcy protection, or is in financial distress and moving toward the aforementioned situations in the near future. Why? The greater the level of risk you assume, the higher the potential return. Investors purchase these bonds at a steep discount of their face value in the anticipation that the company will successfully emerge from bankruptcy as a viable enterprise.

Private debt also includes specialized strategies, such as CLOs, unitranche debt, real estate debt, and infrastructure debt.

Investing in private debt is riskier than investing in traditional bonds. It can certainly add diversity to a traditional portfolio because it has less than perfect correlation with those investments.

6. Natural Resources

```
    <p> </p>Natural Resources include commodities (hard and soft), agricultural land, and
```

Characteristics of Natural Resources

Commodities

Commodity returns are based on changes in price and do not include an income stream such as dividends, interests, or rent.

Commodity derivatives are financial instruments that derive their value from the value of the underlying commodities. They include commodity futures, forwards, options and swaps.

Commodity spot prices are determined by market supply and demand.

The price of a commodity futures contract is determined by the spot price, risk-free rate, storage costs and convenience yield.

Futures Price â‰ˆ Spot Price (1 + r) + Storage Costs - Convenience Yield

Convenience yield is the additional value gained by holding the commodity rather than having a long forward or futures contract on it, such as the ability to take advantage of shortages.

Under normal market conditions, the spot price would be lower than the futures price due to the cost to carry. This is referred as normal carry charge market or **contango**.

Under abnormal circumstances such as shortage of supply of a commodity in the short term, the contango can be reduced or even reversed. In this case the spot price can become higher than the futures price. This process is known as **backwardation**.

Timberland and Farmland

Timberland offers an income stream based on the sale of trees, wood and other products. It can be thought of as both a factory and a warehouse. Plus, it is a sustainable investment that migrates climate-related risks.

Farmland has an income stream component related to harvest quantities and agricultural commodity prices. However, farmland does not have the production facility of timberland, because farm products must be harvested when ripe.

Risk/Returns of Natural Resources

Commodity investment objectives:

- Potential for returns;
- Portfolio diversification.
- Inflation protection.

Commodity spot prices are determined by supply and demand. Supplies of commodities are determined by production and inventory levels. The supply levels cannot be adjusted quickly. Demand levels are influenced by global manufacturing dynamics and economic growth.

Weather is a unique and exogenous risk for farmland and timberland. Global international competition environment can also cause interruptions in crop prices.

Diversification Benefits

Commodities are effective inflation hedges. In fact, some commodity prices are included in inflation calculations.

Because of the low correlation between commodities returns and traditional asset classes' returns, commodities are effective in portfolio diversification.

For some investors, there are ESG considerations by including timberland and farmland in their portfolios, as timber and crops consume carbon as part of the plant life cycle.

Instruments

Commodity investment may involve investing in actual physical commodities or in producers of commodities. However, most investors do not want to get involved in storing commodities such as cattle or crude oil. Typically, such investments are made using commodity derivatives (futures or swaps), which can be traded on exchanges or OTC.

There are also other means of achieving commodity exposure: Exchange-traded products (ETPs), CTAs, and funds that specialize in specific commodity sectors.

Investment funds are primary investment vehicles for timber and farmland.

7. Real estate

```
    <p> </p>In general, real estate refers to buildable lands and buildings, including res
```

There are plenty of ways to invest in real estate without owning, financing, and operating physical properties.

Overview of the Real Estate Market

Characteristics of real estate as an investable asset class:

- Properties are immovable and basically indivisible so they are illiquid.
- Every property is unique, primarily because no two properties can share the same location. In addition, terms and conditions of transactions may differ significantly. Therefore, properties are only approximately comparable to other properties.
- There is no national, or international, auction market for properties. Therefore it is difficult to assess the market value of a given property.
- Transaction costs and management fees for real estate investments are high.
- Real estate markets suffer inefficiencies because of the nature of real estate itself and because information is not freely available.

Forms of Real Estate Ownership

They may be classified along two dimensions, debt or equity based, and in private or public markets. There are many variations within the basic forms.

Direct ownership refers to full ownership rights for an indefinite period of time, giving the owner the right, for example, to lease the property to tenants and resell the property at will. Owners can borrow money. **Leveraged ownership** refers to the same ownership rights but subject to debt (such as a promissory note) and/or a pledge (mortgage) to hand over real estate ownership rights if the loan terms are not met. The benefits of of direct ownership are control, and tax benefits.

Mortgages represent a type of debt investment as a mortgage provides the investor a stream of bondlike payments. This is a form of real estate investment as the creditor may end up with owning the property being mortgaged. To diversify risks a typical investor often invests in securities issued against a pool of mortgages.

Aggregation vehicles aggregate investors and serve the purpose of giving investors collective access to real estate investments.

Real Estate Partnerships (RELPs). A RELP is a professionally managed real estate syndicate that invests in various types of real estate. The purpose of the RELP varies from raw land speculation to investments in income producing properties. Managers assume the role of general partner with unlimited liability, while other investors are treated like limited partners with limited liability.

Real Estate Investment Trusts (REITs). A REIT is a type of closed-end investment company that sells shares to investors and invests the proceeds in various types of real estate and real estate mortgages.

- It allows small investors to receive both the capital appreciation and the income returns without the headache of property management.
- Its shares are traded on a stock market.
- It provides a tax shelter.
- It also has strong restriction on the use, and distribution of funds.

Investment Categories

Residential properties: an individual or a family purchases a home. In most cases a financial institution makes a direct debt investment in the home by offering a mortgage. The largest sector of the real estate market by value and size is the residential real estate.

Commercial real estate: a direct equity and/or debt investment is made into a property which is then managed to generate economic benefit to the parties. It includes office building, shopping centers, and warehouses.

REIT investing: mortgage REITs invest in mortgages and equity REITs invest in commercial and residential properties.

Mortgage-backed securities (MBS): securitization of mortgages. MBS maybe issued privately or publicly.

Risk and Return Characteristics

There are different types of indices: **appraisal indices**, **repeat sales indices** and **REIT indices**. They are constructed differently and have their own limitations such as sample selection biases and understated volatility. They are not necessarily representative of the entire real estate universe due to different geographic concentrations, property types, and asset quality. The low volatility and low correlation with other asset classes may result from these limitations.

Real estate investing can be lucrative, but it's important to understand the risks. Key risks include bad locations, negative cash flow, high vacancies, and problem tenants. Other risks to consider are the lack of liquidity, hidden structural problems, and the unpredictable nature of the real estate market. Leverage also magnifies the effects of both gains and losses for both equity and debt investors.

Diversification Benefits

Due to real estate's moderate correlation with other asset classes, it has been demonstrated to reduce portfolio risk. However, during steep market downturns the correlations between equity REIT and other asset classes can be high.

Valuation

This last section was in previous years' required readings but not in this year's. It is presented here for reference purpose only.

There are three commonly used approaches to real estate market value:

The **comparison sales approach** uses as the basic input the sales prices of properties (benchmark value) that are similar to the subject property. The price must be adjusted to reflect its superiority or inferiority to comparable properties. This approach can give a good feel for the market.

The **income approach** calculates a property's value as the present value of all its future income. It assumes that the annual net operating income (NOI) of a property can be maintained at a constant level forever (that is, NOI is a perpetuity). The most popular income approach is called direct capitalization:

Net operating income (NOI) equals the amount left after subtracting vacancy and collection losses and property operating expenses from an income property's gross potential rental income.

The market capitalization rate is obtained by looking at recent market sales figures to determine the rate of return required by investors.

The discounted cash flow approach is a variation of the income approach.

The **cost approach** is based on the idea that an investor should not pay more for a property than it would cost to rebuild it at today's prices. It generally works well for new or relatively new buildings. Most experts use it as a check against a price estimate. Limitations:

- An appraisal of the land value is not always an easy task.
- The market value of an existing property could differ significantly from its construction cost.

An income-based or asset-based approach can be used to value a REIT.

