

The challenge with computing a confidence interval for a predicted value is calculating  $s_f^2$ . It's highly unlikely that you will have to calculate the standard error of the forecast (it will probably be provided if you need to compute a confidence interval for the dependent variable). However, if you do need to calculate  $s_f^2$ , it can be done with the following formula for the variance of the forecast:

$$s_f^2 = \text{SER}^2 \left[ 1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{(n-1)s_x^2} \right]$$

where:

$\text{SER}^2$  = variance of the residuals = the square of the standard error of the regression

$s_x^2$  = variance of the independent variable

$X$  = value of the independent variable for which the forecast was made

#### Example: Confidence interval for a predicted value

Calculate a 95% prediction interval on the predicted value of WPO from the previous example. Assume the standard error of the forecast is 3.67, and the forecasted value of S&P 500 excess returns is 10%.

Answer:

The predicted value for WPO is:

$$\widehat{\text{WPO}} = -2.3\% + (0.64)(10\%) = 4.1\%$$

The 5% two-tailed critical  $t$ -value with 34 degrees of freedom is 2.03. The prediction interval at the 95% confidence level is:

$$\widehat{\text{WPO}} \pm (t_c \times s_f) \Rightarrow [4.1\% \pm (2.03 \times 3.67\%)] = 4.1\% \pm 7.5\%$$

or

-3.4% to 11.6%

This range can be interpreted as, given a forecasted value for S&P 500 excess returns of 10%, we can be 95% confident that the WPO excess returns will be between -3.4% and 11.6%.

## DUMMY VARIABLES

Observations for most independent variables (e.g., firm size, level of GDP, and interest rates) can take on a wide range of values. However, there are occasions when the independent variable is binary in nature—it is either “on” or “off.” Independent variables that fall into this category are called **dummy variables** and are often used to quantify the impact of qualitative events.



*Professor's Note: We will address dummy variables in more detail when we demonstrate how to model seasonality in Topic 25.*

## WHAT IS HETEROSKEDASTICITY?

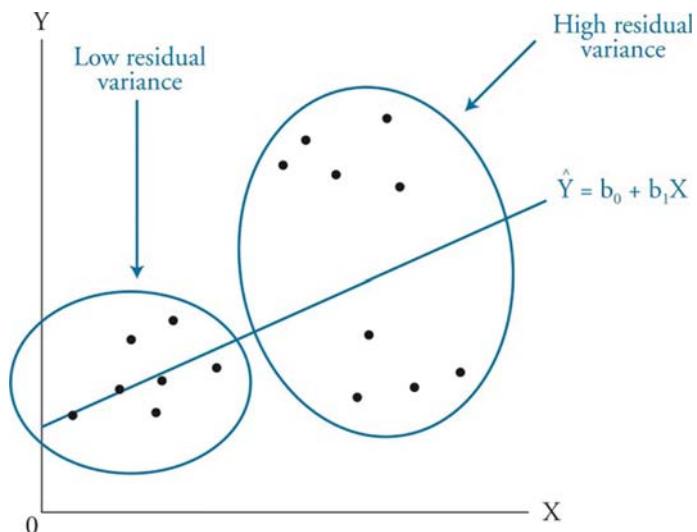
### LO 21.4: Evaluate the implications of homoskedasticity and heteroskedasticity.

If the variance of the residuals is constant across all observations in the sample, the regression is said to be **homoskedastic**. When the opposite is true, the regression exhibits **heteroskedasticity**, which occurs when the variance of the residuals is not the same across all observations in the sample. This happens when there are subsamples that are more spread out than the rest of the sample.

Unconditional heteroskedasticity occurs when the heteroskedasticity is not related to the level of the independent variables, which means that it doesn't systematically increase or decrease with changes in the value of the independent variable(s). While this is a violation of the equal variance assumption, *it usually causes no major problems with the regression.*

Conditional heteroskedasticity is heteroskedasticity that is related to the level of (i.e., conditional on) the independent variable. For example, conditional heteroskedasticity exists if the variance of the residual term increases as the value of the independent variable increases, as shown in Figure 1. Notice in this figure that the residual variance associated with the larger values of the independent variable,  $X$ , is larger than the residual variance associated with the smaller values of  $X$ . Conditional heteroskedasticity *does create significant problems for statistical inference.*

Figure 1: Conditional Heteroskedasticity



## Effect of Heteroskedasticity on Regression Analysis

There are several effects of heteroskedasticity you need to be aware of:

- The standard errors are usually unreliable estimates.
- The coefficient estimates (the  $b_j$ ) aren't affected.
- If the standard errors are too small, but the coefficient estimates themselves are not affected, the  $t$ -statistics will be too large and the null hypothesis of no statistical significance is rejected too often. The opposite will be true if the standard errors are too large.

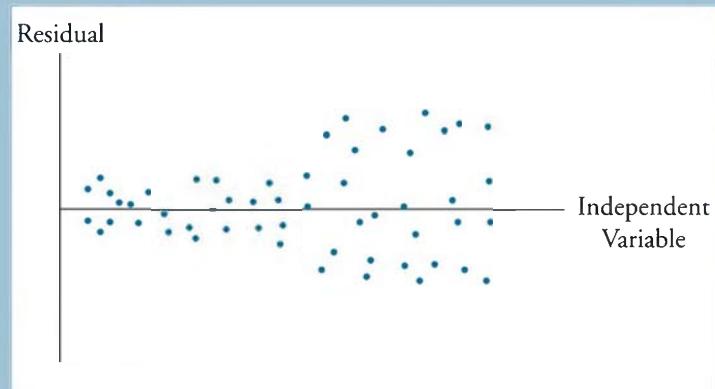
## Detecting Heteroskedasticity

As was shown in Figure 1, a scatter plot of the residuals versus one of the independent variables can reveal patterns among observations.

### Example: Detecting heteroskedasticity with a residual plot

You have been studying the monthly returns of a mutual fund over the past five years, hoping to draw conclusions about the fund's average performance. You calculate the mean return, the standard deviation, and the portfolio's beta by regressing the fund's returns on S&P 500 index returns (the independent variable). The standard deviation of returns and the fund's beta don't seem to fit the firm's stated risk profile. For your analysis, you have prepared a scatter plot of the error terms (actual return – predicted return) for the regression using five years of returns, as shown in the following figure. Determine whether the residual plot indicates that there may be a problem with the data.

### Residual Plot



### Answer:

The residual plot in the previous figure indicates the presence of conditional heteroskedasticity. Notice how the variation in the regression residuals increases as the independent variable increases. This indicates that the variance of the fund's returns about the mean is related to the level of the independent variable.

## Correcting Heteroskedasticity

Heteroskedasticity is not easy to correct, and the details of the available techniques are beyond the scope of the FRM curriculum. The most common remedy, however, is to calculate **robust standard errors**. These robust standard errors are used to recalculate the *t*-statistics using the original regression coefficients. On the exam, use robust standard errors to calculate *t*-statistics if there is evidence of heteroskedasticity. By default, many statistical software packages apply *homoskedastic* standard errors unless the user specifies otherwise.

## THE GAUSS-MARKOV THEOREM

---

**LO 21.5:** Determine the conditions under which the OLS is the best linear conditionally unbiased estimator.

**LO 21.6:** Explain the Gauss-Markov Theorem and its limitations, and alternatives to the OLS.

---

The **Gauss-Markov theorem** says that if the linear regression model assumptions are true and the regression errors display homoskedasticity, then the OLS estimators have the following properties.

1. The OLS estimated coefficients have the minimum variance compared to other methods of estimating the coefficients (i.e., they are the most precise).
2. The OLS estimated coefficients are based on linear functions.
3. The OLS estimated coefficients are unbiased, which means that in repeated sampling the averages of the coefficients from the sample will be distributed around the true population parameters [i.e.,  $E(b_0) = B_0$  and  $E(b_1) = B_1$ ].
4. The OLS estimate of the variance of the errors is unbiased [i.e.,  $E(\hat{\sigma}^2) = \sigma^2$ ].

The acronym for these properties is “BLUE,” which indicates that OLS estimators are the best linear unbiased estimators.

One limitation of the Gauss-Markov theorem is that its conditions may not hold in practice, particularly when the error terms are heteroskedastic, which is sometimes observed in economic data. Another limitation is that alternative estimators, which are not linear or unbiased, may be more efficient than OLS estimators. Examples of these alternative estimators include: the weighted least squares estimator (which can produce an estimator with a smaller variance—to combat heteroskedastic errors) and the least absolute deviations estimator (which is less sensitive to extreme outliers given that rare outliers exist in the data).

## SMALL SAMPLE SIZES

---

### LO 21.7: Apply and interpret the $t$ -statistic when the sample size is small.

---

The central limit theorem is important when analyzing OLS results because it allows for the use of the  $t$ -distribution when conducting hypothesis testing on regression coefficients. This is possible because the central limit theorem says that the means of individual samples will be normally distributed when the sample size is large. However, if the sample size is small, the distribution of a  $t$ -statistic becomes more complicated to interpret.

In order to analyze a regression coefficient  $t$ -statistic when the sample size is small, we must assume the assumptions underlying linear regression hold. In particular, in order to apply and interpret the  $t$ -statistic, error terms must be homoskedastic (i.e., constant variance of error terms) and the error terms must be normally distributed. If this is the case, the  $t$ -statistic can be computed using the default standard error (i.e., the homoskedasticity-only standard error), and it follows a  $t$ -distribution with  $n - 2$  degrees of freedom.

In practice, it is rare to assume that error terms have a constant variance and are normally distributed. However, it is generally the case that sample sizes are large enough to apply the central limit theorem meaning that we can calculate  $t$ -statistics using homoskedasticity-only standard errors. In other words, with a large sample size, differences between the  $t$ -distribution and the standard normal distribution can be ignored.

## KEY CONCEPTS

### LO 21.1

The confidence interval for the regression coefficient,  $B_1$ , is calculated as:

$$b_1 \pm (t_c \times s_{b_1}), \text{ or } [b_1 - (t_c \times s_{b_1}) < B_1 < b_1 + (t_c \times s_{b_1})]$$

### LO 21.2

The  $p$ -value is the smallest level of significance for which the null hypothesis can be rejected. Interpreting the  $p$ -value offers an alternative approach when testing for statistical significance.

