

2026
CEA[®]
Exam Prep

Schweser's
Secret Sauce[®]

Level I

KAPLAN) SCHWESER

Schweser's Secret Sauce[®]

Level I CFA[®]

2026

KAPLAN  **SCHWESER**

This book will be a valuable addition to the study tools of any CFA™ exam candidate. It offers a concise and readable explanation of the major parts of the Level I CFA curriculum. Here is the disclaimer: this book does not cover every Learning Outcome Statement (LOS) and, as you are aware, any LOS is "fair game" for the exam. We have tried to include those LOS that are key concepts in finance and accounting, have application to other LOS, are complex and difficult for candidates, require memorization of characteristics or relationships, or are a prelude to LOS at Levels II and III.

We suggest you use this book as a companion to your other, more comprehensive study materials. It is easier to carry with you and will allow you to study these key concepts, definitions, and techniques over and over, which is an important part of mastering the material. When you get to topics where the coverage here appears too brief or raises questions in your mind, this is your clue to go back to your SchweserNotes™ or the textbooks to fill in the gaps in your understanding. There is no shortcut to learning the broad array of subjects covered by the Level I curriculum, but this volume should be a valuable tool for reviewing and retaining the material as you complete your preparations in the weeks leading up to your exam day.

Pass rates for most Level I exams in recent years have been in the 40% to 45% range, and returning candidates make comments such as, "I was surprised at how difficult the exam was." You should not despair because of this, but you should definitely not underestimate the task at hand. Our study materials are all designed to help you study as efficiently as possible, grasp and retain the material, and apply it with confidence come exam day.

Best regards,

Craig S. Prochaska, CFA

Senior Manager, Content, Advanced Designations

Kaplan Schweser

Topic Focus: Quantitative Methods

SECRET SAUCE

Weight on Exam	6% to 9%
SchweserNotes™ Reference	Book 1, Pages 1–139

Understanding time value of money (TVM) computations is essential for success, not only for quantitative methods, but also other sections of the Level I exam. Candidates who are unfamiliar with TVM calculations, or simply need a refresher, should review them in the CFA Institute prerequisite material for Quantitative Methods. TVM is actually a larger portion of the exam than just quantitative methods because of its integration with other topics. Any portion of the exam that requires discounting cash flows will require TVM calculations. Examples include evaluating capital projects, valuing bonds, and using dividend discount models to value equities. No matter where it shows up on the exam, the key to any TVM problem is to draw a timeline and be certain of when the cash flows will occur so you can discount those cash flows appropriately.

Reading 1: Rates and Returns

SECRET SAUCE

Interest rates measure the time value of money. Equilibrium interest rates are the **required rate of return** for a particular investment. Interest rates are also referred to as **discount rates**. We can also view interest rates as the opportunity cost of current consumption.

The **real risk-free rate** is a theoretical interest rate on a single-period loan with no risk of inflation or default. The real risk-free rate represents **time preference**, the degree to which current consumption is preferred to equal future consumption.

A **nominal risk-free rate** contains an inflation premium:

$$(1 + \text{nominal risk-free rate}) = (1 + \text{real risk-free rate})(1 + \text{expected inflation rate})$$

This relation is approximated as:

$$\text{nominal risk-free rate} \approx \text{real risk-free rate} + \text{expected inflation rate}$$

Risk increases the required rate of return. Types of risks include **default risk**, that a borrower will not make the promised payments on time; **liquidity risk** of receiving less than fair value if an investment must be sold quickly; and **maturity risk**, because the prices of longer-term bonds are more volatile than those of shorter-term bonds. Thus, we can state:

$$\begin{aligned} \text{nominal interest rate} \approx & \text{real risk-free rate} \\ & + \text{inflation premium} \\ & + \text{default risk premium} \\ & + \text{liquidity risk premium} \\ & + \text{maturity risk premium} \end{aligned}$$

Return Measures

A **holding period return (HPR)** is the percentage increase in the value of an investment over a given period.

$$\text{holding period return} = \frac{\text{end-of-period value}}{\text{beginning-of-period value}} - 1$$

To **annualize** an HPR that is realized over a specific number of days, use the following formula:

$$\text{annualized return} = (1 + \text{HPR})^{365/\text{days}} - 1$$

The **arithmetic mean return** is the simple average of a series of periodic returns. The **geometric mean return** is a compound rate:

$$\text{geometric mean return} = \sqrt[n]{(1 + R_1) \times (1 + R_2) \times \dots \times (1 + R_n)} - 1$$

A **harmonic mean return** is used for certain computations, such as the average cost of shares purchased over time. Some refer to this practice as **cost averaging**.

Returns data may include **outliers**, extreme values that distort mean return measures. A **trimmed mean** and a **winsorized mean** are techniques for dealing with outliers. A trimmed mean excludes a stated percentage of the most extreme observations. A 1% trimmed mean, for example, would discard the lowest 0.5% and the highest 0.5% of the observations. To compute a winsorized mean, instead of discarding the highest and lowest observations, we substitute a value for them. To calculate a 90% winsorized mean, for example, we would determine the 5th and 95th percentile of the observations, substitute the 5th percentile for any values lower than that, substitute the 95th percentile for any values higher than that, and then calculate the mean of the revised dataset.

A **money-weighted return** is the internal rate of return on a portfolio that includes all cash inflows and outflows. The beginning value of the account and all deposits into the account are inflows. All withdrawals from the account and its ending value are outflows.

A **time-weighted return** measures compound growth over a specified time horizon. In the investment management industry, time-weighted return is the preferred method of performance measurement because portfolio managers typically do not control the timing of deposits to and withdrawals from the accounts they manage. The annual time-weighted return for an investment may be computed as follows:

Step 1: Form subperiods that correspond to the dates of deposits and withdrawals.

Step 2: Compute an HPR of the portfolio for each subperiod.

Step 3: Compute the product of $(1 + \text{HPR})$ for each subperiod to obtain a total return for the entire measurement period, and annualize the result.

If funds are contributed to an investment portfolio just before a period of relatively poor portfolio performance, the money-weighted rate of return will tend to be lower than the time-weighted rate of return. On the other hand, if funds are contributed to a portfolio at a favorable time (just before a period of relatively high returns), the money-weighted rate of return will be higher than the time-weighted rate of return.

Gross return refers to the total return on a security portfolio before deducting fees for management and administration. **Net return** refers to the return after these fees have been deducted.