8. Infrastructure

```
    <p> </p>The assets underlying infrastructure investments are real, capital intensive,
```

Investment characteristics:

1. Significant capital investment. Funding is often done on a public-private partnership basis.
2. Monopolistic and regulated. Both 1 and 2 create high barriers to entry.
3. Highly leveraged financial structure, which enhances investor returns.
4. Stable long-term cash flows (adjust for economic growth and inflation).
5. Long operational lives.
6. Strategically important: this is most likely valuable to investors aiming to sell newly constructed assets to the government.
7. Defined risks: investment risks are clear and well defined.

Infrastructure investments may be categorized on the basis of the underlying assets.

- **Economic infrastructure assets** include transportation assets, information and communication technology assets, and utility and energy assets.
- **Social infrastructure assets** are directed toward human activities and include such assets as educational, health care, social housing, and correctional facilities, with the focus on providing, operating, and maintaining the asset infrastructure.

Infrastructure investments may also be categorized by the underlying asset's stage of development. Investing in infrastructure assets that are to be constructed is generally referred to as **greenfield investment**. Investing in existing infrastructure assets may be referred to as **brownfield investment**.

Infrastructure investments may also be categorized by location (local, regional, national) associated with the government entity directly involved with the assets.

There are different forms of infrastructure investments, such as direct investment, infrastructure funds and EFTs, and company shares.

Compared with other investments, infrastructure investments typically have lower expected risk, more stable cash flows and lower risk level. Risks tend to be well defined. They may match the longer-term liability structure of certain investors.

9. Issues in Performance Appraisal

```
    <p> </p>Alternative investments are only as good as their performance ratios. Conducti
```

Achieving portfolio diversification and reaping alpha are the end goals of many investors. However, traditional risk and return measures (such as mean return, standard deviation of returns, and beta) may provide an inadequate picture of alternative investments' risk and return characteristics. Moreover, these measures may be unreliable or not representative of specific investments.

There are several performance ratios that investors can use to evaluate alternative investment opportunities.

Sharpe ratio is the industry standard for measuring the risk-adjusted return of an investment. It is the average return earned in excess of the risk-free rate per unit of volatility, or total risk.

Limitation: The Sharpe ratio uses the standard deviation of returns in its formula as a proxy of total portfolio risk. It assumes that the returns are normally distributed. However, alternative investment returns do not tend to be normally distributed.

The **Sortino ratio** is a modified version of the Sharpe ratio that measures a risk-adjusted performance that only penalizes returns that fall below a target rate of return. Since the Sortino ratio does not use the standard deviation, it adjusts the performance for risk by only using the downside risk/deviation. It is especially helpful when investors are analyzing asset classes and portfolios that are highly volatile, since the ratio focuses on whether returns are negative or below a certain threshold.

Limitation: it fails to consider the correlation of alternative assets with the traditional assets, and thus fail to include the alternative investments' diversification potential.

The **Treynor ratio** is a measure of the excess average return of an investment relative to its beta to a relevant benchmark. It is different from Sharpe ratio in that it uses â€œmarketâ€ risk, or beta, instead of total risk (standard deviation). The lower the beta of the alternative asset, the higher the Treynor Ratio, and the better the portfolioâ€™s performance under analysis.

Limitation: it is based on historical beta data that may change in the future.

Other ratios (e.g. Calmar ratio, MAR ratio, batting average and slugging percentage) can also be used in performance appraisal.

Private equity and real estate investments often have variable cash flows, depending on the investment stage. The **IRR calculation** is often used to evaluate these investments. The **cap rate** is often used to evaluate real estate investments.

Many PE funds appear to have low volatility, This is because accounting conventions may simply leave longer-lived investments marked at their initial cost for some time until known impairments begin to transpire. As they are not easily marked to market, their returns can appear somewhat smoothed.

Hedge funds:

- Leverage must be considered when evaluating hedge funds.
- If the hedge fund has invested in thinly-traded small-capitalization stocks, liquidity can be a concern: which price should be used for valuations?
- Redemption rules and lockup periods can bring special challenges to performance appraisal of alternative investments.

10. Calculating Fees and Returns

```
<p> </p>Two and twenty (or "2 and 20") is a fee arrangement that is standard in the he
```

Analysts need to be aware of any custom fee arrangements in place that will affect the calculation of fees and performance.

- Fee discounts based on custom liquidity terms or significant asset size. For example, in exchange for reduced fees, an investor may agree to a longer lock-up period. There are share classes with lower fees for larger investors.

- Special share classes. For example, "Founder Class" investors can pay a 25% to 50% lower performance fee than other classes in the fund until the fund reaches a certain asset size or an agreed upon period of time has elapsed.

- "Either/or" fees. Manager can either charge a fixed management fee or a higher performance fee but not both. The exact mechanics of the arrangement can be calibrated such that the economics of the fee structure remain similar, whilst the alignment with the investor is strengthened.

When comparing the performance of alternative investments versus an index, the analyst must be aware that indexes for alternative investments may be subject to a variety of biases.

- **Survivorship bias** is the situation where unsuccessful funds are removed from the index, and the past index values are adjusted to remove the data of the dropped fund. Since a fund is more likely to be dropped from an index because of poor performance, such actions create bias in the index.

- **Backfill biases** means that when a new hedge fund is added to an index, the past performance of the fund is back-filled in the index. For example, if the hedge fund is 3 years old, it's record for the past three years will be added to the index, and the index values will be adjusted accordingly. The successful funds are more likely to be added to an index than an unsuccessful one, which creates a bias in the index.

Portfolio Management

Portfolio Management (1)

Portfolio Management: An Overview

1. A Portfolio Perspective on Investing

```
Why should investors take a portfolio approach instead of investing in individual stoc
```

Portfolio theory is used to maximize an investment's expected rate of return for a given level of risk or minimize the level of risk for a given expected rate of return.

For the purpose of investing, risk is defined as the variation of the return from what was expected (volatility). It is represented by a measure such as standard deviation.

Diversification is used to reduce a portfolio's overall volatility. Building a portfolio out of many unrelated (uncorrelated) investments minimizes total volatility (risk). The idea is that most assets will provide a return similar to their expected return and will offset those in the portfolio that perform poorly. The **diversification ratio** is the ratio of the standard deviation of an equally weighted portfolio to the standard deviation of a randomly selected security.

- The composition of a portfolio matters a great deal. Different portfolios have different risk-return trade-offs.
- Portfolio diversification does not necessarily provide the same level of risk reduction during times of severe market turmoil as it does when the economy and markets are operating normally.
- The modern portfolio theory says that the value of an additional security to a portfolio ought to be measured along with its relationship to all of the other securities in the portfolio.

2. Steps in the Portfolio Management Process

```
<b>Step One: The Planning Step</b><p> </p>
```

The first step in the portfolio management process is to understand the client's needs and develop an **investment policy statement** (IPS).

The IPS covers the types of risks the investor is willing to assume along with the investment goals and constraints. It should focus on the investor's short-term and long-term needs, familiarity with capital market history, and investor expectations and constraints. Periodically the investor will need to review, update, and change the policy statement.

A policy statement is like a road map: it forces investors to understand their own needs and constraints and to articulate them within the construct of realistic goals. It not only helps investors understand the risks and costs of investing, but also guides the actions of portfolio managers.

Step Two: The Execution Step

The second step is to construct the portfolio. The portfolio manager and the investor determine how to allocate available funds across different countries, asset classes, and securities. This involves constructing a portfolio that will minimize the investor's risk while meeting the needs specified in the policy statement.

Step Three: The Feedback Step

The process of managing an investment portfolio never stops. Once the funds are initially invested according to plan, the real work begins: monitoring and updating the status of the portfolio and the investor's needs.

The last step is the continual monitoring of the investor's needs, capital market conditions, and, when necessary, updating the policy statement. One component of the monitoring process is evaluating a portfolio's performance and comparing the relative results to the expectations and requirements listed in the policy statement. Some rebalancing may be required.

3. Types of Investors

```
There are different types of investment clients.<p> </p>
```

Different **individual investors** have different investment goals, levels of risk tolerance, and constraints. Some seek growth while others may invest to get regular income.

An **institutional investor**'s role is to act as a highly specialized investor on behalf of others. There are many types of institutional investors.