### LO 21.3

A  $t$ -test with  $n - 2$  degrees of freedom is used to conduct hypothesis tests of the estimated regression parameters:

$$t = \frac{b_1 - B_1}{s_{b_1}}$$

A predicted value of the dependent variable,  $\hat{Y}$ , is determined by inserting the predicted value of the independent variable,  $X_p$ , in the regression equation and calculating

$$\hat{Y}_p = b_0 + b_1 X_p.$$

The confidence interval for a predicted  $Y$ -value is  $[\hat{Y} - (t_c \times s_f) < Y < \hat{Y} + (t_c \times s_f)]$ , where  $s_f$  is the standard error of the forecast.

Qualitative independent variables (dummy variables) capture the effect of a binary independent variable:

- Slope coefficient is interpreted as the change in the dependent variable for the case when the dummy variable is one.
- Use one less dummy variable than the number of categories.

### LO 21.4

Homoskedasticity refers to the condition of constant variance of the residuals.

Heteroskedasticity refers to a violation of this assumption.

The effects of heteroskedasticity are as follows:

- The standard errors are usually unreliable estimates.
- The coefficient estimates (the  $b_j$ ) aren't affected.
- If the standard errors are too small, but the coefficient estimates themselves are not affected, the  $t$ -statistics will be too large and the null hypothesis of no statistical significance is rejected too often. The opposite will be true if the standard errors are too large.

**LO 21.5**

The Gauss-Markov theorem says that if linear regression assumptions are true, then OLS estimators are the best linear unbiased estimators.

---

**LO 21.6**

The limitations of the Gauss-Markov theorem are that its conditions may not hold in practice and alternative estimators may be more efficient. Examples of alternative estimators include the weighted least squares estimator and the least absolute deviations estimator.

---

**LO 21.7**

In order to interpret  $t$ -statistics of regression coefficients when a sample size is small, we must assume the assumptions underlying linear regression hold. In practice, it is generally the case that sample sizes are large, meaning that  $t$ -statistics can be computed using homoskedasticity-only standard errors.

## CONCEPT CHECKERS

1. What is the appropriate alternative hypothesis to test the statistical significance of the intercept term in the following regression?

$$Y = a_1 + a_2(X) + \varepsilon$$

- A.  $H_A: a_1 \neq 0$ .
- B.  $H_A: a_1 > 0$ .
- C.  $H_A: a_2 \neq 0$ .
- D.  $H_A: a_2 > 0$ .

Use the following information for Questions 2 through 4.

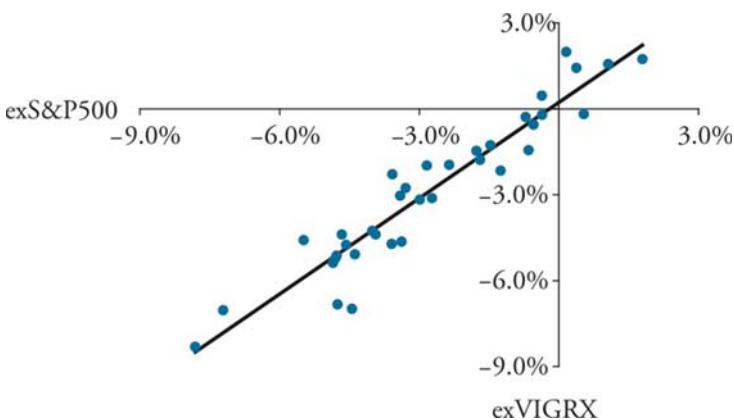
Bill Coldplay is analyzing the performance of the Vanguard Growth Index Fund (VIGRX) over the past three years. The fund employs a passive management investment approach designed to track the performance of the MSCI US Prime Market Growth index, a broadly diversified index of growth stocks of large U.S. companies.

Coldplay estimates a regression using excess monthly returns on VIGRX (exVIGRX) as the dependent variable and excess monthly returns on the S&P 500 index (exS&P) as the independent variable. The data are expressed in decimal terms (e.g., 0.03, not 3%).

$$\text{exVIGRX}_t = b_0 + b_1(\text{exS\&P}_t) + \varepsilon_t$$

A scatter plot of excess returns for both return series from June 2004 to May 2007 are shown in the following figure.

Analysis of Large Cap Growth Fund



Results from that analysis are presented in the following figures.

<i>Coefficient</i>	<i>Coefficient Estimate</i>	<i>Standard Error</i>
$b_0$	0.0023	0.0022
$b_1$	1.1163	0.0624

<i>Source of Variation</i>	<i>Sum of Squares</i>
Explained	0.0228
Residual	0.0024

2. The 90% confidence interval for  $b_0$  is closest to:
  - A. -0.0014 to +0.0060.
  - B. -0.0006 to +0.0052.
  - C. +0.0001 to +0.0045.
  - D. -0.0006 to +0.0045.
  
3. Are the intercept term and the slope coefficient statistically significantly different from zero at the 5% significance level?
 

<u>Intercept term significant?</u>	<u>Slope coefficient significant?</u>
A. Yes	Yes
B. Yes	No
C. No	Yes
D. No	No
  
4. Coldplay would like to test the following hypothesis:  $H_0: B_1 \leq 1$  vs.  $H_A: B_1 > 1$  at the 1% significance level. The calculated  $t$ -statistic and the appropriate conclusion are:
 

<u>Calculated <math>t</math>-statistic</u>	<u>Appropriate conclusion</u>
A. 1.86	Reject $H_0$
B. 1.86	Fail to reject $H_0$
C. 2.44	Reject $H_0$
D. 2.44	Fail to reject $H_0$
  
5. Consider the following statement: In a simple linear regression, the appropriate degrees of freedom for the critical  $t$ -value used to calculate a confidence interval around both a parameter estimate and a predicted Y-value is the same as the number of observations minus two. The statement is:
  - A. justified.
  - B. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a parameter estimate is the number of observations.
  - C. not justified, because the appropriate degrees of freedom used to calculate a confidence interval around a predicted Y-value is the number of observations.
  - D. not justified, because the appropriate degrees of freedom used to calculate a confidence interval depends on the explained sum of squares.

## CONCEPT CHECKER ANSWERS

1. A In this regression,  $a_1$  is the intercept term. To test the statistical significance means to test the null hypothesis that  $a_1$  is equal to zero versus the alternative that it is not equal to zero.
2. A Note that there are 36 monthly observations from June 2004 to May 2007, so  $n = 36$ . The critical two-tailed 10%  $t$ -value with  $34$  ( $n - 2 = 36 - 2 = 34$ ) degrees of freedom is approximately 1.69. Therefore, the 90% confidence interval for  $b_0$  (the intercept term) is  $0.0023 \pm (0.0022)(1.69)$ , or  $-0.0014$  to  $+0.0060$ .
3. C The critical two-tailed 5%  $t$ -value with 34 degrees of freedom is approximately 2.03. The calculated  $t$ -statistics for the intercept term and slope coefficient are, respectively,  $0.0023 / 0.0022 = 1.05$  and  $1.1163 / 0.0624 = 17.9$ . Therefore, the intercept term is not statistically different from zero at the 5% significance level, while the slope coefficient is.
4. B Notice that this is a one-tailed test. The critical one-tailed 1%  $t$ -value with 34 degrees of freedom is approximately 2.44. The calculated  $t$ -statistic for the slope coefficient is  $(1.1163 - 1) / 0.0624 = 1.86$ . Therefore, the slope coefficient is not statistically different from one at the 1% significance level, and Coldplay should fail to reject the null hypothesis.
5. A In simple linear regression, the appropriate degrees of freedom for both confidence intervals is the number of observations in the sample ( $n$ ) minus two.

# LINEAR REGRESSION WITH MULTIPLE REGRESSORS

Topic 22

## EXAM FOCUS

Multiple regression is, in many ways, simply an extension of regression with a single regressor. The coefficient of determination, t-statistics, and standard errors of the coefficients are interpreted in the same fashion. There are some differences, however; namely that the formulas for the coefficients and standard errors are more complicated. The slope coefficients are called partial slope coefficients because they measure the effect of changing one independent variable, assuming the others are held constant. For the exam, understand the implications of omitting relevant independent variables from the model, the adjustment to the coefficient of determination when adding additional variables, and the effect that heteroskedasticity and multicollinearity have on regression results.

## OMITTED VARIABLE BIAS

**LO 22.1: Define and interpret omitted variable bias, and describe the methods for addressing this bias.**

Omitting relevant factors from an ordinary least squares (OLS) regression can produce misleading or biased results. **Omitted variable bias** is present when two conditions are met: (1) the omitted variable is correlated with the movement of the independent variable in the model, and (2) the omitted variable is a determinant of the dependent variable. When relevant variables are absent from a linear regression model, the results will likely lead to incorrect conclusions as the OLS estimators may not accurately portray the actual data.

Omitted variable bias violates the assumptions of OLS regression when the omitted variable is in fact correlated with current independent (explanatory) variable(s). The reason for this violation is because omitted factors that partially describe the movement of the dependent variable will become part of the regression's error term since they are not properly identified within the model. If the omitted variable is correlated with the regression's slope coefficient, then the error term will also be correlated with the slope coefficient. Recall, that according to the assumptions of linear regression, the independent variable must be uncorrelated with the error term.

The issue of omitted variable bias occurs regardless of the size of the sample and will make OLS estimators inconsistent. The correlation between the omitted variable and the independent variable will determine the size of the bias (i.e., a larger correlation will lead to a larger bias) and the direction of the bias (i.e., whether the correlation is positive or negative). In addition, this bias can also have a dramatic effect on the test statistics used to determine whether the independent variables are statistically significant.

Testing for omitted variable bias would check to see if the two conditions addressed earlier are present. If a bias is found, it can be addressed by dividing data into groups and examining one factor at a time while holding other factors constant. However, in order to understand the full effects of all relevant independent variables on the dependent variable, we need to utilize multiple independent coefficients in our model. Multiple regression analysis is therefore used to eliminate omitted variable bias since it can estimate the effect of one independent variable on the dependent variable while holding all other variables constant.

## MULTIPLE REGRESSION BASICS

---

### LO 22.2: Distinguish between single and multiple regression.

---

**Multiple regression** is regression analysis with more than one independent variable. It is used to quantify the influence of two or more independent variables on a dependent variable. For instance, simple (or univariate) linear regression explains the variation in stock returns in terms of the variation in systematic risk as measured by beta. With multiple regression, stock returns can be regressed against beta and against additional variables, such as firm size, equity, and industry classification, that might influence returns.