Pretax nominal return refers to the return before paying taxes. **After-tax nominal return** refers to the return after the tax liability is deducted.

Real return is nominal return adjusted for inflation. Consider a nominal return of 7% in a year when inflation is 2%. The approximate real return is $7\% - 2\% = 5\%$. The exact real return is slightly lower: $1.07 / 1.02 - 1 = 4.9\%$.

Leveraged return is a gain or loss as a percentage of an investor's cash investment. For some investments, such as derivatives or real estate, the cash invested is only a fraction of the value of the underlying assets. Leverage magnifies both gains and losses on an investment.

Compounding Periods

More frequent compounding increases the effective return on an investment, increasing the future value of a given cash flow and decreasing the present value of a given cash flow. The general formula for present value given a compounding frequency is as follows:

$$PV = FV_N \left(1 + \frac{r}{m}\right)^{-mN}$$

where:

r = quoted annual interest rate

N = number of years

m = compounding periods per year

The mathematical limit of shortening the compounding period is **continuous compounding**. We use the natural logarithm (\ln) to state a continuously compounded rate of return:

$$R_{CC} = \ln(1 + HPR)$$

Reading 2: Time Value of Money in Finance

SECRET SAUCE

With discrete compounding periods, the relationship between present values and future values is:

$$FV = PV(1 + r)^t \text{ or } PV = FV(1 + r)^{-t}$$

where:

r = interest rate per compounding period

t = number of compounding periods

With continuous compounding, the relationship is:

$$FV = PV \times e^{rt} \text{ or } PV = FV \times e^{-rt}$$

Common applications of TVM in finance include valuing:

- A single future cash flow, such as a zero-coupon bond.
- A finite series of future cash flows, such as a coupon bond.
- An infinite series of equal future cash flows, such as a perpetuity or a preferred stock.
- A growing series of future cash flows, such as a common stock with a constant growth rate of dividends.

A **zero-coupon bond** pays a single cash flow equal to its face value at maturity. Assuming a positive yield, the investor pays less than the face value to buy the bond.

It is possible for the yield to be negative. In this case an investor would pay more than the face value to buy a zero-coupon bond.

A **fixed-coupon bond** is only slightly more complex. The investor receives a cash interest payment each period and the face value (along with the final interest payment) at maturity. The bond's **coupon rate** is a percentage of the face value and determines the amount of the interest payments. *The coupon rate and the yield are two different things.* We only use the coupon rate to determine the coupon payment (PMT). The yield to maturity (I/Y) is the discount rate implied by the bond's price.

The relationship between prices and yields is inverse. When the price decreases, the yield to maturity increases. When the price increases, the yield to maturity decreases. Or, equivalently, when the yield increases, the price decreases. When the yield decreases, the price increases.

An **amortizing bond** is one that pays a level amount each period, including its maturity period. We can determine an annuity payment using a financial calculator.

A bond with no maturity date is a **perpetuity**. The present value of a perpetuity is simply the payment divided by the yield (as a decimal).

A **preferred stock** is similar to a perpetuity and we calculate its present value the same way:

$$\text{preferred stock value} = \frac{\text{dividend per period}}{\text{the market's required return}}$$

Common stock typically does not promise a fixed dividend payment. Instead, the company's management decides whether and when to pay common dividends. Because the future cash flows are uncertain, we must use models to estimate a common stock's value.

- If we assume a *constant future dividend*, we can value a common stock the same way we value a preferred stock.
- If we assume a *constant growth rate of dividends*, we can use a model called the constant growth dividend discount model or the Gordon growth model. We will illustrate this model later in the Equity Investments section of this book.
- If we assume a *changing growth rate of dividends*, we must choose an appropriate model. One example (which we will also demonstrate in the Equity Investments section) is a multistage dividend discount model, in which we assume a period of rapid growth followed by a period of constant growth.

The **cash flow additivity principle** states that the PV of any stream of cash flows equals the sum of the PVs of the cash flows. If we have two series of cash flows, the sum of the PVs of the two series is the same as the PVs of the two series taken together. We can also divide up a series of cash flows any way we like, and the PV of the "pieces" will equal the PV of the original series.

The cash flow additivity principle is the basis for the **no-arbitrage principle**, or "law of one price," which says that if two sets of future cash flows are identical under all conditions, they will have the same price today. (If they don't, investors will quickly buy the lower-priced one and sell the higher-priced one, which will drive their prices together). We apply the no-arbitrage condition when we calculate forward exchange rates or forward interest rates, and when we value options using a binomial model.

Measures of Central Tendency

The **arithmetic mean** is the sum of the observation values divided by the number of observations. It is the most widely used measure of central tendency. An example of an arithmetic mean is a **sample mean**, which is the sum of all the values in a sample divided by the number of observations in the sample.

The **median** is the midpoint of a dataset, where the data are arranged in ascending or descending order. Half of the observations lie above the median, and half are below.

The **mode** is the value that occurs most frequently in a dataset. A dataset may have more than one mode, or even no mode. When a distribution has one value that appears most frequently, it is said to be **unimodal**. For continuous data, such as investment returns, we typically do not identify a single outcome as the mode. Instead we divide the relevant range of outcomes into intervals and identify the **modal interval** as the one into which the largest number of observations fall.

As mentioned earlier, we can use a *trimmed mean* or a *winsorized mean* to lessen the effects of outliers.

Measures of Location and Dispersion

Examples of **quantiles** include the following:

- **Quartile.** A distribution is divided into quarters. The difference between the third quartile and the first quartile is known as the **interquartile range**.
- **Quintile.** A distribution is divided into fifths.
- **Decile.** A distribution is divided into tenths.
- **Percentile.** A distribution is divided into hundredths (percentages). Any quantile can be expressed as a percentile. For example, the third quartile is the 75th percentile.

Dispersion is variability around the central tendency. In finance and investments, central tendency measures expected return and dispersion measures risk.

The **range** of a dataset is the difference between its largest and its smallest value.

The **mean absolute deviation (MAD)** is the average of the absolute values of the deviations of individual observations from the arithmetic mean. It uses absolute values because actual deviations from the arithmetic mean always sum to zero.