A **pension plan** is a fund that provides retirement income to employees. It is typically considered a long-term investor with high risk tolerance and low liquidity needs.

- In a **defined contribution plan**, the employer agrees to contribute a certain sum each period based on a formula. Only the employer's contribution is defined; no promise is made regarding the ultimate benefits paid out to the employee. The employee accepts the investment risk.
- A **defined benefit plan** defines the benefits that the employee will receive at the time of retirement. That is, the employer assumes the risk of the investment, and is responsible for the payment of the defined benefits regardless of what happens in the investment.

An **endowment** or a **foundation** is an investment fund set up by an institution in which regular withdrawals from the invested capital are used for ongoing operations. Endowments and foundations are often used by universities, hospitals, and churches. They are funded by donations. A typical investment object is to maintain the real value of the fund while generating income to fund the objectives of the institution.

A **bank** typically has a very short investment horizon and low risk tolerance. Its investments are usually conservative. The investment objective of a bank's excess reserves is to earn a return that is higher than the interest rate it pays on its deposit.

Investments made by **insurance companies** are relatively conservative. Although the income needs are typically low, the liquidity needs of such investments are usually high (in order for insurance companies to pay claims).

Both the risk tolerance and the return requirement of **mutual funds** are predefined for each fund and can vary sharply between funds. They are more specialized than pension funds or insurance companies. Study Session 19 discusses mutual funds in more detail.

A **sovereign wealth fund** is a state-owned investment fund. There are two types of funds: saving funds and stabilization funds. Stabilization funds are created to reduce the volatility of government revenues, to counter the boom-bust cycles' adverse effect on government spending and the national economy.

4. The Asset Management Industry

```
   <p> </p>The asset management industry is an integral component of the global financial
```

A **sell-side firm** sells securities, and provides independent investment research and recommendations to its clients, which are **buy-side firms**. A buy-side firm buys securities for the purpose of money or fund management.

Active management requires frequent buying and selling in an effort to outperform a specific benchmark or index. Active management portfolios strive for superior returns but take greater risks and entail larger fees. **Passive management** replicates a specific benchmark or index in order to match its performance.

Traditional asset managers focus on long-only equity and fixed-income strategies, while **alternative asset managers** focus on non-traditional assets such as hedge funds, private equities, and venture capital investments.

Most asset managers are privately owned, structured as limited liability companies or limited partnerships. Some asset managers are publicly traded now.

Industry trends:

- Growth of passive investing, due to low cost and challenges faced by active managers to beat the market.
- Use of "big data" in the investment process. The challenge is how to convert various forms of data into alpha-generating portfolio and security-level decisions.
- Robo-advisers: an expanding wealth management channel. This is based on several industry trends, including growing demand from mass affluent and younger investors, lower fees and new entrants.

5. Pooled Investments

```
   "Pooled investments" is a term given to a wide range of investment types, such as mutu
```

When you invest in a pooled investment, your money goes into an investment fund. You pool your money with others to help spread the risk. Professional fund managers then invest the money on your behalf in a highly competitive environment.

Mutual Funds

An investment company invests a pool of funds belonging to many individuals in a single portfolio of securities. In exchange for this commitment of capital, the investment company issues to each investor new shares representing his or her proportional ownership of the mutually held securities portfolio (commonly known as a mutual fund).

Mutual funds are classified according to whether or not they stand ready to redeem investor shares.

- An **open-end mutual fund** continues to sell and repurchase shares after its initial public offering. It stands ready to redeem investor shares at market value.
- A **closed-end mutual fund** operates like any other public firm. It is initiated through a stock offering to raise capital. Its stock trades on the regular secondary market and the market price is determined by

supply and demand. A typical closed-end fund offers no further shares and does not repurchases the shares on demand (no funds can be withdrawn). Therefore, investors must trade in public secondary markets (e.g., NASDAQ) to buy or sell shares.

Various fees charged by mutual funds:

- They charge fees for their efforts of setting up funds. Sales commissions are charged at purchase (**front-end load**) as a percentage of the investment.
    - A **load fund** has sales commission charges. A load fund's offering price = NAV of the share + a sales charge (7.5 - 8% of the NAV). The NAV price is the redemption (bid) price and the offering (ask) price equals the NAV divided by 1 minus the percent load.
    - A **no-load fund** imposes no initial sales charge.
- **Redemption fee** (**back-end load**). A charge to exit the fund. This discourages quick trading turnover; these funds are set up so that the fees decline the longer the shares are held (in this case, the fees are sometimes called contingent deferred sales charges). Load funds generally charge no redemption fees.
- All mutual funds charge annual fees.

There are four types of mutual funds based on portfolio makeup.

- Money Market Funds. These funds attempt to provide current income, safety of principal, and liquidity by investing in diversified portfolios of short-term securities, such as T-bills, banker certificates of deposit, bank acceptances, and commercial paper. They generally allow holders to write checks again their account, so they are essentially cash holdings for holders. However, they are not insured in the same way as bank deposits.
- Bond Mutual Funds. Bond funds concentrate on various types of bonds to generate high current income with minimal risk. Bonds held include government bonds, high-grade corporate bonds, and junk bonds.
- Stock Mutual Funds. These funds invest almost solely in common stocks. Some funds focus on growth companies while others specialize in specific industries. Different stock mutual funds can suit almost any taste or investment objective.

  There are two investment styles.

    - **Passive mutual fund management** is a long-term buy-and-hold strategy. Usually stocks are purchased with the intention that the portfolio's returns will track those of an index over time. The purpose is not to beat the index but to match its performance.
    - **Active mutual fund management** is an attempt by the fund manager to outperform a passive benchmark portfolio on a risk-adjusted basis. Management fees are usually higher and there are usually more trading activities, which can cause tax consequences for investors.
- Hybrid/Balanced Funds. These funds diversify outside the stock market by combining common stock with fixed-income securities.

Other Investment Products

1. **Exchange Traded Funds**

Refer to Reading 47 (Introduction to Alternative Investments).

1. **Separately Managed Accounts**

The key difference between mutual funds and separate accounts is that, in a separate account, the money manager is purchasing the securities in the portfolio on behalf on the investor, not on behalf of the fund. Therefore, the investor can determine which assets are bought or sold, and when.

A mutual fund investor owns shares of a company (mutual fund) that in turn owns other investments, whereas an SMA investor owns the invested assets directly in his own name.

An investor in an SMA typically has the ability to direct the investment manager to sell individual securities with the objective of raising capital gains or obtaining losses for tax planning purposes. This practice is known as "tax harvesting"; its objective is to attempt to equalize capital gains and losses across all of the investor's accounts for a given year in order to reduce capital gains taxes owed.

Another major advantage of individual cost basis is the ability to customize the portfolio by choosing to avoid investing in certain stocks or certain economic sectors (technology, sin stocks, etc.).

1. **Hedge Funds**

A hedge fund is an investment fund, open to a limited range of investors, that undertakes a wider range of investment and trading activities than traditional long-only investment funds, and that, in general, pays a performance fee to its investment manager. Every hedge fund has its own investment strategy that determines the type of investments and the methods of investment it undertakes.

Unlike mutual funds, most hedge funds are not regulated. The net effect is that the hedge fund investor base is generally very different from that of the typical mutual fund.

Hedge funds employ many different trading strategies, which are classified in many different ways, with no standard system used. A hedge fund will typically commit itself to a particular strategy, particular investment types, and leverage limits via statements in its offering documentation, thereby giving investors some indication of the nature of the particular fund.

1. **Buyout and Venture Capital Funds**

Both funds take equity positions and plan a very active role in the management of the company. The equity they hold is private, and they don't have a long investment horizon.

- Buyout funds make only a few large investments in public companies with the intent of selling the restructured companies in three to five years.
- Venture capital funds buy start-up companies and grow them. They play a very active role in managing these companies.

Portfolio Risk and Return: Part I

1. Major Return Measures

```
There are various types of return measures.<p> </p>
```

Holding Period Return

Refer to Reading 2 for a detailed discussion of this return measure.

Arithmetic or Mean Return

Refer to Reading 2 for a detailed discussion of this return measure.

Geometric Mean Return

Refer to Reading 2 for a detailed discussion of this return measure.