The general multiple linear regression model is:

$$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_k X_{ki} + \epsilon_i$$

where:

$Y_i$  =  $i$ th observation of the dependent variable  $Y$ ,  $i = 1, 2, \dots, n$

$X_j$  = independent variables,  $j = 1, 2, \dots, k$

$X_{ji}$  =  $i$ th observation of the  $j$ th independent variable

$B_0$  = intercept term

$B_j$  = slope coefficient for each of the independent variables

$\epsilon_i$  = error term for the  $i$ th observation

$n$  = number of observations

$k$  = number of independent variables

---

### LO 22.5: Describe the OLS estimator in a multiple regression.

---

The multiple regression methodology estimates the intercept and slope coefficients such that the sum of the squared error terms,  $\sum_{i=1}^n \epsilon_i^2$ , is minimized. The estimators of these coefficients are known as **ordinary least squares (OLS) estimators**. The OLS estimators are typically found with statistical software, but can also be computed using calculus or a trial-and-error method. The result of this procedure is the following regression equation:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki}$$

where the lowercase  $b_i$ 's indicate an estimate for the corresponding regression coefficient

The residual,  $e_i$ , is the difference between the observed value,  $Y_i$ , and the predicted value from the regression,  $\hat{Y}_i$ :

$$e_i = Y_i - \hat{Y}_i = Y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki})$$

### LO 22.3: Interpret the slope coefficient in a multiple regression.

Let's illustrate multiple regression using research by Arnott and Asness (2003).<sup>1</sup> As part of their research, the authors test the hypothesis that future 10-year real earnings growth in the S&P 500 (EG10) can be explained by the trailing dividend payout ratio of the stocks in the index (PR) and the yield curve slope (YCS). YCS is calculated as the difference between the 10-year T-bond yield and the 3-month T-bill yield at the start of the period. All three variables are measured in percent.

#### Formulating the Multiple Regression Equation

The authors formulate the following regression equation using annual data (46 observations):

$$\text{EG10} = B_0 + B_1 \text{PR} + B_2 \text{YCS} + \varepsilon$$

The results of this regression are shown in Figure 1.

**Figure 1: Estimates for Regression of EG10 on PR and YCS**

	Coefficient	Standard Error
Intercept	-11.6%	1.657%
PR	0.25	0.032
YCS	0.14	0.280

#### Interpreting the Multiple Regression Results

The interpretation of the estimated regression coefficients from a multiple regression is the same as in simple linear regression for the intercept term but significantly different for the slope coefficients:

- The **intercept term** is the value of the dependent variable when the independent variables are all equal to zero.
- Each slope coefficient is the estimated change in the dependent variable for a one-unit change in that independent variable, *holding the other independent variables constant*. That's why the slope coefficients in a multiple regression are sometimes called **partial slope coefficients**.

For example, in the real earnings growth example, we can make these interpretations:

- *Intercept term:* If the dividend payout ratio is zero and the slope of the yield curve is zero, we would expect the subsequent 10-year real earnings growth rate to be -11.6%.
- *PR coefficient:* If the payout ratio increases by 1%, we would expect the subsequent 10-year earnings growth rate to increase by 0.25%, *holding YCS constant*.
- *YCS coefficient:* If the yield curve slope increases by 1%, we would expect the subsequent 10-year earnings growth rate to increase by 0.14%, *holding PR constant*.

1. Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal*, vol. 59, no. 1 (January/February): 70–87.

Let's discuss the interpretation of the multiple regression slope coefficients in more detail. Suppose we run a regression of the dependent variable  $Y$  on a single independent variable  $X_1$  and get the following result:

$$Y = 2.0 + 4.5X_1$$

The appropriate interpretation of the estimated slope coefficient is that if  $X_1$  increases by 1 unit, we would expect  $Y$  to increase by 4.5 units.

Now suppose we add a second independent variable  $X_2$  to the regression and get the following result:

$$Y = 1.0 + 2.5X_1 + 6.0X_2$$

Notice that the estimated slope coefficient for  $X_1$  changed from 4.5 to 2.5 when we added  $X_2$  to the regression. We would expect this to happen most of the time when a second variable is added to the regression, unless  $X_2$  is uncorrelated with  $X_1$ , because if  $X_1$  increases by 1 unit, then we would expect  $X_2$  to change as well. The multiple regression equation captures this relationship between  $X_1$  and  $X_2$  when predicting  $Y$ .

Now the interpretation of the estimated slope coefficient for  $X_1$  is that if  $X_1$  increases by 1 unit, we would expect  $Y$  to increase by 2.5 units, *holding X<sub>2</sub> constant*.

#### **LO 22.4: Describe homoskedasticity and heteroskedasticity in a multiple regression.**

In multiple regression, homoskedasticity and heteroskedasticity are just extensions of their definitions discussed in the previous topic. Homoskedasticity refers to the condition that the variance of the error term is constant for all independent variables,  $X_i$ , from  $i = 1$  to  $n$ :  $\text{Var}(\varepsilon_i | X_i) = \sigma^2$ . Heteroskedasticity means that the dispersion of the error terms varies over the sample. It may take the form of conditional heteroskedasticity, which says that the variance is a function of the independent variables.

#### **MEASURES OF FIT**

#### **LO 22.6: Calculate and interpret measures of fit in multiple regression.**

The standard error of the regression (SER) measures the uncertainty about the accuracy of the predicted values of the dependent variable,  $\hat{Y}_i = b_0 + b_1 X_i$ . Graphically, the relationship is stronger when the actual x,y data points lie closer to the regression line (i.e., the  $e_i$  are smaller).

Formally, SER is the standard deviation of the predicted values for the dependent variable about the regression line. Equivalently, it is the standard deviation of the error terms in the regression. SER is sometimes specified as  $s_e$ .

Recall that regression minimizes the sum of the squared vertical distances between the predicted value and actual value for each observation (i.e., prediction errors). Also, recall that the sum of the squared prediction errors,  $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ , is called the sum of squared residuals, SSR (not to be confused with SER). If the relationship between the variables in the regression is very strong (actual values are close to the line), the prediction errors, and the SSR, will be small. Thus, as shown in the following equations, the standard error of the regression is a function of the SSR:

$$SER = \sqrt{s_e^2} = \sqrt{\frac{SSR}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^n [Y_i - (b_0 + b_i X_i)]^2}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-k-1}}$$

where:

$n$  = number of observations

$k$  = number of independent variables

$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$  = SSR = the sum of squared residuals

$\hat{Y}_i = b_0 + b_i X_i$  = a point on the regression line corresponding to a value of  $X_i$ . It is the expected (predicted) value of  $Y$ , given the estimated relation between  $X$  and  $Y$ .

Similar to the standard deviation for a single variable, SER measures the degree of variability of the actual  $Y$ -values relative to the estimated  $Y$ -values. The SER gauges the “fit” of the regression line. *The smaller the standard error, the better the fit.*

## COEFFICIENT OF DETERMINATION, $R^2$

The multiple coefficient of determination,  $R^2$ , can be used to test the overall effectiveness of the entire set of independent variables in explaining the dependent variable. Its interpretation is similar to that for simple linear regression: the percentage of variation in the dependent variable that is *collectively* explained by all of the independent variables. For example, an  $R^2$  of 0.63 indicates that the model, as a whole, explains 63% of the variation in the dependent variable.

$R^2$  is calculated the same way as in simple linear regression.

$$R^2 = \frac{\text{total variation} - \text{unexplained variation}}{\text{total variation}} = \frac{TSS - SSR}{TSS} = \frac{\text{explained variation}}{\text{total variation}} = \frac{ESS}{TSS}$$

## Adjusted R<sup>2</sup>

Unfortunately, R<sup>2</sup> by itself *may not be a reliable measure of the explanatory power of the multiple regression model*. This is because R<sup>2</sup> almost always increases as independent variables are added to the model, even if the marginal contribution of the new variables is not statistically significant. Consequently, a relatively high R<sup>2</sup> may reflect the impact of a large set of independent variables rather than how well the set explains the dependent variable. This problem is often referred to as overestimating the regression.

To overcome the problem of overestimating the impact of additional variables on the explanatory power of a regression model, many researchers recommend adjusting R<sup>2</sup> for the number of independent variables. The *adjusted R<sup>2</sup>* value is expressed as:

$$R_a^2 = 1 - \left[ \left( \frac{n-1}{n-k-1} \right) \times (1 - R^2) \right]$$

where:

n = number of observations

k = number of independent variables

R<sub>a</sub><sup>2</sup> = adjusted R<sup>2</sup>

R<sub>a</sub><sup>2</sup> is less than or equal to R<sup>2</sup>. So while adding a new independent variable to the model will increase R<sup>2</sup>, it may either *increase or decrease* the R<sub>a</sub><sup>2</sup>. If the new variable has only a small effect on R<sup>2</sup>, the value of R<sub>a</sub><sup>2</sup> may decrease. In addition, R<sub>a</sub><sup>2</sup> may be less than zero if the R<sup>2</sup> is low enough.

### Example: Calculating R<sup>2</sup> and adjusted R<sup>2</sup>

An analyst runs a regression of monthly value-stock returns on five independent variables over 60 months. The total sum of squares for the regression is 460, and the sum of squared errors is 170. Calculate the R<sup>2</sup> and adjusted R<sup>2</sup>.

Answer:

$$R^2 = \frac{460 - 170}{460} = 0.630 = 63.0\%$$

$$R_a^2 = 1 - \left[ \left( \frac{60-1}{60-5-1} \right) \times (1 - 0.63) \right] = 0.596 = 59.6\%$$

The R<sup>2</sup> of 63% suggests that the five independent variables together explain 63% of the variation in monthly value-stock returns.

**Example: Interpreting adjusted R<sup>2</sup>**

Suppose the analyst now adds four more independent variables to the regression, and the R<sup>2</sup> increases to 65.0%. Identify which model the analyst would most likely prefer.

**Answer:**

With nine independent variables, even though the R<sup>2</sup> has increased from 63% to 65%, the adjusted R<sup>2</sup> has decreased from 59.6% to 58.7%:

$$R_a^2 = 1 - \left[ \left( \frac{60-1}{60-9-1} \right) \times (1 - 0.65) \right] = 0.587 = 58.7\%$$

The analyst would prefer the first model because the adjusted R<sup>2</sup> is higher and the model has five independent variables as opposed to nine.