The **sample variance** is the measure of dispersion that applies when we are evaluating a sample of n observations from a population. Sample variance is calculated using the following formula:

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

The computed variance is in squared units of measurement. We can address this problem by using the **standard deviation**, which is the square root of the variance. The units of standard deviation are the same as the units of the data:

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Relative dispersion is the amount of variability in a distribution around a reference point or benchmark. Relative dispersion is commonly measured with the **coefficient of variation (CV)**, which measures the amount of dispersion in a distribution relative to its mean:

$$CV = \frac{\text{standard deviation}}{\text{arithmetic mean}}$$

In an investments setting, the CV is used to measure the risk (variability) per unit of expected return (mean). A lower CV is better.

In some situations, it may be more appropriate to consider only **downside risk**, or outcomes less than the mean (or some other specific value). One measure of downside risk is **target downside deviation**, which is also known as **target semideviation**:

$$s_{\text{target}} = \sqrt{\frac{\sum_{\text{all } X_i < B} (X_i - B)^2}{n-1}}$$

where B = the target

Skewness, or skew, refers to the extent to which a distribution is not symmetrical. Nonsymmetrical distributions may be either positively or negatively skewed and result from the occurrence of outliers in the dataset. A *positively skewed* distribution has outliers greater than the mean (in the upper region, or right tail) and is said to be skewed right. A *negatively skewed* distribution has outliers less than the mean that fall within its lower (left) tail and is said to be skewed left.

Skewness affects the location of the mean, median, and mode:

- For a symmetrical distribution, the mean, median, and mode are equal.
- For a positively skewed distribution, the mode is less than the median, which is less than the mean.
- For a negatively skewed distribution, the mean is less than the median, which is less than the mode.

Kurtosis is the degree to which a distribution is more or less peaked than a normal distribution. **Leptokurtic** describes a distribution that is more peaked than a normal distribution, whereas **platykurtic** refers to a distribution that is less peaked, or flatter than a normal one. A distribution is **mesokurtic** if it has the same kurtosis as a normal distribution.

A distribution is said to exhibit **excess kurtosis** if it has either more or less kurtosis than the normal distribution. The computed kurtosis for all normal distributions is three. Statisticians, however, sometimes report excess kurtosis, which is defined as kurtosis minus three. Thus, a normal distribution has excess kurtosis equal to zero, a leptokurtic distribution has excess kurtosis greater than zero, and platykurtic distributions have excess kurtosis less than zero.

Securities returns tend to exhibit both skewness and kurtosis. If returns are modeled using an assumed normal distribution, the predictions will not take into account the potential for extremely large, negative outcomes. Most risk managers put little emphasis on the mean and standard deviation and focus more on returns in the tails of the distribution—that is where the risk is.

Covariance measures how two variables move together. Sample covariance is calculated as follows:

$$s_{XY} = \frac{\sum_{i=1}^n \{ [X_i - \bar{X}] [Y_i - \bar{Y}] \}}{n-1}$$

where:

X_i = an observation of variable X

Y_i = an observation of variable Y

\bar{X} = mean of variable X

\bar{Y} = mean of variable Y

n = number of periods

Like the variance, the units of covariance are the square of the units used for the data. The value of covariance depends on the units of the variables, and we cannot use it to interpret the relative strength of the relationship. To address these issues, we use a standardized measure of the linear relationship between two variables

called the **correlation coefficient**. The correlation between two variables, X and Y, is:

$$\rho_{XY} = \frac{S_{XY}}{S_X S_Y}$$

Correlation has no units. Its values range from -1 to $+1$. If correlation is $+1$ (perfect positive correlation), a change in one random variable results in a proportional change in the other. If correlation is -1 (perfect negative correlation), a change in one random variable results in a proportional but opposite change in the other. If correlation is zero, the two variables have no linear relationship.

Two variables might have a *nonlinear* relationship that the correlation coefficient does not reflect. Analysts often use **scatter plots**, with one variable on the vertical axis and the other on the horizontal axis, to reveal nonlinear relationships.

Variables can exhibit **spurious correlation** that results from chance or from their association with a third variable.

Reading 4: Probability Trees and Conditional Expectations

SECRET SAUCE

The **expected value** of a random variable is the weighted average of the possible outcomes for the variable:

$$E(X) = \sum P(x_i)x_i = P(x_1)x_1 + P(x_2)x_2 + \dots + P(x_n)x_n$$

where:

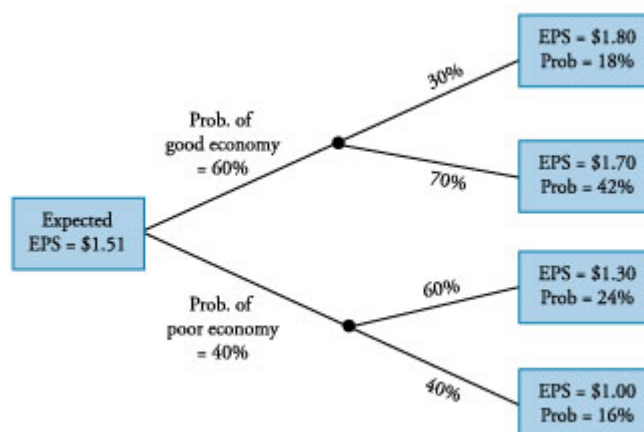
$$P(x_i) = \text{probability of outcome } x_i$$

Variance (from a probability model) can be calculated as the probability-weighted sum of the squared deviations from the expected value. The standard deviation is the positive square root of the variance.

When we estimated standard deviation from sample data, we divided the sum of the squared deviations from the mean by the sample size minus one. With a probability model, by contrast, we have no past observations (and therefore no "n - 1" to divide by). We use probability weights instead, as they describe the entire distribution of outcomes.

A general framework called a **probability tree** is used to show the probabilities of various outcomes. The following **A Probability Tree for an Investment Problem** shows an example.

A Probability Tree for an Investment Problem



Conditional expected values depend on the outcome of some other event. An analyst can use a conditional expected value to revise his expectations when new information arrives.