Money-Weighted Return or Internal Rate of Return

The **dollar-weighted rate of return** is essentially the internal rate of return (IRR) on a portfolio. This approach considers the *timing* and *amount* of cash flows. It is affected by the timing of cash flows. If funds are added to a portfolio when the portfolio is performing well (poorly), the dollar-weighted rate of return will be inflated (depressed).

The **time-weighted rate of return** measures the compound growth rate of $1 initial investment over the measurement period. *Time-weighted* means that returns are averaged over time. This approach is not affected by the timing of cash flows; therefore, it is the preferred method of performance measurement.

Example

Jayson bought a share of IBM stock for $100 on December 31, 2000. On December 31, 2001, he bought another share for $150. On December 31, 2002, he sold both shares for $140 each. The stock paid a dividend of $10 per share at the end of each year.

To calculate the dollar-weighted rate of return, you need to determine the timing and amount of cash flows for each year, and then set the present value of net cash flows to be 0: $- 100 - 140/(1 + r) + 300/(1 + r)^2 = 0$. You can use the IRR function on a financial calculator to solve for r to get the dollar-weighted rate of return: r = 17%.

To calculate the time-weighted rate of return:

- Split the overall measurement period into equal sub-periods on the dates of cash flows.

For the first year:

-- beginning price: $100

-- dividends: $10

-- ending price: $150

- For the second year:

-- beginning price: $300 (150 x 2)

-- dividends: $20 (10 x 2)

-- ending price: $280 (140 x 2)

Calculate the holding period return (HPR) on the portfolio for each sub-period: HPR = (Dividends + Ending Price)/Beginning Price - 1. For the first year, HPR_1: (150 + 10)/100 - 1 = 0.60. For the second year, HPR_2: (280 + 20)/300 - 1 = 0.

Calculate the time-weighted rate of return:

- If the measurement period < 1 year, compound holding period returns to get an annualized rate of return for the year.
- If the measurement period > 1 year, take the geometric mean of the annual returns.

Annualized Return

Annualizing returns allows for comparison among different assets and over different time periods.

$$r_{annual} = (1 + r_{period})^c - 1$$

where c is the number of periods in a year and r_{period} is the rate of return per period.

Example

Monthly return: 0.6%. The annualized return is $(1 + 0.6\%)^{12} - 1 = 7.44\%$.

Portfolio Return

The expected return on a portfolio of assets is the market-weighted average of the expected returns on the individual assets in the portfolio.

where Rp is the return on the portfolio, Ri is the return on asset i and wi is the weighting of component asset i (that is, the share of asset i in the portfolio).

Other Major Return Measures

1. A **gross return** is the return before any fees, expense, taxes, etc. A **net return** is the return after deducting all fees and expenses from the gross return.

2. Different types of investments generate different types of income and have different tax implications. For example, in the U.S. the interest income is fully taxable at an investor's marginal tax rate while capital gains are taxed at a much lower rate. Therefore, many investors therefore use the **after-tax return** to evaluate mutual fund performance.

3. The **nominal return** and the **real return** are two ways to measure how well an investment is performing. The real return takes into consideration the effects of inflation when calculating how much buying power has changed.

4. An investor can also use **leverage** to amplify his expected return (and risk).

2. Historical Return and Risk

```
   The textbook examines the historical risk and return for the three main asset categori
```

T-bills: the safest investment on earth. The price paid for this safety is steep: the return is only 3.7%, which is barely above the inflation rate of 3.0% for the period. Further, although many academicians consider T-bills to be "riskless," a quick perusal of the T-bill graph shows considerable variation of return, meaning that you cannot depend on a constant income stream. This risk is properly reflected in the standard deviation of 3.1%. The best that can be said for the performance of T-bills is that they keep pace with inflation in the long run.

Long-term bonds carry one big risk: interest rates risk. The longer the maturity of the bond the worse the damage. For bearing this risk, investors are rewarded with another 1.5% of long-term return. In the long run, investors can expect a real return (inflation-adjusted) of about 2% with a standard deviation of 10%.

The rewards of stocks are considerable: a real return of greater than 6%. This return does not come free, of course. The standard deviation is 20%. You can lose more than 40% in a bad year; during the calendar years 1929-32 the inflation-adjusted ("real") value of this investment class decreased by almost two-thirds.

$1 in 1900 would have grown to $582 in 2008 if invested in stocks, only $9.90 if invested in bonds, and to $2.90 if invested in T-bills. The message is clear: stocks are to be held for the long term. Don't worry too much about the short-term volatility of the markets; in the long run stocks will almost always have higher returns than bonds.

Stocks have outperformed bonds consistently over long periods of time. However, stocks are much riskier and investors demand compensation for bearing the risk. The question is: is the premium too big?

Other Investment Characteristics

Two assumptions are usually made when investors perform investment analysis using mean and variance.

- Returns are normally distributed.
- Markets are operationally efficient.

Is normality a good approximation of returns? In fact, returns are not quite normally distributed. The biggest departure from normality is that extremely bad returns are more likely than predicted by the normal distribution (fat tails).

There are operational limitations of the market that affect the choice of investments. One such limitation is liquidity, which affects the cost of trading.

3. Variance and Covariance of Returns

```
    Investment is all about reward versus variability (risk). The return measures  the rew
```

The **variance** is a measure of how spread out a distribution is. It is computed as the average squared deviation of each number from its mean. The formula for the variance in a population is:

where Î¼ is the mean and N is the number of scores.

To compute variance in a sample:

where m is the sample mean.

The formula for the **standard deviation** is very simple: it is the square root of the variance. It is the most commonly used measure of spread.

The **standard deviation of a portfolio** is a function of:

- The weighted average of the individual variances, plus
- The weighted covariances between all the assets in the portfolio.

In a two-asset portfolio:

The maximum amount of risk reduction is predetermined by the correlation coefficient. Thus, the correlation coefficient is the engine that drives the whole theory of portfolio diversification.

Example with perfect positive correlation (assume equal weights):

What is the standard deviation of a portfolio (E), assuming the following data?

Ïƒ$_1$ = 0.1, w$_1$ = 0.5, Ïƒ$_2$ = 0.1, w$_2$ = 0.5, Ï$_{12}$ = 1

Solution:

Cov$_{12}$ = Ïƒ$_1$ x Ïƒ$_2$ x Ï$_{12}$ = 0.1 x 0.1 x 1 = 0.01

Standard Deviation of Portfolio $[0.5^2 \text{ x } 0.1^2 + 0.5^2 \text{ x } 0.1^2 + 2 \text{ x } 0.5 \text{ x } 0.5 \text{ x } 0.01]^{1/2} = 0.10$ (perfect correlation)

If there are three securities in the portfolio, its standard deviation is:

4. Risk Aversion and Portfolio Selection

```
    <b>Risk Aversion</b><p> </p>
```

Every investor wants to maximize the investment returns for a given level of risk. Risk refers to the uncertainty of future outcomes. **Risk aversion** relates to the notion that investors as a rule would rather avoid risk. Given a choice of two investments with equal returns, risk-averse investors will select the investment with lower risk. Investors are risk-averse. Consequently, investors will demand a *risk premium* for taking on additional levels of risk. The more risk-averse the investor, the more of a premium he or she will demand prior to taking on risk.

Investors who do not demand a premium for risk are said to be **risk-neutral** (e.g., willing to place both a large and small bet on the flip of a coin and be indifferent) and those investors that enjoy risk are said to be **risk seekers** (e.g., people who buy lottery tickets despite the knowledge that for every $1 spent, on average they will get less than $0.1 back).

Example

Three investors, Sam, Mike, and Mary are considering two investments: A and B. Investment A is the less risky of the two, requiring an investment of $1,000 with an expected rate of return at 10%. Investment B also requires an investment of $1,000 and has an expected return of 10% but appears to have considerably more variability in potential returns than A. Sam requires a return of 14%, Mike requires 10%, and Mary seeks only an 8% return.

Question: Given the information above, which of the three investors is considered risk-averse?