## ASSUMPTIONS OF MULTIPLE REGRESSION

---

### LO 22.7: Explain the assumptions of the multiple linear regression model.

---

As with simple linear regression, most of the assumptions made with the multiple regression pertain to  $\epsilon$ , the model's error term:

- A linear relationship exists between the dependent and independent variables. In other words, the model in LO 22.2 correctly describes the relationship.
- The independent variables are not random, and there is no exact linear relation between any two or more independent variables.
- The expected value of the error term, conditional on the independent variables, is zero [i.e.,  $E(\epsilon | X_1, X_2, \dots, X_k) = 0$ ].
- The variance of the error terms is constant for all observations [i.e.,  $E(\epsilon_i^2) = \sigma_\epsilon^2$ ].
- The error term for one observation is not correlated with that of another observation [i.e.,  $E(\epsilon_i \epsilon_j) = 0, j \neq i$ ].
- The error term is normally distributed.

## MULTICOLLINEARITY

---

### LO 22.8: Explain the concepts of imperfect and perfect multicollinearity and their implications.

---

Multicollinearity refers to the condition when two or more of the independent variables, or linear combinations of the independent variables, in a multiple regression are highly correlated with each other. This condition distorts the standard error of the regression and the coefficient standard errors, leading to problems when conducting *t*-tests for statistical significance of parameters.

The degree of correlation will determine the difference between perfect and imperfect multicollinearity. If one of the independent variables is a perfect linear combination of the other independent variables, then the model is said to exhibit **perfect multicollinearity**. In this case, it will not be possible to find the OLS estimators necessary for the regression results.

An important consideration when performing multiple regression with dummy variables is the choice of the number of dummy variables to include in the model. Whenever we want to distinguish between  $n$  classes, we must use  $n - 1$  dummy variables. Otherwise, the regression assumption of no exact linear relationship between independent variables would be violated. In general, if every observation is linked to only one class, all dummy variables are included as regressors, and an intercept term exists, then the regression will exhibit perfect multicollinearity. This problem is known as the **dummy variable trap**. As mentioned, this issue can be avoided by excluding one of the dummy variables from the regression equation (i.e.,  $n - 1$  dummy variables). With this approach, the intercept term will represent the omitted class.

**Imperfect multicollinearity** arises when two or more independent variables are highly correlated, but less than perfectly correlated. When conducting regression analysis, we need to be cognizant of imperfect multicollinearity since OLS estimators will be computed, but the resulting coefficients may be improperly estimated. In general, when using the term multicollinearity, we are referring to the *imperfect case*, since this regression assumption violation requires detecting and correcting.

### Effect of Multicollinearity on Regression Analysis

As a result of multicollinearity, there is a *greater probability that we will incorrectly conclude that a variable is not statistically significant* (e.g., a Type II error). Multicollinearity is likely to be present to some extent in most economic models. The issue is whether the multicollinearity has a significant effect on the regression results.

### Detecting Multicollinearity

The most common way to detect multicollinearity is the situation where  $t$ -tests indicate that none of the individual coefficients is significantly different than zero, while the  $R^2$  is high. This suggests that the variables together explain much of the variation in the dependent variable, but the individual independent variables do not. The only way this can happen is when the independent variables are highly correlated with each other, so while their common source of variation is explaining the dependent variable, the high degree of correlation also “washes out” the individual effects.

High correlation among independent variables is sometimes suggested as a sign of multicollinearity. In fact, as a general rule of thumb: If the absolute value of the sample correlation between any two independent variables in the regression is greater than 0.7, multicollinearity is a potential problem. However, this only works if there are exactly two independent variables. If there are more than two independent variables, while individual variables may not be highly correlated, linear combinations might be, leading to multicollinearity. High correlation among the independent variables suggests the possibility of multicollinearity, but low correlation among the independent variables *does not necessarily* indicate multicollinearity is *not* present.

**Example: Detecting multicollinearity**

Bob Watson runs a regression of mutual fund returns on average P/B, average P/E, and average market capitalization, with the following results:

Variable	Coefficient	p-Value
Average P/B	3.52	0.15
Average P/E	2.78	0.21
Market Cap	4.03	0.11
R <sup>2</sup>	89.6%	

Determine whether or not multicollinearity is a problem in this regression.

**Answer:**

The R<sup>2</sup> is high, which suggests that the three variables as a group do an excellent job of explaining the variation in mutual fund returns. However, none of the independent variables individually is statistically significant to any reasonable degree, since the p-values are larger than 10%. This is a classic indication of multicollinearity.

**Correcting Multicollinearity**

The most common method to correct for multicollinearity is to omit one or more of the correlated independent variables. Unfortunately, it is not always an easy task to identify the variable(s) that are the source of the multicollinearity. There are statistical procedures that may help in this effort, like stepwise regression, which systematically remove variables from the regression until multicollinearity is minimized.

## KEY CONCEPTS

---

### LO 22.1

Omitted variable bias is present when two conditions are met: (1) the omitted variable is correlated with the movement of the independent variable in the model, and (2) the omitted variable is a determinant of the dependent variable.

---

### LO 22.2

The multiple regression equation specifies a dependent variable as a linear function of two or more independent variables:

$$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_k X_{ki} + \epsilon_i$$

The intercept term is the value of the dependent variable when the independent variables are equal to zero. Each slope coefficient is the estimated change in the dependent variable for a one-unit change in that independent variable, holding the other independent variables constant.

---

### LO 22.3

In a multivariate regression, each slope coefficient is interpreted as a partial slope coefficient in that it measures the effect on the dependent variable from a change in the associated independent variable holding other things constant.

---

### LO 22.4

Homoskedasticity means that the variance of error terms is constant for all independent variables, while heteroskedasticity means that the variance of error terms varies over the sample. Heteroskedasticity may take the form of conditional heteroskedasticity, which says that the variance is a function of the independent variables.

---

### LO 22.5

Multiple regression estimates the intercept and slope coefficients such that the sum of the squared error terms is minimized. The estimators of these coefficients are known as ordinary least squares (OLS) estimators. The OLS estimators are typically found with statistical software.

**LO 22.6**

The standard error of the regression is the standard deviation of the predicted values for the dependent variable about the regression line:

$$\text{SER} = \sqrt{\frac{\text{SSR}}{n - k - 1}}$$

The coefficient of determination,  $R^2$ , is the percentage of the variation in Y that is explained by the set of independent variables.

- $R^2$  increases as the number of independent variables increases—this can be a problem.
  - The adjusted  $R^2$  adjusts the  $R^2$  for the number of independent variables.
  - $R_a^2 = 1 - \left[ \left( \frac{n-1}{n-k-1} \right) \times (1 - R^2) \right]$
- 

**LO 22.7**

Assumptions of multiple regression mostly pertain to the error term,  $\varepsilon_i$

- A linear relationship exists between the dependent and independent variables.
  - The independent variables are not random, and there is no exact linear relation between any two or more independent variables.
  - The expected value of the error term is zero.
  - The variance of the error terms is constant.
  - The error for one observation is not correlated with that of another observation.
  - The error term is normally distributed.
- 

**LO 22.8**

Perfect multicollinearity exists when one of the independent variables is a perfect linear combination of the other independent variable. Imperfect multicollinearity arises when two or more independent variables are highly correlated, but less than perfectly correlated.

## CONCEPT CHECKERS

Use the following table for Question 1.

<i>Source</i>	<i>Sum of Squares (SS)</i>
Explained	1,025
Residual	925

1. The total sum of squares (TSS) is closest to:
- 100.
  - 1.108.
  - 1,950.
  - 0.9024.

Use the following information to answer Questions 2 and 3.

Multiple regression was used to explain stock returns using the following variables:

Dependent variable:

RET = annual stock returns (%)

Independent variables:

MKT = market capitalization = market capitalization / \$1.0 million

IND = industry quartile ranking (IND = 4 is the highest ranking)

FORT = Fortune 500 firm, where {FORT = 1 if the stock is that of a Fortune 500 firm, FORT = 0 if not a Fortune 500 stock}

The regression results are presented in the following table.

	<i>Coefficient</i>	<i>Standard Error</i>	<i>t-Statistic</i>	<i>p-Value</i>
Intercept	0.5220	1.2100	0.430	0.681
Market capitalization	0.0460	0.0150	3.090	0.021
Industry ranking	0.7102	0.2725	2.610	0.040
Fortune 500	0.9000	0.5281	1.700	0.139

2. Based on the results in the table, which of the following most accurately represents the regression equation?
- $0.43 + 3.09(\text{MKT}) + 2.61(\text{IND}) + 1.70(\text{FORT})$ .
  - $0.681 + 0.021(\text{MKT}) + 0.04(\text{IND}) + 0.139(\text{FORT})$ .
  - $0.522 + 0.0460(\text{MKT}) + 0.7102(\text{IND}) + 0.9(\text{FORT})$ .
  - $1.21 + 0.015(\text{MKT}) + 0.2725(\text{IND}) + 0.5281(\text{FORT})$ .

**Topic 22**

**Cross Reference to GARP Assigned Reading – Stock & Watson, Chapter 6**

3. The expected amount of the stock return attributable to it being a Fortune 500 stock is closest to:
- A. 0.522.
  - B. 0.046.
  - C. 0.710.
  - D. 0.900.
4. Which of the following situations is not possible from the results of a multiple regression analysis with more than 50 observations?
- | <u>R<sup>2</sup></u> | <u>Adjusted R<sup>2</sup></u> |
|----------------------|-------------------------------|
| A. 71%               | 69%                           |
| B. 83%               | 86%                           |
| C. 54%               | 12%                           |
| D. 10%               | -2%                           |
5. Assumptions underlying a multiple regression are most likely to include:
- A. The expected value of the error term is  $0.00 < i < 1.00$ .
  - B. Linear and non-linear relationships exist between the dependent and independent variables.
  - C. The error for one observation is not correlated with that of another observation.
  - D. The variance of the error terms is not constant for all observations.

## CONCEPT CHECKER ANSWERS

1. C  $TSS = 1,025 + 925 = 1,950$
2. C The coefficients column contains the regression parameters.
3. D The regression equation is  $0.522 + 0.0460(MKT) + 0.7102(IND) + 0.9(FORT)$ . The coefficient on FORT is the amount of the return attributable to the stock of a Fortune 500 firm.
4. B Adjusted  $R^2$  must be less than or equal to  $R^2$ . Also, if  $R^2$  is low enough and the number of independent variables is large, adjusted  $R^2$  may be negative.
5. C Assumptions underlying a multiple regression include: the error for one observation is not correlated with that of another observation; the expected value of the error term is zero; a linear relationship exists between the dependent and independent variables; the variance of the error terms is constant.