Bayes' formula is used to update a given set of prior probabilities for a given event in response to the arrival of new information. The rule for updating prior

probability of an event is as follows:

$$\text{updated probability} = \frac{\text{probability of new information for a given event}}{\text{unconditional probability of new information}} \times \text{prior probability of event}$$

Reading 5: Portfolio Mathematics

SECRET SAUCE

The **expected return and variance for a portfolio** of assets can be determined using the expected returns and weights of the individual assets in the portfolio:

$$E(R_P) = w_A R_A + w_B R_B$$

$$\text{Var}_P = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2w_A w_B \sigma_A \sigma_B \rho_{A,B}$$

or:

$$\text{Var}_P = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2w_A w_B \text{COV}_{A,B}$$

We can calculate **covariance of returns** two assets in a portfolio as follows:

$$\text{COV}_{A,B} = E\{[R_A - E(R_A)][R_B - E(R_B)]\}$$

Shortfall risk is the probability that a portfolio's return or value will be below a specified target return or value over a specified period. **Roy's safety-first criterion** states that the optimal portfolio minimizes the probability that the return of the portfolio falls below some minimum acceptable "threshold" level.

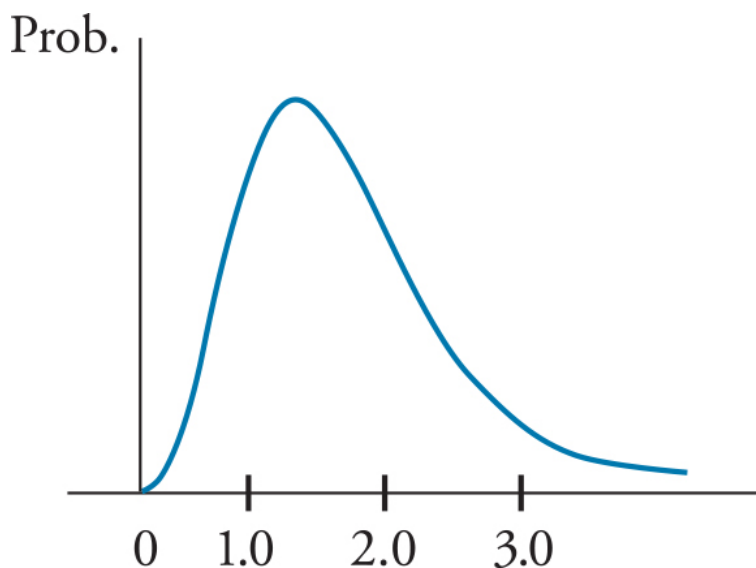
Roy's **safety-first ratio (SFRatio)** shows how many standard deviations the expected return is above the threshold return (R_L):

$$\text{SFRatio} = \frac{E(R_p) - R_L}{\sigma_p}$$

A portfolio with a higher SFRatio is preferred. The greater the SFRatio, the lower the probability that returns will be below the threshold return, and therefore the lower the shortfall risk.

If x is normally distributed, $Y = e^x$ is lognormally distributed. The **lognormal distribution** is positively skewed as shown in the following **Lognormal Distribution**.

Lognormal Distribution



The lognormal distribution is useful for modeling asset prices if we think of an asset's future price as the result of a continuously compounded return on its current price. That is:

$$P_T = P_0 e^{r_{0,T}}$$

where:

P_T = asset price at time T

P_0 = asset price at time 0 (today)

$r_{0,T}$ = continuously compounded return on the asset from time 0 to time T

In many pricing models, we assume returns are **independently and identically distributed**.

- If returns are *independently* distributed, past returns are not useful for predicting future returns.
- If returns are *identically* distributed, they exhibit "stationarity," in that their mean and variance do not change over time.

Monte Carlo simulation is performed by making assumptions about the distributions of prices or risk factors and using a large number of computer-generated random values for the relevant risk factors or prices to generate a distribution of possible outcomes (e.g., project NPVs, portfolio values). An advantage of Monte Carlo simulation is that its inputs are not limited to the range of historical data. Its limitations are that it is fairly complex, can provide answers that are no better than the assumptions used, and is a statistical method that cannot offer the insights provided by an analytic method.

Resampling is another method for generating data inputs to use in a simulation. To conduct resampling, we start with the observed sample and repeatedly draw subsamples from it, each with the same number of observations. From these samples, we can infer parameters for the population, such as its mean and variance.

In **bootstrap resampling**, we draw repeated samples from the full dataset, replacing the sampled observations each time so that they might be redrawn in another sample. We can then directly calculate the standard deviation of these sample means as our estimate of the standard error of the sample mean. Simulation using data from bootstrap resampling follows the same procedure as Monte Carlo simulation. The difference is the source and scope of the data. For example, if a simulation uses bootstrap resampling of historical returns data, its inputs are limited by the distribution of actual outcomes.

Probability sampling refers to selecting a sample when we know the probability of each sample member in the overall population.

- With **simple random sampling**, each item or person in the population has the same likelihood of being included in the sample.
- Another way to form an approximately random sample is **systematic sampling**, selecting every n th member from a population.
- **Stratified random sampling** separates a population into groups based on one or more characteristics, then takes a random sample from each group based on its size. For example, in constructing bond index portfolios, we may first divide the bonds into groups by maturity, rating, and embedded options, then pick bonds from each group in proportion to the number of index bonds in that group. This ensures that our "random" sample has similar maturity, rating, and call characteristics to the index.
- **Cluster sampling** is also based on subsets of a population, but assumes that each subset (cluster) is representative of the overall population. In **one-stage cluster sampling**, a random sample of clusters is selected and all the data in those clusters comprise the sample. In **two-stage cluster sampling**, random samples from each of the selected clusters comprise the sample. To the extent that the subgroups do not have the same distribution as the entire population of the characteristic we are interested in, cluster sampling will have greater sampling error than simple random sampling and two-stage cluster sampling can be expected to have greater sampling error than one-stage cluster sampling.

Non-probability sampling is based on either low cost or easy access to some data items, or on using the judgment of the researcher in selecting specific data items. Less randomness in selection may lead to greater sampling error.

- **Convenience sampling** refers to selecting sample data based on its ease of access, using data that are readily available.
- **Judgmental sampling** refers to samples for which each observation is selected from a larger data set by the researcher, based on her experience and judgment.

Central Limit Theorem

The **central limit theorem** of statistics states that in selecting simple random samples of size n from a population with a mean μ and a finite variance σ^2 , the sampling distribution of the sample mean approaches a normal probability

distribution with mean μ and a variance equal to σ^2/n as the sample size becomes sufficiently large ($n \geq 30$).

The **standard error of the sample mean** is the standard deviation of the sample divided by the square root of n , the number of observations in the sample. As the sample size increases, the sample mean gets closer, on average, to the true mean of the population.