Solution: Only Sam would be considered risk-averse. He is the only investor who demands a premium of return given the higher risk level. Mike would be considered risk-neutral since he demands no premium in return (despite the higher risk) and Mary would be considered a risk-seeker since she, in fact, will accept less return for a riskier situation.

Risk aversion implies that there is a positive relationship between expected returns (ER) and expected risk (Es), and that the risk return line (CML and SML) is upward-sweeping.

Evidence that suggests that individuals are generally risk-averse:

- **Purchase of insurance.** Most investors purchase various types of insurance (e.g., life insurance, car insurance, etc.). By buying insurance, an investor avoids the uncertainty of a potential large future cost by paying the current known cost of the insurance policy.
- **Difference in the promised yield for different grades of bonds.** The promised yield of a bond is its required rate of return. Different grades of bonds have different degrees of credit risk. The promised yield increases as you go from the lowest-risk grade (e.g., AAA) to a grade with higher risk (e.g., AA). That is, as the credit risk of a bond increases, investors will require a higher rate of return.

Utility Theory

Although investors differ in their risk tolerance, they should be consistent in their selection of any portfolio in terms of the risk-return trade-off. Because risk can be quantified as the sum of the variance of the returns over time, it is possible to assign a utility score (aka utility value, utility function) to any portfolio by subtracting its variance from its expected return to yield a number that would be commensurate with an investor's tolerance for risk, or a measure of their satisfaction with the investment. Because risk aversion is not an objectively measurable quantity, there is no unique equation that would yield such a quantity, but an equation can be selected, not for its absolute measure, but for its comparative measure of risk tolerance. One such equation is the following utility formula:

Utility Score = Expected Return - 0.5 x Ïƒ^{2} A

where A is the risk aversion coefficient (a number proportionate to the amount of risk aversion of the investor). It is positive for a risk-averse investor, zero for a risk-neutral investor, and negative for a risk seeker.

For example, if a T-bill pays 4%, and XYZ stock has a return of 12% and a standard deviation of 25%, and an investor's risk aversion coefficient is 2, his utility score of XYZ stock is equal to: 12% - 0.5 x 0.25^2 x 2 = 5.75%.

If someone were more risk-averse, we might use 3 instead of 2 to indicate the investor's greater aversion to risk. In this case, the above equation yields: 12% - 0.5 x 25^2 x 3 = 2.63%.

Since 2.63% is less than the 4% yield of risk-free T-bills, this risk-averse investor will reject XYZ stock in favor of T-bills while the other investor will invest in XYZ stock, since he assigns a utility score of 5.75% to the investment (which is higher than the T-bill yield).

A few conclusions about utility:

- It is unbounded on both sides.
- Higher return contributes to higher utility.
- Higher variance reduces utility.
- Utility does not measure satisfaction. It can be used to rank different investments.
- Utility cannot be compared among individuals; it's a very personal concept.

Indifference Curves

The set of all portfolios with the same utility score plots as an **indifference curve**.

An investor's indifference curves specify his or her preferences when making risk-return trade-offs. He or she will accept any portfolio with a utility score on his or her risk-indifference curve as being equally acceptable.

- The investments along each curve are equally attractive to the investor.
- The slope of the curves represents how risk-averse the investor is. Steep indifference curves indicate a conservative investor while flat indifference curves indicate a less risk-averse investor.

For a risk-averse investor:

- Every indifference curve runs from the southwest to the northeast.
- Every indifference curve is convex.
- The slope coefficient of an indifference curve is closely related to the risk aversion coefficient.

The Capital Allocation Line

Suppose we construct a portfolio (P) that combines a risky asset i with an expected return of r_i and standard deviation of Ïƒ$_i$, and a riskless asset with a return of r_f. Let w_1 represent the fraction of the total portfolio value placed in the riskless asset.

- The portfolio return $E(r_p)$ is given by $E(r_p) = w_1 r_f + (1 - w_1) E(r_i)$.
- The portfolio standard deviation is given by Ïƒ$_p$ = $(1 - w_1)$ Ïƒ$_i$

Example

Combine the S&P and a T-bill in a portfolio. $E(r_{S\&P})$ = 13%, Ïƒ$_{S\&P}$ = 20.3%, and r_f= 3.8%. Some of the possible portfolios are:

- w_1 = 0, $E(r_p) = E(r_{S\&P})$ = 13%, and Ïƒ$_p$ = 20.3%.
- w_1 = 0.5, $E(r_p)$ = 0.5 x 0.038 + 0.5 x 0.13 = 8.4%, and Ïƒ$_p$ = 0.5 x 20.3% = 10.1%.
- w_1 = 1, $E(r_p) = r_f)$ = 3.8%, and Ïƒ$_p$ = 0%.
- w_1 = -0.5, $E(r_p)$ = -0.5 x 0.038 + 1.5 x 0.13 = 17.6%, and Ïƒ$_p$ = 1.5 x 0.203 = 30.5%. This means there is negative investment in the riskfree asset: the investor borrowed at the risk-free rate. This is called a leveraged position in the risky asset - some of the investment is financed by borrowing.

The portfolio's expected return and standard deviation obey a liner relation:

- The slope, [(E(r$_i$ - r$_f$)/Ïƒ$_i$], is known as the portfolio's **Sharpe measure** or **reward-to-variability ratio**.
- The intercept is the risk-free rate.
- The line is called the **capital allocation line** (CAL). When i is a marker index portfolio, the line is called the capital market line (CML).
- A leveraged position is to the right of P.

CAL shows one simple fact: increasing the amount invested in the risky asset increases the expected return by a certain risk premium.

Now the investor must find the point of highest utility on CAL.

The optimal choice for an investor is the point of tangency of the highest indifference curve to the CAL - slope of the indifference curve is equal to the slope of the CAL.

The optimal w$_1$ will be higher for investors with higher A.

5. Portfolio Risk

```
   Consider two mutual funds, D (specialized in bonds and debt securities) and E (special
```

Expected return of the portfolio: E(r$_p$) = w$_D$ E(r$_D$) + w$_E$ E(r$_E$)

Variance of the portfolio: Ïƒ$_p^2$ = w$_D$

2Ïƒ$_D^2$ + w$_E^2$ Ïƒ$_E^2$ + 2 w$_D$ w$_E$ Cov(r$_D$, r$_E$)

= (w$_D$Ïƒ$_D$)^2 + (w$_E$ÏƒE)^2 + 2 (w$_D$ÏƒD) (w$_E$ÏƒE) Ï$_{DE}$

= (w$_D$Ïƒ$_D$ + w$_E$Ïƒ$_E$)^2 + 2 (w$_D$Ïƒ$_D$) (w$_E$Ïƒ$_E$) (Ï$_{DE}$ - 1)

= (w$_D$Ïƒ$_D$ - w$_E$Ïƒ$_E$)^2 + 2 (w$_D$Ïƒ$_D$) (w$_E$Ïƒ$_E$) (Ï$_{DE}$ + 1)

If the two assets are not perfectly positively correlated, the standard deviation of the portfolio is less than the weighted average of the standard deviations of the assets.

Covariance of returns measures the degree to which the rates of return on two securities *move together* over time.

- A *positive* covariance indicates that the rates of return on the two securities tend to move in the *same* direction.
- A *negative* covariance indicates that the rates of return on the two securities tend to move in *opposite* directions.
- A covariance of *zero* indicates that there is *no* relationship between the rates of return on the two securities.

The magnitude of the covariance depends on the magnitude of the individual stocks' standard deviations and the relationship between their co-movements. The **covariance** is an absolute measure of movement and is measured in return units squared. As the magnitude of the covariance is affected by the variability of return of each individual security, covariance cannot be used to compare across different pairs of securities.

The measure can be standardized by dividing the covariance by the standard deviations of the two securities being tested.

$p_{(1,2)} = \text{cov}_{(1,2)}/\text{Ï}f_1\text{Ï}f_2$

Rearranging the terms gives: $\text{cov}_{(1,2)} = p_{(1,2)}\text{Ï}f_1\text{Ï}f_2$.

The term $p_{(1,2)}$ is called the correlation coefficient between the returns of securities 1 and 2. The correlation coefficient has no units. It is a pure measure of the co-movement of the two stocks' returns. It varies in the range of -1 to 1.