---

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

# HYPOTHESIS TESTS AND CONFIDENCE INTERVALS IN MULTIPLE REGRESSION

---

Topic 23

## EXAM FOCUS

This topic addresses methods for dealing with uncertainty in a multiple regression model. Hypothesis tests and confidence intervals for single- and multiple-regression coefficients will be discussed. For the exam, you should know how to use a *t*-test to assess the significance of the individual regression parameters and an *F*-test to assess the effectiveness of the model as a whole in explaining the dependent variable. Also, be able to identify the common model misspecifications. Focus on interpretation of the regression equation and the test statistics. Remember that most of the test and descriptive statistics discussed (e.g., *t*-stat, *F*-stat, and  $R^2$ ) are provided in the output of statistical software. Hence, application and interpretation of these measurements are more likely than actual computations on the exam.

---

---

### LO 23.1: Construct, apply, and interpret hypothesis tests and confidence intervals for a single coefficient in a multiple regression.

---

#### Hypothesis Testing of Regression Coefficients

As with simple linear regression, the magnitude of the coefficients in a multiple regression tells us nothing about the importance of the independent variable in explaining the dependent variable. Thus, we must conduct hypothesis testing on the estimated slope coefficients to determine if the independent variables make a significant contribution to explaining the variation in the dependent variable.

The *t*-statistic used to test the significance of the individual coefficients in a multiple regression is calculated using the same formula that is used with simple linear regression:

$$t = \frac{b_j - B_j}{s_{b_j}} = \frac{\text{estimated regression coefficient} - \text{hypothesized value}}{\text{coefficient standard error of } b_j}$$

The *t*-statistic has  $n - k - 1$  degrees of freedom.



*Professor's Note: An easy way to remember the number of degrees of freedom for this test is to recognize that "k" is the number of regression coefficients in the regression, and the "1" is for the intercept term. Therefore, the degrees of freedom is the number of observations minus k minus 1.*

## Determining Statistical Significance

The most common hypothesis test done on the regression coefficients is to test statistical significance, which means testing the null hypothesis that the coefficient is zero versus the alternative that it is not:

$$\text{"testing statistical significance"} \Rightarrow H_0: b_j = 0 \text{ versus } H_A: b_j \neq 0$$

### Example: Testing the statistical significance of a regression coefficient

Consider again, from the previous topic, the hypothesis that future 10-year real earnings growth in the S&P 500 (EG10) can be explained by the trailing dividend payout ratio of the stocks in the index (PR) and the yield curve slope (YCS). Test the statistical significance of the independent variable PR in the real earnings growth example at the 10% significance level. Assume that the number of observations is 46. The results of the regression are reproduced in the following figure.

**Coefficient and Standard Error Estimates for Regression of EG10 on PR and YCS**

	<i>Coefficient</i>	<i>Standard Error</i>
Intercept	-11.6%	1.657%
PR	0.25	0.032
YCS	0.14	0.280

### Answer:

We are testing the following hypothesis:

$$H_0: PR = 0 \text{ versus } H_A: PR \neq 0$$

The 10% two-tailed critical *t*-value with  $46 - 2 - 1 = 43$  degrees of freedom is approximately 1.68. We should reject the null hypothesis if the *t*-statistic is greater than 1.68 or less than -1.68.

The *t*-statistic is:

$$t = \frac{0.25}{0.032} = 7.8$$

Therefore, because the *t*-statistic of 7.8 is greater than the upper critical *t*-value of 1.68, we can reject the null hypothesis and conclude that the PR regression coefficient is statistically significantly different from zero at the 10% significance level.

## Interpreting $p$ -Values

The  $p$ -value is the smallest level of significance for which the null hypothesis can be rejected. An alternative method of doing hypothesis testing of the coefficients is to compare the  $p$ -value to the significance level:

- If the  $p$ -value is less than significance level, the null hypothesis can be rejected.
- If the  $p$ -value is greater than the significance level, the null hypothesis cannot be rejected.

### Example: Interpreting $p$ -values

Given the following regression results, determine which regression parameters for the independent variables are statistically significantly different from zero at the 1% significance level, assuming the sample size is 60.

Variable	Coefficient	Standard Error	t-Statistic	p-Value
Intercept	0.40	0.40	1.0	0.3215
X1	8.20	2.05	4.0	0.0002
X2	0.40	0.18	2.2	0.0319
X3	-1.80	0.56	-3.2	0.0022

### Answer:

The independent variable is statistically significant if the  $p$ -value is less than 1%, or 0.01. Therefore  $X1$  and  $X3$  are statistically significantly different from zero.

Figure 1 shows the results of the  $t$ -tests for each of the regression coefficients of our 10-year earnings growth example, including the  $p$ -values.

Figure 1: Regression Results for Regression of EG10 on PR and YCS

	Coefficient	Standard Error	t-statistic	p-value
Intercept	-11.6%	1.657%	-7.0	< 0.0001
PR	0.25	0.032	7.8	< 0.0001
YCS	0.14	0.280	0.5	0.62

As we determined in a previous example, we can reject the null hypothesis and conclude that PR is statistically significant. We can also draw the same conclusion for the intercept term because -7.0 is less than the lower critical value of -1.68 (because it is a two-tailed test). However, we fail to reject the null hypothesis for YCS, so we cannot conclude that YCS has a statistically significant effect on the dependent variable, EG10, when PR is also included in the model. The  $p$ -values tell us exactly the same thing (as they always will): the

intercept term and PR are statistically significant at the 10% level because their *p*-values are less than 0.10, while YCS is not statistically significant because its *p*-value is greater than 0.10.

### Other Tests of the Regression Coefficients

You should also be prepared to formulate one- and two-tailed tests in which the null hypothesis is that the coefficient is equal to some value other than zero, or that it is greater than or less than some value.

#### Example: Testing regression coefficients (two-tail test)

Using the data from Figure 1, test the null hypothesis that PR is equal to 0.20 versus the alternative that it is not equal to 0.20 using a 5% significance level.

#### Answer:

We are testing the following hypothesis:

$$H_0: PR = 0.20 \text{ versus } H_A: PR \neq 0.20$$

The 5% two-tailed critical *t*-value with  $46 - 2 - 1 = 43$  degrees of freedom is approximately 2.02. We should reject the null hypothesis if the *t*-statistic is greater than 2.02 or less than -2.02.

The *t*-statistic is:

$$t = \frac{0.25 - 0.20}{0.032} = 1.56$$

Therefore, because the *t*-statistic of 1.56 is between the upper and lower critical *t*-values of -2.02 and 2.02, we cannot reject the null hypothesis and must conclude that the PR regression coefficient is not statistically significantly different from 0.20 at the 5% significance level.

**Example: Testing regression coefficients (one-tail test)**

Using the data from Figure 1, test the null hypothesis that the intercept term is greater than or equal to  $-10.0\%$  versus the alternative that it is less than  $-10.0\%$  using a 1% significance level.

**Answer:**

We are testing the following hypothesis:

$$H_0: \text{Intercept} \geq -10.0\% \text{ versus } H_A: \text{Intercept} < -10.0\%$$

The 1% one-tailed critical  $t$ -value with  $46 - 2 - 1 = 43$  degrees of freedom is approximately 2.42. We should reject the null hypothesis if the  $t$ -statistic is less than -2.42.

The  $t$ -statistic is:

$$t = \frac{-11.6\% - (-10.0\%)}{1.657\%} = -0.96$$

Therefore, because the  $t$ -statistic of -0.96 is not less than -2.42, we cannot reject the null hypothesis.

### Confidence Intervals for a Regression Coefficient

The confidence interval for a regression coefficient in multiple regression is calculated and interpreted the same way as it is in simple linear regression. For example, a 95% confidence interval is constructed as follows:

$$b_j \pm (t_c \times s_{b_j})$$

or

$$\text{estimated regression coefficient} \pm (\text{critical } t\text{-value})(\text{coefficient standard error})$$

The critical  $t$ -value is a two-tailed value with  $n - k - 1$  degrees of freedom and a 5% significance level, where  $n$  is the number of observations and  $k$  is the number of independent variables.

**Example: Calculating a confidence interval for a regression coefficient**

Calculate the 90% confidence interval for the estimated coefficient for the independent variable PR in the real earnings growth example.

**Answer:**

The critical  $t$ -value is 1.68, the same as we used in testing the statistical significance at the 10% significance level (which is the same thing as a 90% confidence level). The estimated slope coefficient is 0.25 and the standard error is 0.032. The 90% confidence interval is:

$$0.25 \pm (1.68)(0.032) = 0.25 \pm 0.054 = 0.196 \text{ to } 0.304$$



*Professor's Note: Notice that because zero is not contained in the 90% confidence interval, we can conclude that the PR coefficient is statistically significant at the 10% level. Constructing a confidence interval and conducting a  $t$ -test with a null hypothesis of "equal to zero" will always result in the same conclusion regarding the statistical significance of the regression coefficient.*

**PREDICTING THE DEPENDENT VARIABLE**

We can use the regression equation to make predictions about the dependent variable *based on forecasted values of the independent variables*. The process is similar to forecasting with simple linear regression, only now we need predicted values for more than one independent variable. The predicted value of dependent variable  $Y$  is:

$$\hat{Y}_i = b_0 + b_1 \hat{X}_{1i} + b_2 \hat{X}_{2i} + \dots + b_k \hat{X}_{ki}$$

where:

$\hat{Y}_i$  = the predicted value of the dependent variable

$b_j$  = the estimated slope coefficient for the  $j$ th independent variable

$\hat{X}_{ji}$  = the forecast of the  $j$ th independent variable,  $j = 1, 2, \dots, k$

*Professor's Note: The prediction of the dependent variable uses the estimated intercept and all of the estimated slope coefficients, regardless of whether the estimated coefficients are statistically significantly different from zero.*

*For example, suppose you estimate the following regression equation:*



$\hat{Y} = 6 + 2X_1 + 4X_2$ , and you determine that only the first independent variable ( $X_1$ ) is statistically significant (i.e., you rejected the null that  $B_1 = 0$ ). To predict  $Y$  given forecasts of  $X_1 = 0.6$  and  $X_2 = 0.8$ , you would use the complete model:  $\hat{Y} = 6 + (2 \times 0.6) + (4 \times 0.8) = 10.4$ . Alternatively, you could drop  $X_2$  and reestimate the model using just  $X_1$ , but remember that the coefficient on  $X_1$  will likely change.

**Example: Calculating a predicted value for the dependent variable**

An analyst would like to use the estimated regression equation from the previous example to calculate the predicted 10-year real earnings growth for the S&P 500, assuming the payout ratio of the index is 50%. He observes that the slope of the yield curve is currently 4%.