There are alternatives to calculating the standard error of the sample mean:

- With the **jackknife method**, we calculate multiple sample means, each with one of the observations removed from the sample. The standard deviation of these sample means can then be used as an estimate of the standard error of sample means.
- With the more computationally demanding **bootstrap method**, we draw repeated samples of size n from the full data set (replacing the sampled observations each time) and then calculate the standard deviation of these means. The bootstrap method can be used to estimate the distributions of complex statistics, including those that do not have an analytic form.

Reading 8: Hypothesis Testing

SECRET SAUCE

A **hypothesis** is a statement about a population parameter that is to be tested. For example, "The mean return on the S&P 500 Index is equal to zero."

Steps in Hypothesis Testing

The steps in hypothesis testing are as follows:

- Step 1:* State the hypothesis.
- Step 2:* Select a test statistic.
- Step 3:* Specify the level of significance.
- Step 4:* State the decision rule for the hypothesis.
- Step 5:* Collect the sample and calculate statistics.
- Step 6:* Make a decision about the hypothesis.
- Step 7:* Make a decision based on the test results.

Null and Alternative Hypotheses

The **null hypothesis**, designated as H_0 , is the hypothesis the researcher wants to reject. It is the hypothesis that is actually tested and is the basis for the selection of the test statistics. Thus, if you seek to show that the mean return on the S&P 500 Index is different from zero, the null hypothesis will be that the mean return on the index *equals* zero.

The **alternative hypothesis**, designated H_a , is what is concluded if there is sufficient evidence to reject the null hypothesis. It is usually the alternative hypothesis you are really trying to support. Why? Since you can never really prove anything with statistics, when the null hypothesis is rejected, the implication is that the (mutually exclusive) alternative hypothesis is valid.

A **test statistic** is calculated from sample data and is compared to a critical value to evaluate H_0 . The most common test statistics are the z-statistic and the t-statistic. **Critical values** come from tables and are based on the researcher's desired level of significance. If the test statistic exceeds the critical value (or is outside the range of critical values), the researcher rejects H_0 . As the level of significance gets smaller, the critical value gets larger and it becomes more difficult to reject the null hypothesis. The **p-value** of a test statistic is the lowest significance value that will result in rejection of the null hypothesis.

Type I and Type II Errors

When testing a hypothesis, there are two possible types of errors:

- **Type I error.** Rejection of the null hypothesis when it is actually true.
- **Type II error.** Failure to reject the null hypothesis when it is actually false.

The **power of a test** is $1 - P(\text{Type II error})$. The more likely that a test will reject a false null, the more powerful the test. A test that is unlikely to reject a false null hypothesis has little power.

The significance level is the probability of making a Type I error (rejecting the null when it is true) and is designated by the Greek letter alpha (α). You can think of this as the probability that the test statistic will exceed or fall below the critical values by chance even though the null hypothesis is true. A significance level of 5% ($\alpha = 0.05$) means there is a 5% chance of rejecting a true null hypothesis.

Type I and Type II Errors in Hypothesis Testing

Decision	True Condition	
	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Incorrect decision Type II error
Reject H_0	Incorrect decision Type I error Significance level, α , = $P(\text{Type I error})$	Correct decision Power of the test = $1 - P(\text{Type II error})$

Types of Hypothesis Tests

- For a hypothesis concerning the *value of a population mean*, we use a *t*-test (or a *z*-test if the sample size is large enough).
- To test a hypothesis concerning the *equality of two population means*, we use a *t*-test. The nature of that test depends on whether the samples are independent (a difference in means test) or dependent (a paired comparisons test).
- For a hypothesis concerning the *value of a population variance*, we use a chi-square test.
- To test a hypothesis concerning the *equality of two population variances*, we use an *F*-test.

The following table summarizes the test statistics used for each type of hypothesis test.

Types of Test Statistics

Hypothesis tests of:	Use a:
One population mean	t -statistic
Two population means	t -statistic
One population variance	Chi-square statistic
Two population variances	F -statistic
Correlation	t -statistic
Independence	Chi-square statistic

Reading 9: Parametric and Non-Parametric Tests of Independence

SECRET SAUCE

Parametric tests, like the *t*-test, *F*-test, and chi-square test, make assumptions regarding the distribution of the population from which samples are drawn. **Non-parametric tests** either do not consider a particular population parameter or have few assumptions about the sampled population.

A parametric test of whether two random variables are independent uses a hypothesis that their correlation coefficient equals zero. Its test statistic follows a *t*-distribution with $n - 2$ degrees of freedom and increases with the sample size as well as with the sample correlation coefficient:

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where:

r = sample correlation

n = sample size

The **Spearman rank correlation test**, a nonparametric test, can be used to test whether two sets of ranks are correlated. Ranks are simply ordered values.

A **contingency table** is a two-dimensional array with rows that represent attributes of one of the variables and with columns that represent attributes of the other variable. The values in each cell are the frequencies with which we observe two attributes simultaneously. A hypothesis test of whether two characteristics of a sample are independent uses a Chi-square statistic calculated from a contingency table. The test compares the actual table values to what the values would be if the two characteristics were independent.

Reading 10: Simple Linear Regression

SECRET SAUCE

Simple linear regression is used to estimate the linear relationship between two variables and evaluate the significance of the relationship. A researcher determines which variable likely explains the variation in the other.

The **dependent variable** (the Y variable) is also referred to as the explained variable, the endogenous variable, or the predicted variable.

The **independent variable** (the X variable) is also referred to as the explanatory variable, the exogenous variable, or the predicting variable.

The following **linear regression model** is used to describe the relationship between variables X (independent) and Y (dependent):

$$Y_i = b_0 + b_1X_i + \varepsilon_i, \text{ with } i = 1 \text{ to } n$$

where:

Y_i = i th observation of the dependent variable, Y

X_i = i th observation of the independent variable, X

b_0 = regression intercept term

b_1 = regression slope coefficient

ε_i = **residual** for the i th observation (also referred to as the error term)

Based on this regression model, the regression process estimates an equation for a line through a scatter plot of the data that best explains the observed values for Y in terms of the observed values for X . This **regression line** takes the following form:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1X_i, \text{ for } i = 1 \text{ to } n$$

where:

\hat{Y}_i = estimated value of Y_i given X_i

\hat{b}_0 = estimated intercept term

\hat{b}_1 = estimated **slope coefficient**

The regression line is the line through the scatter plot of X and Y that minimizes the sum of the squared errors [differences between the Y -values predicted by the regression equation \hat{Y}_i and the observed Y -values (Y_i)]. The sum of these squared differences is called the **sum of squared errors (SSE)**.