How should you interpret the correlation coefficient?

- A correlation coefficient of +1 means that returns always move together in the same direction. They are perfectly positively correlated.
- A correlation coefficient of -1 means that returns always move in completely opposite directions. They are perfectly negatively correlated.
- A correlation coefficient of zero means that there is no relationship between the two stocks' returns. They are uncorrelated.

Example

Two risky assets, A and B, have the following scenarios of returns:

What is the covariance between the returns of A and B?

The expected return is a probability-weighted average of the returns. Using this definition, the expected return of A is 0.35 * 10% + 0.35_(-4%) + 0.3_(9%) = 4.8%. The expected return of B is 0.35 * 7% + 0.35_(4%) + 0.3_(-6%) = 2.05%.

The covariance between the returns equals the expected value of the product of the deviations of the individual returns from their means. *Remember this!*

To calculate this, we construct the following table:

The expected value of the product of the deviations is 0.35 * 5.2% * 4.95% + 0.35 * (-8.8%) * 1.95% + 0.3_4.2%_(-8.05%) = -0.0714%.

Portfolio Risk and Return

The **standard deviation of a portfolio** is a function of:

- The weighted average of the individual variances, plus
- The weighted covariances between all the assets in the portfolio.

When an asset is added to a large portfolio with many assets, the new asset affects the portfolio's standard deviation in two ways. It affects:

- The asset's own variance, and
- Covariance between this asset and *every other asset* in the portfolio. The effect of these numerous covariances will outweigh the effect of the asset's own variance. The more assets in the portfolio, the more this is true.

Therefore, the important factor to consider when adding an investment to a portfolio is not the investment's own variance, but its *average covariance* with all the other investments in the portfolio.

Adding securities that are not perfectly positively correlated with each other will reduce the standard deviation of the portfolio. The lower (higher) the correlations between the returns of assets in the portfolio, the lower (higher) the portfolio risk, and thus the higher (lower) the diversification benefits. The ultimate benefit of diversification occurs when the correlation between two assets is -1.00.

The graph below shows a portfolio with an overall standard deviation of zero, thus creating a risk-free portfolio. This occurs when two assets are combined, each moving in an opposite direction. For example, imagine a portfolio of investments, one of which moves with sun-related activities (e.g., sunglasses) and the other in the direction of rain-related activities (e.g., umbrellas). The combined portfolio of sunglasses and umbrellas ought to negate weather-related issues (theoretically speaking), as the two assets move in opposite directions.

- Returns on A move in the opposite direction from returns on B.
- The mid-point between A and B represents the mean return with no variability in returns.
- The overall standard deviation of this portfolio is zero.
- This is considered a risk-free portfolio.

In a portfolio with two assets, the ideal scenario provides a contrast in asset returns similar to the "saw tooth" diagram shown above. Thus, one asset would completely offset the other asset (in terms of risk) providing a smooth rate of return with no variability. This, of course, could only occur if the two assets had a perfect negative correlation.

The maximum amount of risk reduction is predetermined by the correlation coefficient. Thus, the correlation coefficient is the engine that drives the whole theory of portfolio diversification.

Let's re-visit our very first example. The portfolio opportunity set shows the portfolio return and risk for different Ïs.

Example with perfect positive correlation (assume equal weights)

What is the standard deviation of a portfolio (E) assuming the following data?

Ïƒ$_1$ = 0.1, w$_1$ = 0.5, Ïƒ$_2$ = 0.1, w$_2$ = 0.5, Ï$_{12}$ = 1

Solution:

Cov$_{12}$ = Ïƒ$_1$ x Ïƒ$_2$ x Ï$_{12}$ = 0.1 x 0.1 x 1 = 0.01

Standard Deviation of Portfolio = 0.10 (perfect correlation)

If there are three securities in the portfolio, its standard deviation is:

A **diversification** benefit is a reduction in a portfolio's standard deviation of return through diversification without an accompanying decrease in expected return. Portfolio diversification is affected by the number of assets in the portfolio and the correlation between these assets.

There are many ways to diversify investment risk.

- Diversify with asset classes.
- Diversify with index funds.
- Diversify among countries.
- Diversify by not owning your employer's stock.
- Evaluate each asset before adding it to a portfolio.
- Buy insurance for risky portfolios.

6. Efficient Frontier

The mean-variance portfolio theory says that any investor will choose the optimal port

- Maximize expected return for a given level of risk; and
- Minimize risks for a given level of expected returns.

Again, consider a situation where you have two stocks to choose from: A and B. You can invest your entire wealth in one of these two securities. Or you can invest 10% in A and 90% in B, or 20% in A and 80%in B, or 70% in A and 30% in B, or... There are a huge number of possible combinations even in the simple case of two securities. Imagine the different combinations you have to consider when you have thousands of stocks.

The **minimum-variance frontier** shows the minimum variance that can be achieved for a given level of expected return. To construct the minimum-variance frontier of a portfolio:

- Use historical data to estimate the mean, variance of each individual stock in the portfolio, and the correlation of each pair of stocks.
- Use a computer program to find the weights of all stocks that minimize the portfolio variance for each pre-specified expected return.
- Calculate the expected returns and variances for all the minimum-variance portfolios determined in step 2 and then graph the two variables.

The outcome of risk-return combinations generated by portfolios of risky assets gives you the minimum variance for a given rate of return. Logically, any set of combinations formed by two risky assets with less than perfect correlation will lie inside the triangle XYZ and will be convex.

Investors will never want to hold a portfolio below the minimum variance point. They will always get higher returns along the positively sloped part of the minimum-variance frontier.

The **efficient frontier** is the set of mean-variance combinations from the minimum-variance frontier where, for a given risk, no other portfolio offers a higher expected return.

Any point beneath the efficient frontier is inferior to points above. Moreover, any points along the efficient frontier are, by definition, superior to all other points for that combined risk-return tradeoff.

Portfolios on the efficient frontier have different return and risk measures. As you move upward along the efficient frontier, both risk and the expected rate of return of the portfolio increase, and no one portfolio can dominate any other on the efficient frontier. An investor will target a portfolio on the efficient frontier on the basis of his attitude toward risk and his utility curves.

The concept of efficient frontier narrows down the options of the different portfolios from which the investor may choose. For example, portfolios at points A and B offer the same risk, but the one at point A offers a higher return for the same risk. No rational investor will hold the portfolio at point B and therefore we can ignore it. In this case, A dominates B. In the same way, C dominates D.

7. Optimal Portfolio

The efficient frontier only considers the investments in risky assets. However, invest

When a risk-free asset is combined with a risky portfolio, a graph of possible portfolio risks-return combinations becomes a *straight line* between the two assets.

Assume the proportion of the portfolio the investor places in the tangency portfolio P is w_P:

- The expected rate of return for the new portfolio is the weighted average of the two returns: $E(R) = (1 - w_P) R_f + w_P E(R_T)$

- The standard deviation of the new portfolio is the linear proportion of the standard deviation of the risky asset portfolio P: Ïƒ$_{\text{portfolio}}$ = w$_P$ Ïƒ$_P$

The introduction of a risk-free asset changes the efficient frontier into a straight line. This straight efficient frontier line is called the **Capital Market Line (CML)** for all investors and the **Capital Allocation Line (CAL)** for one investor.

- Investors at point r$_f$ have 100% of their funds invested in the risk-free asset.
- Investors at point P have 100% of their funds invested in portfolio P.
- Between r$_f$ and P, investors hold both the risk-free asset and portfolio P. This means investors are lending some of their funds (buying the risk-free asset).
- To the right of P, investors hold more than 100% of portfolio P. This means they are borrowing funds to buy more of portfolio P. This represents a levered position.

Investors will choose the highest CAL (i.e., the CAL tangent to the efficient frontier). This portfolio is the solution to the optimization problem of maximizing the slope of the CAL.

Now, the line r$_f$-P dominates all portfolios on the original efficient frontier. Thus, this straight line becomes the new efficient frontier.