**Answer:**

$$\widehat{EG10} = -11.6\% + 0.25(50\%) + 0.14(4\%) = 1.46\%$$

The model predicts a 1.46% real earnings growth rate for the S&P 500, assuming a 50% payout ratio, when the slope of the yield curve is 4%.

## JOINT HYPOTHESIS TESTING

---

**LO 23.2: Construct, apply, and interpret joint hypothesis tests and confidence intervals for multiple coefficients in a multiple regression.**

**LO 23.3: Interpret the F-statistic.**

**LO 23.5: Interpret confidence sets for multiple coefficients.**

---

A joint hypothesis tests two or more coefficients at the same time. For example, we could develop a null hypothesis for a linear regression model with three independent variables that sets two of these coefficients equal to zero:  $H_0: b_1 = 0$  and  $b_2 = 0$  versus the alternative hypothesis that one of them is not equal to zero. That is, if just one of the equalities in this null hypothesis does not hold, we can reject the entire null hypothesis. Using a joint hypothesis test is preferred in certain scenarios since testing coefficients individually leads to a greater chance of rejecting the null hypothesis. For example, instead of comparing one t-statistic to its corresponding critical value in a joint hypothesis test, we are testing two t-statistics. Thus, we have an additional opportunity to reject the null. A robust method for applying joint hypothesis testing, especially when independent variables are correlated, is known as the *F*-statistic.

## THE *F*-STATISTIC

An *F*-test assesses how well the set of independent variables, as a group, explains the variation in the dependent variable. That is, the *F*-statistic is used to test whether *at least one* of the independent variables explains a significant portion of the variation of the dependent variable.

For example, if there are four independent variables in the model, the hypotheses are structured as:

$$H_0: B_1 = B_2 = B_3 = B_4 = 0 \text{ versus } H_A: \text{at least one } B_j \neq 0$$

The *F*-statistic, *which is always a one-tailed test*, is calculated as:

$$\frac{\frac{\text{ESS}}{k}}{\frac{\text{SSR}}{n - k - 1}}$$

where:

ESS = explained sum of squares

SSR = sum of squared residuals



*Professor's Note: The explained sum of squares and the sum of squared residuals are found in an analysis of variance (ANOVA) table. We will analyze an ANOVA table from a multiple regression shortly.*

To determine whether at least one of the coefficients is statistically significant, the calculated *F*-statistic is compared with the **one-tailed** critical *F*-value,  $F_c$ , at the appropriate level of significance. The degrees of freedom for the numerator and denominator are:

$$df_{\text{numerator}} = k$$

$$df_{\text{denominator}} = n - k - 1$$

where:

n = number of observations

k = number of independent variables

The decision rule for the *F*-test is:

Decision rule: reject  $H_0$  if  $F$  (test-statistic) >  $F_c$  (critical value)

Rejection of the null hypothesis at a stated level of significance indicates that at least one of the coefficients is significantly different than zero, which is interpreted to mean that at least one of the independent variables in the regression model makes a significant contribution to the explanation of the dependent variable.



*Professor's Note: It may have occurred to you that an easier way to test all of the coefficients simultaneously is to just conduct all of the individual t-tests and see how many of them you can reject. This is the wrong approach, however, because if you set the significance level for each t-test at 5%, for example, the significance level from testing them all simultaneously is NOT 5%, but rather some higher percentage. Just remember to use the F-test on the exam if you are asked to test all of the coefficients simultaneously.*

**Example: Calculating and interpreting the *F*-statistic**

An analyst runs a regression of monthly value-stock returns on five independent variables over 60 months. The total sum of squares is 460, and the sum of squared residuals is 170. Test the null hypothesis at the 5% significance level (95% confidence) that all five of the independent variables are equal to zero.

**Answer:**

The null and alternative hypotheses are:

$$H_0: B_1 = B_2 = B_3 = B_4 = B_5 = 0 \text{ versus } H_A: \text{at least one } B_j \neq 0$$

$$\text{ESS} = \text{TSS} - \text{SSR} = 460 - 170 = 290$$

$$F = \frac{58.0}{3.15} = 18.41$$

The critical *F*-value for 5 and 54 degrees of freedom at a 5% significance level is approximately 2.40. Remember, it's a **one-tailed test**, so we use the 5% *F*-table! Therefore, we can reject the null hypothesis and conclude that at least one of the five independent variables is significantly different than zero.



*Professor's Note: When testing the hypothesis that all the regression coefficients are simultaneously equal to zero, the F-test is always a one-tailed test, despite the fact that it looks like it should be a two-tailed test because there is an equal sign in the null hypothesis.*

## INTERPRETING REGRESSION RESULTS

Just as in simple linear regression, the variability of the dependent variable or **total sum of squares** (TSS) can be broken down into **explained sum of squares** (ESS) and **sum of squared residuals** (SSR). As shown previously, the coefficient of determination is:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\text{SSR}}{\text{TSS}} = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

Regression results usually provide  $R^2$  and a host of other measures. However, it is useful to know how to compute  $R^2$  from other parts of the results. Figure 2 is an ANOVA table of the results of a regression of hedge fund returns on lockup period and years of experience of the manager. In the ANOVA table, the value of 90 represents TSS, the ESS equals 84.057, and the SSR is 5.943. Although the output results provide the value  $R^2 = 0.934$ , it can also be computed using TSS, ESS, and SSR like so:

$$R^2 = \frac{84.057}{90} = 1 - \frac{5.943}{90} = 0.934$$

The coefficient of multiple correlation is simply the square root of  $R$ -squared. In the case of a multiple regression, the coefficient of multiple correlation is always positive.

Figure 2: ANOVA Table

$R$ -squared	0.934						
Adj $R$ -squared	0.890						
Standard error	1.407						
Observations	6						
<i>Degrees of Freedom</i>							
		<i>SS</i>	<i>MS</i>	<i>F</i>			
Explained	2	84.057	42.029	21.217			
Residual	3	5.943	1.981				
Total	5	90					
<i>Variables</i>		<i>Coeff</i>	<i>Std Error</i>	<i>t-stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept		-4.4511	3.299	-1.349	0.270	-14.950	6.048
Lockup		2.057	0.337	6.103	0.009	0.984	3.130
Experience		2.008	0.754	2.664	0.076	-0.391	4.407

The results in Figure 2 produce the following equation:

$$\hat{Y}_i = -4.451 + 2.057 \times X_{1i} + 2.008 \times X_{2i}$$

This equation tells us that holding other variables constant, increasing the lockup period will increase the expected return of a hedge fund by 2.057%. Also, holding other variables constant, increasing the manager's experience one year will increase the expected return of a hedge fund by 2.008%. A hedge fund with an inexperienced manager and no lockup period will earn a negative return of -4.451%.

The ANOVA table outputs the standard errors, *t*-statistics, probability values (*p*-values), and confidence intervals for the estimated coefficients. These can be used in a hypothesis test for each coefficient. For example, for the independent variable experience ( $b_2$ ), the output indicates that the standard error is  $se(b_2) = 0.754$ , which yields a *t*-statistic of:  $2.008 / 0.754 = 2.664$ . The critical *t*-value at a 5% level of significance is  $t_{0.025} = 3.182$ . Thus, a hypothesis stating that the number of years of experience is not related to returns could not be rejected. In other words, the result is to not reject the null hypothesis that  $B_2 = 0$ . This is also seen with the provided confidence interval. Upper and lower limits of the confidence interval can be found in the ANOVA results.

$$[b_2 - t_{\alpha/2} \times se(b_2)] < B_2 < [b_2 + t_{\alpha/2} \times se(b_2)]$$

$$(2.008 - 3.182 \times 0.754) < B_2 < (2.008 + 3.182 \times 0.754)$$

$$-0.391 < B_2 < 4.407$$

Since the confidence interval contains the value zero, then the null hypothesis:  $H_0: B_2 = 0$  cannot be rejected in a two-tailed test at the 5% level of significance. Figure 2 provides a third way of performing a hypothesis test by providing a *p*-value. The *p*-value indicates the

minimum level of significance at which the two-tailed hypothesis test can be rejected. In this case, the p-value is 0.076 (i.e., 7.6%), which is greater than 5%.

The statistics for  $b_1$  indicate that a null hypothesis can be rejected at a 5% level using a two-tailed test. The  $t$ -statistic is 6.103, and the confidence interval is 0.984 to 3.13. The p-value of 0.9% is less than 5%.

The statistics in the ANOVA table also allow for the testing of the joint hypothesis that both slope coefficients equal zero.

$$H_0: B_1 = B_2 = 0$$

$$H_A: B_1 \neq 0 \text{ or } B_2 \neq 0$$

The test statistic in this case is the  $F$ -statistic where the degrees of freedom are indicated by two numbers: the number of slope coefficients (2) and the sample size minus the number of slope coefficients minus one ( $6 - 2 - 1 = 3$ ). The  $F$ -statistic given the hedge fund data can be calculated as follows:

$$F = \frac{\text{ESS}/\text{df}}{\text{SSR}/\text{df}} = \frac{84.057/2}{5.943/3} = \frac{42.029}{1.981} = 21.217$$

The critical  $F$ -statistic at a 5% significance level is  $F_{0.05} = 9.55$ . Since the value from the regression results is greater than that value:  $F = 21.217 > 9.55$ , a researcher would reject the null hypothesis:  $H_0: B_1 = B_2 = 0$ . It should be noted that rejecting the null hypothesis indicates one or both of the coefficients are significant.

## SPECIFICATION BIAS

Specification bias refers to how the slope coefficient and other statistics for a given independent variable are usually different in a simple regression when compared to those of the same variable when included in a multiple regression. To illustrate this point, the following three OLS results correspond to a two-variable regression using only the indicated independent variable and the results for a three-variable:

$$\hat{Y}_i = 1 + 2 \times (\text{lockup})_i \\ t = 3.742$$

$$\hat{Y}_i = 11.714 + 1.714 \times (\text{experience})_i \\ t = 2.386$$

$$\hat{Y}_i = -4.451 + 2.057 \times (\text{lockup})_i + 2.008 \times (\text{experience})_i \\ t = 6.103 \quad t = 2.664$$

Specification bias is indicated by the extent to which the coefficient for each independent variable is different when compared across equations (e.g., for lockup, the slope is 2 in the two-variable equation, and the slope is 2.057 in the multivariate regression). This is because in the two-variable regression, the slope coefficient includes the effect of the included independent variable in the equation and, to some extent, the indirect effect of the excluded

variable(s). In this case, the bias for the coefficient on the lockup coefficient was not large because the experience variable was not significant as indicated in its two-variable regression ( $t = 2.386 < t_{0.025} = 2.78$ ) and was not significant in the multivariable regression either.