The slope coefficient for the regression line is an estimate of the change in Y for a one-unit change in X . The slope term is calculated as follows:

$$\hat{b}_1 = \frac{\text{Cov}_{XY}}{\sigma^2_X}$$

The **intercept term** is the regression line's intersection with the Y -axis at $X = 0$. The intercept term may be expressed as follows

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$$

In other words, the regression line passes through a point with coordinates equal to the means of the independent and dependent variables.

Assumptions of Simple Linear Regression

Linear regression assumes the following:

1. A linear relationship exists between the dependent and the independent variables.
2. The variance of the residual term is constant for all observations (homoskedasticity). **Heteroskedasticity** refers to the situation when this assumption is violated.
3. The residual term is independently distributed; that is, the residual for one observation is not correlated with that of another observation.
4. The residual term is normally distributed.

Analysis of Variance (ANOVA)

ANOVA is a statistical procedure for analyzing the total variation in the dependent variable. Three of the measures in an ANOVA table are as follows:

- **Sum of squares total (SST)**. Sum of the squared differences between the actual Y -values and the mean of Y . SST measures the total variation in the dependent variable.
- **Sum of squares regression (SSR)**. Sum of the squared differences between the predicted Y -values and the mean of Y . SSR measures the variation in the dependent variable that *is* explained by the independent variable.
- **Sum of squared errors (SSE)**. Sum of the squared differences between the actual Y -values and the predicted Y -values. SSE measures the variation in the dependent variable that *is not* explained by the independent variable.

From these, we can calculate the **coefficient of determination**, or **R-squared**, which is the proportion of the total variation in the dependent variable that is explained by the independent variable:

$$R^2 = SSR / SST$$

The **mean square regression (MSR)** is equal to the SSR divided by the regression's degrees of freedom. For a simple linear regression the degrees of freedom are one, because we have one independent variable.

The **mean squared error (MSE)** is the SSE divided by its degrees of freedom, which for a simple linear regression is the number of observations minus two.

$$MSE = \frac{SSE}{n-2}$$

The **standard error of estimate (SEE)** for a regression is the standard deviation of its residuals. The lower the SEE, the better the model fit.

$$SEE = \sqrt{MSE}$$

Dividing the MSR by the MSE gives us an *F*-statistic that we can use to test whether the slope coefficient is statistically significant.

$$F = MSR / MSE$$

Hypothesis Test of the Slope Coefficient

For a simple linear regression, the *F*-test is equivalent to a *t*-test of the significance of the estimated slope coefficient (b_1). The test statistic is as follows:

$$t = \frac{\hat{b}_1 - b}{S_{\hat{b}_1}}$$

where:

b_1 = hypothesized value of slope coefficient

$$s_{\hat{b}_1} = \text{standard error of slope coefficient} = \frac{SEE}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

Degrees of freedom for this *t*-test are $n - 2$. The null hypothesis is that $b_1 = 0$ (the independent variable has no significant explanatory power for the value of the dependent variable).

Indicator Variables

An **indicator variable** or **dummy variable** takes on a value of 1 when some condition holds, or 0 when that condition does not hold. To determine whether that condition has a significant effect on the dependent variable, an analyst can use an indicator as an independent variable and test the significance of its slope coefficient. Using this technique in a simple linear regression is equivalent to

dividing the dependent variable's observations into two groups (when the condition holds and when it doesn't hold) and performing a difference-in-means test.

Predicted Values and Confidence Intervals

We use a regression model to predict values of the dependent variable, given predicted values for the independent variable. For example, given the following regression equation:

$$\hat{Y} = -2.3\% + 0.64\hat{X}$$

If we have a forecast value for X of 10%, the predicted value of Y is $-2.3\% + 0.64(10\%) = 4.1\%$.

We can construct a confidence interval around the predicted value, as follows:

$$Y \pm (t_c \times s_f)$$

where:

t_c = two-tailed critical t -value at the desired level of significance
with $df = n - 2$

$$s_f = \text{standard error of the forecast} = \text{SEE}^2 \left[1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{(n - 1)s_X^2} \right]$$

On the Level I exam, we believe any question that requires the standard error of the forecast is highly likely to provide a value for it.

Linear Regression with Transformed Variables

If one or both of the variables appear to be growing at a constant rate in percentage terms (constant compound, or exponential, rate of growth), we can transform them using the natural logarithm and then apply simple linear regression. Such a model is called a **log-lin model** if the dependent variable is transformed while the independent variable is linear, a **lin-log model** if the independent variable is transformed while the dependent variable is linear, or a **log-log model** if both variables are transformed.

Reading 11: Introduction to Big Data Techniques

SECRET SAUCE

The term **fintech** refers to developments in technology that can be applied to the financial services industry. Examples include tools and techniques for analyzing very large datasets, such as artificial intelligence.

Big Data is an expression that refers to all the potentially useful information that is generated in the economy, from traditional sources (financial markets, company financial reports, and economic statistics) as well as **alternative data** from nontraditional sources (data from individuals' online activity, **corporate exhaust** such as bank records and retail scanner data, and data from the **Internet of Things** such as smartphones and smart buildings).

Characteristics of Big Data include its volume, velocity, and variety. The *volume* of data continues to grow by orders of magnitude and is now measured in terabytes (1,000 gigabytes) and even petabytes (1,000 terabytes). *Velocity* refers to how quickly data are communicated. Real-time data are said to have low **latency** while data communicated with a lag have high latency. The *variety* of data ranges from structured forms such as spreadsheets and databases, to semistructured forms such as photos and web page code, to unstructured forms such as video.

Data science describes methods for processing and visualizing data. Processing methods include *capture* (collecting data), *curation* (assuring data quality), *storage* (archiving and accessing data), *search* (finding needed information in stored data), and *transfer* (moving data to where they are needed). Visualization techniques include charts and graphs for structured data, and tools such as word clouds or mind maps for less structured data.

Artificial intelligence, or computer systems that can be programmed to simulate human cognition, may be applied to organize unstructured data. **Neural networks** are an example of artificial intelligence. In **machine learning**, a computer algorithm is given inputs of source data and designed to learn how to recognize patterns in the input data or model output data.