Separation Theorem

Investors make *different financing decisions* based on their risk preferences. The separation of the investment decision from the financing decision is called the **separation theorem**. The portfolio choice problem can be broken down into two tasks:

- Choosing P, a technical matter (can be done by the broker)
- Deciding on the proportion to be invested in P and in the riskless asset.

Optimal Investor Portfolio

We can combine the efficient frontier and/or capital allocation line with indifference curves. The **optimal portfolio** is the portfolio that gives the investor the greatest possible utility.

- Two investors will select the same portfolio from the efficient set only if their utility curves are identical.
- Utility curves to the right represent less risk-averse investors; utility curves to the left represent more risk-averse investors.

This is portfolio selection without a risk-free asset:

The optimal portfolio for each investor is the highest indifference curve that is tangent to the efficient frontier.

This is portfolio selection with a risk-free asset:

The optimal portfolio for each investor is the highest indifference curve that is tangent to the capital allocation line.

Portfolio Risk and Return: Part II

1. Capital Market Theory

```
<b>Introduction of a Risk-Free Asset</b><p> </p>
```

Adding a risk-free asset to the investment opportunities present on the efficient frontier effectively adds the opportunity to both borrow and lend. A U.S. Treasury bill (T-bill) is a common risk-free security proxy. Buying a T-bill loans the U.S. government money. Selling a T-bill short effectively borrows money. The concept of a risk-free asset is a major element in developing Capital Market Theory (CMT). Adding risk-free assets integrates investment and financing decisions. With risk-free asset:

- Expected return is entirely certain.
- Standard deviation of return is zero.
- Covariance with any risky asset or portfolio is always zero, as is the correlation.

The Capital Market Line

Introducing risk-free assets creates a set of expected return-risk possibilities that did not exist previously. The new risk-return trade-off is a straight line tangent to the efficient frontier at the market portfolio (point M) with a vertical intercept at the risk-free rate of return, R_f. This line is called the Capital Market Line (CML).

- The capital allocation line (CAL) is the graph of all possible combinations of the risk-free asset and the risky asset for one investor.
- The capital market line is the line formed when the risky asset is a market portfolio rather than a single risky asset or portfolio. The market portfolio is a mutual fund or exchange-traded fund (based on a market index, for instance).

The introduction of the risk-free asset significantly changes the Markowitz efficient set of portfolios. Investors are better off because they have improved investment opportunities.

This new line leads all investors to invest in *the same risky portfolio*, the market portfolio. That is, all investors make *the same investment decision*. They can, however, attain their desired risk preferences by adjusting the weight of the market portfolio in their portfolios.

- A *strongly risk-averse* investor will *lend* some funds at the *risk-free rate* and invest the remainder in the market portfolio.
- A *less risk-averse* investor will *borrow* some funds at the *risk-free rate* and invest all the funds in the market portfolio.

The Market Portfolio

The market portfolio of risky securities, M, is the highest point of tangency between the line emanating from R_f and the efficient frontier and is the singular optimal risky portfolio. In equilibrium, all risky assets must be in portfolio M because all investors are assumed to arrive at, and hold, the same risky portfolio.

All assets are included in portfolio M in proportion to their market value. For example, if the market for Google stock was 2 percent of the market value of all risky assets, Google would constitute 2 percent of the market value of portfolio M. Therefore, 2 percent of the market value of each investor's portfolio of risky assets would be Google. Think of portfolio M as a broad market index such as the S&P 500 Index. The market portfolio is, of course, a risky portfolio; its risk is designated Ïƒ$_{M}$.

Portfolio M in a Global Context. In theory, the market portfolio (M) should include all risky assets worldwide, both financial and real, in their proper proportions. It has been estimated that the value of non-U.S. assets exceeds 60 percent of the world total. Further, U.S. equities make up only about 10 percent of total world assets. Therefore, international diversification is important.

Portfolio M and Diversification. Because the market portfolio includes all risky assets, portfolio M is by definition completely diversified. The market portfolio is the optimal portfolio of risky assets for investors to own, and therefore will form part of the CML. A portfolio that is completely diversified has a correlation with the market of 1.0.

Differential Borrowing and Lending Rates. Most investors can lend unlimited amounts at the risk-free rate by buying government securities, but they must pay a premium relative to the prime rate when borrowing money. The effect of this differential is that there will be two different lines going to the Markowitz efficient frontier.

- The segment RFR-F indicates the investment opportunities available when an investor combines risk-free assets (lending at RFR) and portfolio F on the Markowitz efficient frontier. It is NOT possible to extend this line any further if it is assumed that you cannot borrow at this risk-free rate to acquire further units of Portfolio F.
- If you can borrow at Rb, you can use the proceeds to invest in portfolio K to extend the CML along the line segment K-G.
- Therefore, the CML is made up of RFR-F-K-G. This implies that you can either lend or borrow, but the borrowing portfolio is not as profitable as when it was assumed that you could borrow at RFR. Your net return is less, as the slope of the borrowing line (K-G) is below that for RFR-F.

2. Pricing of Risk and Computation of Expected Return

```
<b>Systematic Risk and Unsystematic Risk</b><p> </p>
```

Total risk is measured as the standard deviation of security returns. It has two components:

- **Systematic risk** is the risk that is inherent in the market that cannot be diversified away. The systematic risk of an asset is the relevant risk for constructing portfolios. Examples of systematic risk or market risk include macroeconomic factors that affect everything (such as the growth in U.S. GNP, inflation, etc.).

  Note that different securities may respond differently to market changes, and thus may have different systematic risks. For example, automobile manufacturers are much more sensitive to market changes than discount retailers (e.g., Wal-Mart). As a result, automobile manufacturers have higher systematic risk.
- Unique, diversifiable, or **unsystematic risk** (or nonsystematic risk) is risk that can be diversified away. This risk is offset by the unique variability of the other assets in a portfolio. An investor should not expect to receive additional return for assuming unsystematic risk.

Systematic risk is priced, and investors are compensated for holding assets or portfolios based only on that investment's systematic risk. Investors do not receive any return for accepting unsystematic risk.

Return-Generating Models

A **return-generating model** tries to estimate the expected return of a security based on certain parameters. Both the market model and CAPM are single-factor models. The common, single factor is the return on the market portfolio. Multifactor models describe the return on an asset in terms of the risk of the asset with respect to a set of factors. Such models generally include systematic factors, which explain the average returns of a large number of risky assets. Such factors represent priced risk, risk which investors require an additional return for bearing.

According to the type of factors used, there are three categories of multifactor models:

- In **macroeconomic factor models**, the factors are surprises in macroeconomic variables that significantly explain equity returns. Surprise is defined as actual minus forecast value and has an expected value of zero. The factors, such as GDP, interest rates, and inflation, can be understood as affecting either the expected future cash flows of companies or the interest rate used to discount these cash flows back to the present.
- In **fundamental factor models**, the factors are attributes of stocks or companies that are important in explaining cross-sectional differences in stock prices. Among the fundamental factors are book-value-to-price ratio, market cap, P/E ratio, financial leverage, and earnings growth rate.

- In **statistical factor models**, statistical methods are applied to a set of historical returns to determine portfolios that explain historical returns in one of two senses. In factor analysis models, the factors are the portfolios that best explain (reproduce) historical return covariances. In principal-components models, the factors are portfolios that best explain (reproduce) the historical return variances.

Here is a two-factor macroeconomic model.

$$R_i = a_i + b_{i1}\ F_{GDP} + b_{i2}\ F_{INT} + \text{Îµ}_i$$

where

- R_i = the return for asset i.
- a_i = expected return for asset i in the absence of any surprises.
- b_{i1} = GDP surprise sensitivity of asset i. This is a slope coefficient which is interpreted as the GDP **factor sensitivity** of asset i.
- F_{GDP} = surprise in GDP growth. This is the GDP **factor surprise**, the difference between the expected value and the actual value of the GDP.
- b_{i2} = interest rate surprise sensitivity of asset i. This is the interest rate factor sensitivity of asset i.
- F_{INT} = surprise in interest rates. This is the interest rate factor surprise.
- Îµ_i = firm-specific surprises (the portion of the return to asset i not explained by the factor model).

The model says stock returns are explained by surprises in GDP growth and interest rates. The regression analysis is usually used to estimate assets' sensitivities to these factors.