## R<sup>2</sup> AND ADJUSTED R<sup>2</sup>

### LO 23.7: Interpret the R<sup>2</sup> and adjusted R<sup>2</sup> in a multiple regression.

To further analyze the importance of an added variable to a regression, we can compute an adjusted coefficient of determination, or **adjusted R<sup>2</sup>**. The reason adjusted R<sup>2</sup> is important is because, mathematically speaking, the coefficient of determination, R<sup>2</sup>, must go up if a variable with any explanatory power is added to the regression, even if the marginal contribution of the new variables is not statistically significant. Consequently, a relatively high R<sup>2</sup> may reflect the impact of a large set of independent variables rather than how well the set explains the dependent variable. This problem is often referred to as overestimating the regression.

When computing both the R<sup>2</sup> and the adjusted R<sup>2</sup>, there are a few pitfalls to acknowledge, which could lead to invalid conclusions.

1. If adding an additional independent variable to the regression improves the R<sup>2</sup>, this variable is not necessarily statistically significant.
2. The R<sup>2</sup> measure may be spurious, meaning that the independent variables may show a high R<sup>2</sup>; however, they are not the exact cause of the movement in the dependent variable.
3. If the R<sup>2</sup> is high, we cannot assume that we have found all relevant independent variables. Omitted variables may still exist, which would improve the regression results further.
4. The R<sup>2</sup> measure does not provide evidence that the most or least appropriate independent variables have been selected. Many factors go into finding the most robust regression model, including omitted variable analysis, economic theory, and the quality of data being used to generate the model.

## RESTRICTED VS. UNRESTRICTED LEAST SQUARES MODELS

A restricted least squares regression imposes a value on one or more coefficients with the goal of analyzing if the restriction is significant. To explain this concept, it is useful to note that there is an implied restriction in each of the two variable regressions:

$$\begin{aligned}\hat{Y}_i &= b_0 + b_{\text{lockup}} \times (\text{lockup})_i \\ \hat{Y}_i &= b_0 + b_{\text{experience}} \times (\text{experience})_i\end{aligned}$$

In essence, each of the two-variable regressions is a restricted regression where the coefficient on the omitted variable is restricted to zero. To help illustrate the concept, the more elaborate subscripts have been used in these expressions. Using the indicated notation, the first specification that only includes "lockup" is restricting  $b_{\text{experience}}$  to 0. In the unrestricted

multivariable regression, both  $b_{\text{lockup}}$  and  $b_{\text{experience}}$  are allowed to assume the values that minimize the SSR. The  $R^2$  from the restricted regression is called a **restricted  $R^2$**  or  $R_r^2$ . For comparison, the **unrestricted  $R^2$**  from the specification that includes both independent variables is given the notation  $R_{\text{ur}}^2$ , and both are included in an  $F$ -statistic that can test if the restriction is significant or not:

$$F = \frac{(R_{\text{ur}}^2 - R_r^2)/m}{(1 - R_{\text{ur}}^2)/(n - k_{\text{ur}} - 1)}$$

The symbol “ $m$ ” refers to the number of restrictions, which in the example discussed would be equal to one. This  $F$ -stat is known as the **homoskedasticity-only  $F$ -statistic** since it can only be derived from  $R^2$  when the error terms display homoskedasticity. An alternative formula for computing this  $F$ -stat is to use the sum of squared residuals in place of the  $R^2$ :

$$F = \frac{(SSR_{\text{ur}} - SSR_r)/m}{SSR_{\text{ur}}/(n - k_{\text{ur}} - 1)}$$

In the event that the error terms are not homoskedastic, a heteroskedasticity-robust  $F$ -stat would be applied. This statistic is used more frequently in practice; however, as the sample size,  $n$ , increases, these two types of  $F$ -statistics will converge.

#### LO 23.4: Interpret tests of a single restriction involving multiple coefficients.

With the  $F$ -statistic, we constructed a null hypothesis that tested multiple coefficients being equal to zero. However, what if we wanted to test whether one coefficient was equal to another such that:  $H_0: b_1 = b_2$ ? The alternative hypothesis in this scenario would be that the two are not equal to each other. Hypothesis tests of single restrictions involving multiple coefficients requires the use of statistical software packages, but we will examine the methodology of two different approaches.

The first approach is to directly test the restriction stated in the null. Some statistical packages can test this restriction and output a corresponding  $F$ -stat. This is the easier of the two methods; however, a second method will need to be applied if your statistical package cannot directly test the restriction.

The second approach transforms the regression and uses the null hypothesis as an assumption to simplify the regression model. For example, in a regression with two independent variables:  $Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \epsilon_i$ , we can add and subtract  $B_2 X_{1i}$  to ultimately transform the regression to:  $B_0 + (B_1 - B_2)X_{1i} + B_2(X_{1i} + X_{2i}) + \epsilon_i$ . One of the coefficients will drop out in this equation when assuming that the null hypothesis of  $B_1 = B_2$  is valid. We can remove the second term from our regression equation so that:  $B_0 + B_2(X_{1i} + X_{2i}) + \epsilon_i$ . We observe that the null hypothesis test changes from a single restriction involving multiple coefficients to a single restriction on just one coefficient.



*Professor's Note: Remember that this process is typically done with statistical software packages, so on the exam, you would simply be asked to describe and/or interpret these tests.*

## MODEL MISSPECIFICATION

### LO 23.6: Identify examples of omitted variable bias in multiple regressions.

Recall from the previous topic that omitting relevant factors from a regression can produce misleading or biased results. Similar to simple linear regression, omitted variable bias in multiple regressions will result if the following two conditions occur:

- The omitted variable is a determinant of the dependent variable.
- The omitted variable is correlated with *at least* one of the independent variables.

As an example of omitted variable bias, consider a regression in which we're trying to predict monthly returns on portfolios of stocks ( $R$ ) using three independent variables: portfolio beta ( $B$ ), the natural log of market capitalization ( $\ln M$ ), and the natural log of the price-to-book ratio  $\ln(PB)$ . The correct specification of this model is as follows:

$$R = b_0 + b_1 B + b_2 \ln M + b_3 \ln PB + \varepsilon$$

Now suppose we did not include  $\ln M$  in the regression model:

$$R = a_0 + a_1 B + a_2 \ln PB + \varepsilon$$

If  $\ln M$  is correlated with any of the remaining independent variables ( $B$  or  $\ln PB$ ), then the error term is also correlated with the same independent variables, and the resulting regression coefficients are biased and inconsistent. That means our hypothesis tests and predictions using the model will be unreliable.

## KEY CONCEPTS

### LO 23.1

A *t*-test is used for hypothesis testing of regression parameter estimates:

$$t = \frac{b_j - B_j}{s_{b_j}}, \text{ with } n - k - 1 \text{ degrees of freedom}$$

Testing for statistical significance means testing  $H_0: B_j = 0$  vs.  $H_A: B_j \neq 0$ .

---

### LO 23.2

The confidence interval for regression coefficient is:

$$\text{estimated regression coefficient} \pm (\text{critical } t\text{-value})(\text{coefficient standard error})$$

The value of dependent variable Y is predicted as:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

---

### LO 23.3

The *F*-distributed test statistic can be used to test the significance of all (or any subset of) the independent variables (i.e., the overall fit of the model) using a one-tailed test:

$$F = \frac{\text{ESS}/k}{\text{SSR}/[n - k - 1]} \text{ with } k \text{ and } n - k - 1 \text{ degrees of freedom}$$

---

### LO 23.4

Hypothesis tests of single restrictions involving multiple coefficients requires the use of statistical software packages.

### LO 23.5

The ANOVA table outputs the standard errors, t-statistics, probability values (*p*-values), and confidence intervals for the estimated coefficients.

Upper and lower limits of the confidence interval can be found in the ANOVA results.

$$[b_2 - t_{\alpha/2} \times se(b_2)] < B_2 < [b_2 + t_{\alpha/2} \times se(b_2)]$$

The statistics in the ANOVA table also allow for the testing of the joint hypothesis that both slope coefficients equal zero.

$$H_0: B_1 = B_2 = 0$$

$$H_A: B_1 \neq 0 \text{ or } B_2 \neq 0$$

The test statistic in this case is the *F*-statistic.

---

### LO 23.6

Omitting a relevant independent variable in a multiple regression results in regression coefficients that are biased and inconsistent, which means we would not have any confidence in our hypothesis tests of the coefficients or in the predictions of the model.

---

### LO 23.7

Restricted least squares models restrict one or more of the coefficients to equal a given value and compare the  $R^2$  of the restricted model to that of the unrestricted model where the coefficients are not restricted. An *F*-statistic can test if there is a significant difference between the restricted and unrestricted  $R^2$ .

## CONCEPT CHECKERS

Use the following table for Question 1.

Source	Sum of Squares (SS)	Degrees of Freedom
Explained	1,025	5
Residual	925	25

1. The  $R^2$  and the  $F$ -statistic, respectively, are closest to:

$$\begin{array}{ll} \underline{R^2} & \underline{F\text{-statistic}} \end{array}$$

- A. 53%      1.1
- B. 47%      1.1
- C. 53%      5.5
- D. 47%      5.5

Use the following information to answer Question 2.

An analyst calculates the sum of squared residuals and total sum of squares from a multiple regression with four independent variables to be 4,320 and 9,105, respectively. There are 65 observations in the sample.

2. The critical  $F$ -value for testing  $H_0 = B_1 = B_2 = B_3 = B_4 = 0$  vs.  $H_A$ : at least one  $B_j \neq 0$  at the 5% significance level is closest to:
- A. 2.37.
  - B. 2.53.
  - C. 2.76.
  - D. 3.24.
3. When interpreting the  $R^2$  and adjusted  $R^2$  measures for a multiple regression, which of the following statements incorrectly reflects a pitfall that could lead to invalid conclusions?
- A. The  $R^2$  measure does not provide evidence that the most or least appropriate independent variables have been selected.
  - B. If the  $R^2$  is high, we have to assume that we have found all relevant independent variables.
  - C. If adding an additional independent variable to the regression improves the  $R^2$ , this variable is not necessarily statistically significant.
  - D. The  $R^2$  measure may be spurious, meaning that the independent variables may show a high  $R^2$ ; however, they are not the exact cause of the movement in the dependent variable.