In **supervised learning**, the input and output data are labeled, the machine learns to model the outputs from the inputs, and then the machine is given new data on which to use the model. In **unsupervised learning**, the input data are not labeled, and the machine learns to describe the structure of the data. **Deep learning** uses layers of neural networks to identify patterns, and may employ supervised or unsupervised learning. Image and speech recognition are among the applications of deep learning.

Overfitting occurs when the machine creates a model that is too complex: the machine learns the input and output data too exactly and identifies spurious

patterns and relationships. **Underfitting** occurs when the model is not complex enough to describe the data: the machine fails to identify actual patterns and relationships.

Applications of fintech that are relevant to investment management include text analytics, natural language processing, risk governance, and algorithmic trading.

- **Text analytics** refers to the analysis of unstructured data in text or voice forms. It has the potential to partially automate tasks such as evaluating company regulatory filings.
- **Natural language processing** refers to the use of computers and artificial intelligence to interpret human language. Possible applications in finance could be to check for regulatory compliance in employee communications, or to evaluate large volumes of research reports to detect subtle changes in sentiment.
- **Risk governance** requires an understanding of a firm's exposure to a wide variety of risks. Machine learning and other techniques related to Big Data can be useful in modeling and testing risk, particularly if firms use real-time data to monitor risk exposures.
- **Algorithmic trading** is computerized securities trading based on predetermined rules. This can be useful for optimizing execution instructions, dividing large orders across exchanges in the best way, or to carry out **high-frequency trading** that takes advantage of intraday securities mispricings.

Topic Focus: Economics

SECRET SAUCE

Weight on Exam	6% to 9%
----------------	----------

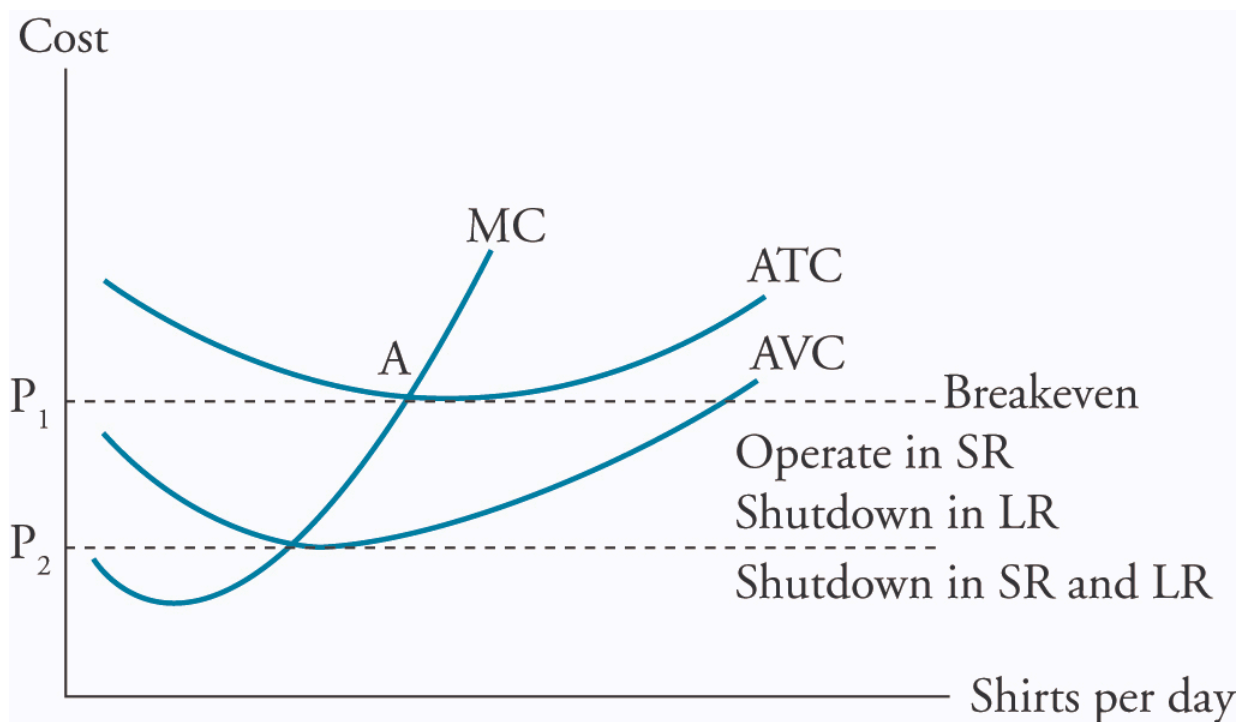
SchweserNotes™ Reference	Book 1, Pages 141–239
--------------------------	-----------------------

Breakeven and Shutdown

In the short run, a firm may be selling at less than average total cost (ATC), generating an economic loss. Such a firm should continue to operate in the short run as long as price is greater than average variable cost (AVC). In this case, the losses from shutting down (producing zero output) in the short run would be greater (equal to total fixed costs [TFC]) than the losses from continued operation. If selling price is less than AVC, the firm will minimize its losses in the short run by ceasing operations.

In the long run, a firm should shut down if price is expected to remain less than the ATC. These cases are illustrated in **Shutdown and Breakeven**. At prices below P_1 but above P_2 , a profit-maximizing (loss-minimizing) firm should continue to operate in the short run but shut down in the long run. At prices below P_2 , the firm should shut down in the short run as well. We refer to this price (minimum AVC) as the **short-run shutdown point**.