Calculation and Interpretation of Beta

Beta (Î²) is the standardized measure of systematic risk.

Since all investors want to hold the market portfolio, a security's covariance with the market portfolio ($Cov_{i,M}$) is the appropriate risk measure. $Cov_{i,M}$ is an absolute measure of the security's systematic risk. Its magnitude is affected by the variability of both the security and the market portfolio (recall that $Cov_{i,j} = p_{i,j}$ x $\text{Ïƒ}_{i,j}$ x $\text{Ïƒ}_{i,j}$). To standardize the measure of systematic risk, divide $Cov_{i,M}$ by the covariance of the market portfolio with itself ($Cov_{M,M}$). Therefore, the standardized measure of systematic risk (beta) is defined as Î² = $Cov_{i,M}$ / $Cov_{M,M} = Cov_{i,M} / \text{Ïƒ}_M^2 = \text{Ï}_{i,M}\ \text{Ïƒ}_i/\text{Ïƒ}_M$.

- The market portfolio has a Î² of 1.
- If Î² > 1, the security is more volatile than the market.
- If Î² < 1, the security is less volatile than the market.

3. The Capital Asset Pricing Model

```
<b>Assumptions of the CAPM</b><p> </p>
```

The assumptions of the CAPM include:

- All investors are *Markowitz efficient investors* who want to target points on the efficient frontier where their utility maps are tangent to the line. The exact location on the efficient frontier and, therefore, the specific portfolio selected, will depend on the individual investor's risk-return utility function.

- Markets are frictionless. There are no taxes or transaction costs involved in buying or selling assets. Investors can borrow and lend any amount of money at the risk-free rate of return.

- All investors have the same one-period time horizon (e.g., one year).

- All investors have homogeneous expectations: that is, they estimate identical probability distributions for future rates of return.

- All investments are infinitely divisible, which means that it is possible to buy or sell fractional shares of any asset or portfolio.
- All investors are price takers. Their trades cannot affect security prices.

The CAPM

Capital market theory builds on portfolio theory. CAPM refers to the **capital asset pricing model**. It is used to determine the required rate of return for any risky asset.

In the discussion about the Markowitz efficient frontier, the assumptions are:

- Investors have examined the set of risky assets and identified the efficient frontier.
- Every investor will choose the optimal portfolio of risky assets on the efficient frontier. The optimal portfolio lies at the point where the highest indifference curve is tangent to the efficient frontier.

The CAPM uses the SML or **security market line** to compare the relationship between risk and return. Unlike the CML, which uses standard deviation as a risk measure on the X axis, the SML uses the market beta, or the relationship between a security and the marketplace.

The use of beta enables an investor to compare the relationship between a single security and the market return rather than a single security with each and every other security (as Markowitz did). Consequently, the risk added to a market portfolio (or a fully diversified set of securities) should be reflected in the security's beta. The expected return for a security in a fully diversified portfolio should be:

E(R_M) - R_f is the **market risk premium**, while the **risk premium of the security** is calculated by Î²[E(R_M) - R_f].

Note that the "expected" and the "required" returns mean the same thing. The expected return based on the CAPM is exactly the return an investor requires on the security.

- To compute the required rate of return: .
- To compute the expected rate of return of an individual security, you need to use forecasted future security price and dividend: R = (Future price - current price + dividend) / Current price.

The SML represents the required rate of return, given the systematic risk provided by the security. If the expected rate of return exceeds this amount, then the security provides an investment opportunity for the investor. The difference between the expected and required return is called the **alpha** (Î±) or **excess rate of return**. The alpha can be positive when a stock is undervalued (it lies above the SML) or negative when the stock is overvalued (it falls below the SML). The alpha becomes zero when the stock falls directly on the SML (properly valued).

Security Market Line vs. Capital Market Line:

- The CML examines the expected returns on *efficient portfolios* and their *total risk* (measured by standard deviation). The SML examines the expected returns on *individual assets* and their *systematic risk* (measured by beta). If the relationship between expected return and beta is valid for any individual securities, it must also be valid for portfolios constructed with any of these securities. So, the SML is valid for *both* efficient portfolios *and* individual assets.
- The CML is the graph of the efficient frontier and the SML is the graph of the CAPM.
- The slope of the CML is the market portfolio's Sharpe ratio and the slope of the SML is the market risk premium.
- All properly priced securities and efficient portfolios lie on the SML. However, only efficient portfolios lie on the CML.

Portfolio Beta. The Î² of a portfolio is the weighted sum of the individual asset betas. For example, if 40% of the money is in stock A with a Î² of 2.0 and 60% of the money is in stock B with a Î² of 0.8, the portfolio Î² is 0.4 x 2.0 + 0.6 x 0.8 = 1.28.

4. Applications of the CAPM

```
    <b>Estimate of Expected Return.</b> Apply the CAPM formula to calculate the expected r
```

Portfolio Performance Evaluation. Four ratios are commonly used for this purpose.

- **Sharpe Ratio**

  It is a measure of the excess return per unit of risk. It defined as the portfolio's risk premium divided by its risk: . This ratio is easy to use. The two limitations are:

  - It uses standard deviation, not beta, as a measure of volatility.
  - The ratio itself is not very informative.

- **Treynor Ratio**

  It measures the excess return on an investment which has no diversifiable risk. Systematic risk is used instead of total risk: . Again, the ratio itself is not very informative, and it cannot be applied to assets with negative Î²s. It is a ranking criterion which is easy to use.

- **M-Squared (M^2)**

  It is a performance measurement using return per unit of total risk as measured by the standard deviation. The investment portfolio's standard deviation is adjusted to reflect the standard deviation of the market benchmark portfolio. The return premiums of the adjusted investment portfolio and the market index portfolio are then compared.

- **Jensen's Alpha**

  It is the abnormal return over the theoretical expected return. The theoretical expected return is calculated using CAPM (and beta): Î± $_p$ = R_p - [R_f+ Î²(R_m - R_f)]. Since the CAPM return is supposed to be risk-adjusted, Jensen's Î± is also risk-adjusted. Investors are constantly seeking investments that have positive Î± or "abnormal returns."

If an investor holds a portfolio that is not fully diversified, total risk matters. Sharpe ratio and M-squared are appropriate performance measures in such cases. On the other hand, if the portfolio is well-diversified, Treynor ratio and Jensen's alpha are relevant, as only the systematic risk of the portfolio matters.

Security Characteristic Line. A security characteristic line (SCL) graphs the relationship between the excess market return and excess security return.

R_i - R_f= Î±$_i$ - Î²$_i$ (R_m - R_f)

If we compare the SML and the SCL:

While there is only one SML, there are many different SCLs for securities with different betas.

Security Selection. Overvalued and undervalued securities are those securities that do not lie on the SML line. By definition, securities that are efficiently priced should fall directly on the (calculated) SML line. If a security is above the line it is deemed undervalued since it is providing more expected return than what is demanded for that risk level. Securities falling below the SML line, on the other hand, provides less return than the market demands. Securities that fall below the SML are considered overvalued. In the former case, the security price will be bid up, such that the expected return declines and the security falls back to the SML line.

In a situation where the security is overvalued, the security price declines until the expected return rises.

All assets and all portfolios should plot on the SML.

- Stock C has an estimated rate of return equal to its systematic risk or required rate of return.
- Stock B is expected to provide a rate of return above the required rate of return.
- Stock A is expected to provide a rate of return below the required rate of return.
- Investors should buy B (undervalued).
- Investors should sell A (overvalued).

Constructing a Portfolio

As the number of securities increases, the portfolio manager can eliminate unsystematic risk (or **diversifiable risk**) and focus on systematic or undiversifiable risk. As you can see, much of the unsystematic risk can be diversified away in as few as 30 securities.

To construct a market portfolio you can start with a portfolio of securities like the S&P 500. You can then evaluate the Î± of any security, using the CAPM and the S&P 500 as the market portfolio. If the Î± is positive, add the security to the portfolio. If a security's Î± is negative and it is already included in the portfolio, remove it. You can also use the information ratio of each individual security to determine the relative weight of the security in the portfolio.