Shutdown and Breakeven

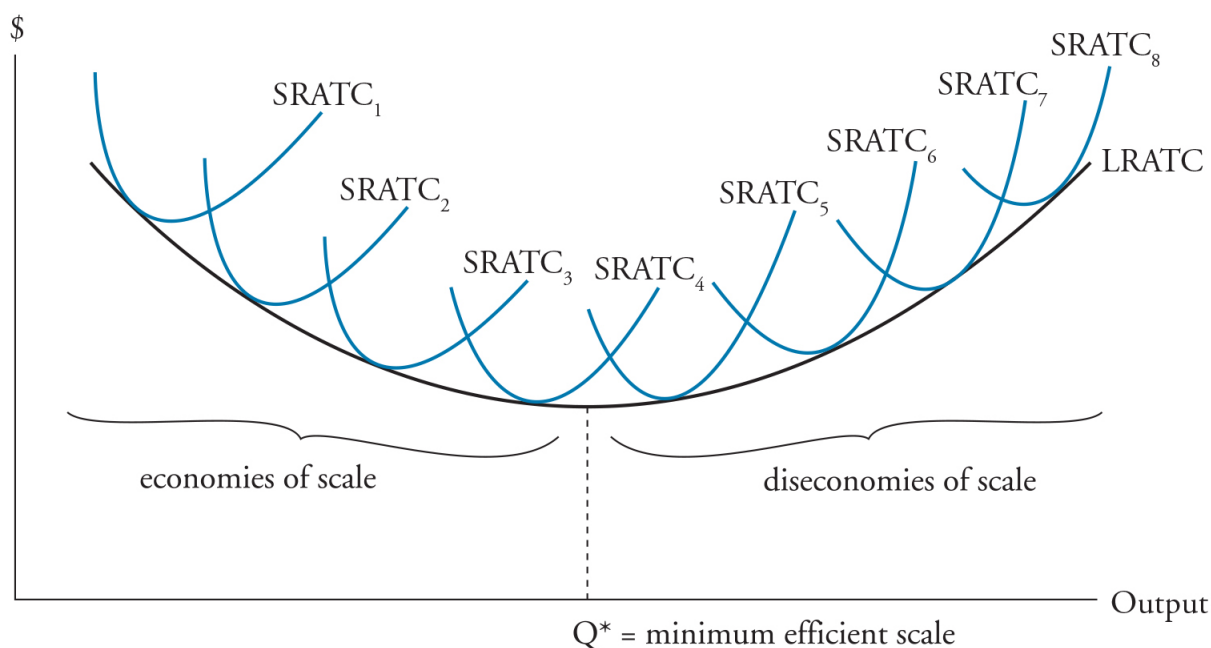


Economies and Diseconomies of Scale

In the long run, firms can adjust their scale of operations (i.e., capital is variable). The minimum ATC at each possible scale of operations is shown on the **long-run average total cost (LRATC)** curve.

The downward-sloping segment of the LRATC curve presented in **Economies and Diseconomies of Scale** indicates that **economies of scale** (or *increasing returns to scale*) are present. Economies of scale result from factors such as labor specialization, mass production, and investment in more efficient equipment and technology. In addition, the firm may be able to negotiate lower input prices with suppliers as firm size increases and more resources are purchased. The lowest point on the LRATC curve corresponds to the scale or plant size at which the ATC of production is at a minimum. This scale is sometimes called the **minimum efficient scale**. At larger firm sizes, minimum ATCs begin to increase, indicating that there are **diseconomies of scale** beyond the minimum efficient scale.

Economies and Diseconomies of Scale



Characteristics of Market Structures

We can differentiate among four types of markets based on the following characteristics:

- Number of firms and their relative sizes
- Degree to which firms differentiate their products
- Bargaining power of firms with respect to pricing
- Barriers to entry into or exit from the industry
- Degree to which firms compete on factors other than price

At one end of the spectrum is **perfect competition**, in which many firms produce identical products and competition forces them all to sell at the market price. At the other extreme, we have **monopoly**, where only one firm is producing the product. In between are **monopolistic competition** (many sellers and differentiated products) and **oligopoly** (few firms that compete in various ways).

Markets can be differentiated by several characteristics, including number of seller firms, their market shares or industry concentration, the degree of product

differentiation, the nature of competition, and barriers to entry into and exit from the industry. We can identify four primary types of market structures. An analyst, however, may be most concerned with the pricing power a particular firm has. To identify the type of market in which a firm operates, an analyst should focus on the number of firms in the market, their market shares, the nature of competition, the availability of substitute goods, and barriers to entry into and exit from the industry.

Perfect competition is characterized by the following:

- Many firms, each small relative to the market
- Very low barriers to entry into or exit from the industry
- Homogeneous products that are perfect substitutes
- No advertising or branding
- No pricing power

Monopolistic competition is characterized by the following:

- Many firms
- Low barriers to entry into or exit from the industry
- Differentiated products, heavy advertising, and high marketing expenditure
- Firms that have some pricing power

Oligopoly markets are characterized by the following:

- Few sellers
- High barriers to entry into or exit from the industry
- Products that may be homogeneous or differentiated by branding and advertising
- Firms that may have significant pricing power

Monopoly is characterized by the following:

- A single firm that comprises the whole market
- Very high barriers to enter or exit the industry
- Advertising used to compete with any substitute products (normally no good substitutes)
- Significant pricing power

Demand Characteristics

Perfect competition

Price = marginal revenue = marginal cost (in equilibrium)

Perfectly elastic firm demand curve

Zero economic profit in equilibrium

Monopolistic competition	<p>Price > marginal revenue = marginal cost (in equilibrium)</p> <p>Downward-sloping firm demand curve</p> <p>Zero economic profit in long-run equilibrium</p>
Oligopoly	<p>Price > marginal revenue = marginal cost (in equilibrium)</p> <p>Downward-sloping firm demand curve</p> <p>May have positive economic profit in long-run equilibrium</p> <p>Tends towards zero economic profit over time</p>
Monopoly	<p>Price > marginal revenue = marginal cost (in equilibrium)</p> <p>Downward-sloping firm demand curve</p> <p>May have positive economic profit in long-run equilibrium</p> <p>Profits may be zero because of expenditures to preserve monopoly</p>

The nature of competition in oligopoly markets can be described by alternative models. At one extreme, competition within an oligopoly market is strong, the product undifferentiated, and the result is very much like perfect competition in the long run. At the other extreme, if oligopolistic firms successfully collude (mostly illegally), they will charge the price a monopolist would and agree to share the economic profits. Between these extremes, we have the following three models:

1. The **kinked demand curve model** is based on an assumption that a firm's competitors will not follow a price increase, but they will cut their prices in response to a price decrease by a competitor. Under this model, each firm faces a demand curve with a kink at the current market price—more elastic above the current price and less elastic below the current price.
2. The **Cournot model** assumes that the firms in a two-firm oligopoly have identical cost structures and react only to the price charged by the other firm in the prior period. Each firm will produce half the industry output and charge the same price in equilibrium. This is a special case of a **Nash equilibrium**, defined as a situation in which no firm can increase profits by changing its price or output choice. The incentive for firms to cheat on a collusive