Use the following information for Questions 4 and 5.

Phil Ohlmer estimates a cross sectional regression in order to predict price to earnings ratios (P/E) with fundamental variables that are related to P/E, including dividend payout ratio (DPO), growth rate (G), and beta (B). In addition, all 50 stocks in the sample come from two industries, electric utilities or biotechnology. He defines the following dummy variable:

IND = 0 if the stock is in the electric utilities industry, or  
= 1 if the stock is in the biotechnology industry

The results of his regression are shown in the following table.

<i>Variable</i>	<i>Coefficient</i>	<i>t-Statistic</i>
Intercept	6.75	3.89*
IND	8.00	4.50*
DPO	4.00	1.86
G	12.35	2.43*
B	-0.50	1.46

\*significant at the 5% level

4. Based on these results, it would be most appropriate to conclude that:
  - A. biotechnology industry PEs are statistically significantly larger than electric utilities industry PEs.
  - B. electric utilities PEs are statistically significantly larger than biotechnology industry PEs, holding DPO, G, and B constant.
  - C. biotechnology industry PEs are statistically significantly larger than electric utilities industry PEs, holding DPO, G, and B constant.
  - D. the dummy variable does not display statistical significance.
5. Ohlmer is valuing a biotechnology stock with a dividend payout ratio of 0.00, a beta of 1.50, and an expected earnings growth rate of 0.14. The predicted P/E on the basis of the values of the explanatory variables for the company is closest to:
  - A. 7.7.
  - B. 15.7.
  - C. 17.2.
  - D. 11.3.

## CONCEPT CHECKER ANSWERS

1. C  $R^2 = \frac{ESS}{TSS} = \frac{1,025}{1,950} = 53\%$
- $$F = \frac{\frac{ESS}{df}}{\frac{SSR}{df}} = \frac{\frac{1,025}{5}}{\frac{925}{25}} = \frac{205}{37} = 5.5$$
2. B This is a one-tailed test, so the critical  $F$ -value at the 5% significance level with 4 and 60 degrees of freedom is approximately 2.53.
3. B If the  $R^2$  is high, we *cannot* assume that we have found all relevant independent variables. Omitted variables may still exist, which would improve the regression results further.
4. C The  $t$ -statistic tests the null that industry PEs are equal. The dummy variable is significant and positive, and the dummy variable is defined as being equal to one for biotechnology stocks, which means that biotechnology PEs are statistically significantly larger than electric utility PEs. Remember, however, this is only accurate if we hold the other independent variables in the model constant.
5. B Note that  $IND = 1$  because the stock is in the biotech industry. Predicted P/E =  $6.75 + (8.00 \times 1) + (4.00 \times 0.00) + (12.35 \times 0.14) - (0.50 \times 1.5) = 15.7$ .

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

# MODELING AND FORECASTING TREND

## Topic 24

### EXAM FOCUS

A trend model captures a time series pattern and allows us to make predictions about a variable in the future. This topic focuses on selecting the best forecasting model to estimate a trend. For the exam, be able to describe the differences between linear and nonlinear trends. Also, understand how mean squared error (MSE) is calculated and how adjusting for degrees of freedom,  $k$ , is accomplished with the unbiased MSE (or  $s^2$ ), Akaike information criterion (AIC), and Schwarz information criterion (SIC). Finally, be able to explain how selection tools compare based on penalty factors and the consistency property.

### LINEAR AND NONLINEAR TRENDS

#### LO 24.1: Describe linear and nonlinear trends.

A *time series* is a set of observations for a variable over successive periods of time (e.g., monthly stock market returns for the past 10 years). The series has a **trend** if a consistent pattern can be seen by plotting the data (i.e., the individual observations) on a graph. A trend in finance or economics can be illustrated with a slow evolution of variables, such as demographics or technologies, over a long time horizon. In this topic, we focus on **deterministic trends**, which are trends that evolve in an expected fashion.

#### Linear Trend Models

A *linear trend* is a time series pattern that can be graphed using a straight line. The simplest form of a linear trend is represented by the following model:

$$y_t = \beta_0 + \beta_1(t)$$

where:

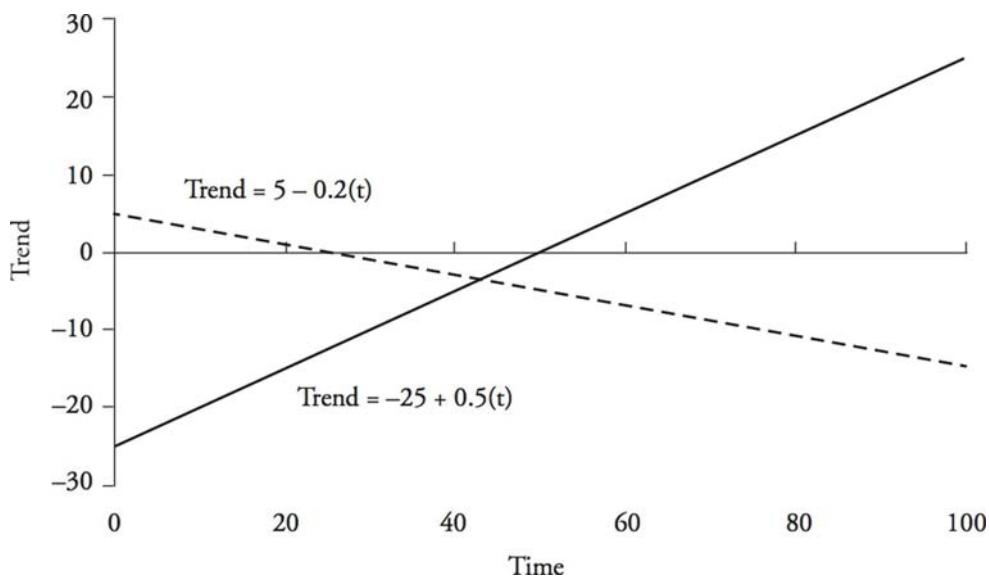
$y_t$  = the value of the time series (the dependent variable at time  $t$ )

$\beta_0$  = regression intercept at the vertical axis

$\beta_1$  = regression slope coefficient (or trend coefficient)

$t$  = time trend or time dummy (the independent variable):  $t = 1, 2, 3, \dots, T - 1, T$

A downward-sloping line (i.e., negative slope coefficient) indicates a negative trend, while an upward-sloping line (i.e., a positive slope coefficient) indicates a positive trend. The steepness of the trend will depend on the magnitude of the slope coefficient. A higher  $\beta_1$  in absolute value terms (e.g., 0.5) indicates a steeper slope, while a lower  $\beta_1$  (e.g., 0.2) indicates a gentler slope. Figure 1 illustrates downward- and upward-sloping linear trends with different levels of steepness.

**Figure 1: Linear Trends**

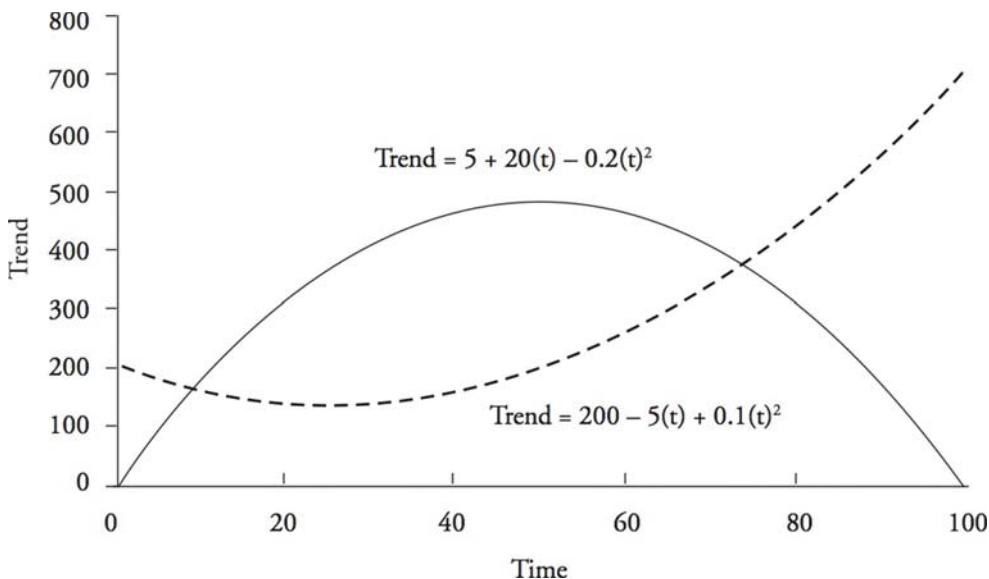
### Nonlinear Trend Models

A *nonlinear trend* is a time series pattern that can be graphed with a curve. For example, a nonlinear trend would result if a variable increases at an increasing rate. When estimating and forecasting trends, a trend is not required to be linear; however, it should exhibit a smooth pattern. Nonlinear trends can be modeled using either quadratic or exponential functions.

As mentioned, a possible way to capture nonlinearities is to use a **quadratic trend** as follows:

$$y_t = \beta_0 + \beta_1(t) + \beta_2(t)^2$$

This function can model various trends by adjusting the sign and level of the coefficients. For example, when both  $\beta_1$  and  $\beta_2$  are positive, the trend increases at an increasing rate over time. Conversely, when both  $\beta_1$  and  $\beta_2$  are negative, the trend decreases at an increasing rate over time. When  $\beta_1$  is negative and  $\beta_2$  is positive, the trend will resemble a “U” shape. Finally, when  $\beta_1$  is positive and  $\beta_2$  is negative, the trend will resemble an “inverted U” shape. Note that U-shaped trends are rare when modeling financial data because most of the data in a time series typically falls on one side of the U. Figure 2 illustrates quadratic trends with different signs and levels for coefficients.

**Figure 2: Quadratic Trends**

While quadratic trends may be adequate for modeling some nonlinear trends, other trends may be better approximated using an **exponential trend**. In particular, financial time series often display exponential growth (i.e., growth with continuous compounding). Positive exponential growth means that the random variable (i.e., the time series) tends to increase at some constant rate of growth (e.g., 2% per year). If we plot the data, the observations will form a convex curve. Negative exponential growth means that the data tends to decrease at some constant rate of decay, and the plotted time series will be a concave curve.

When a series exhibits exponential growth, it can be modeled using an exponential trend as follows:

$$y_t = \beta_0 e^{\beta_1(t)}$$

where:

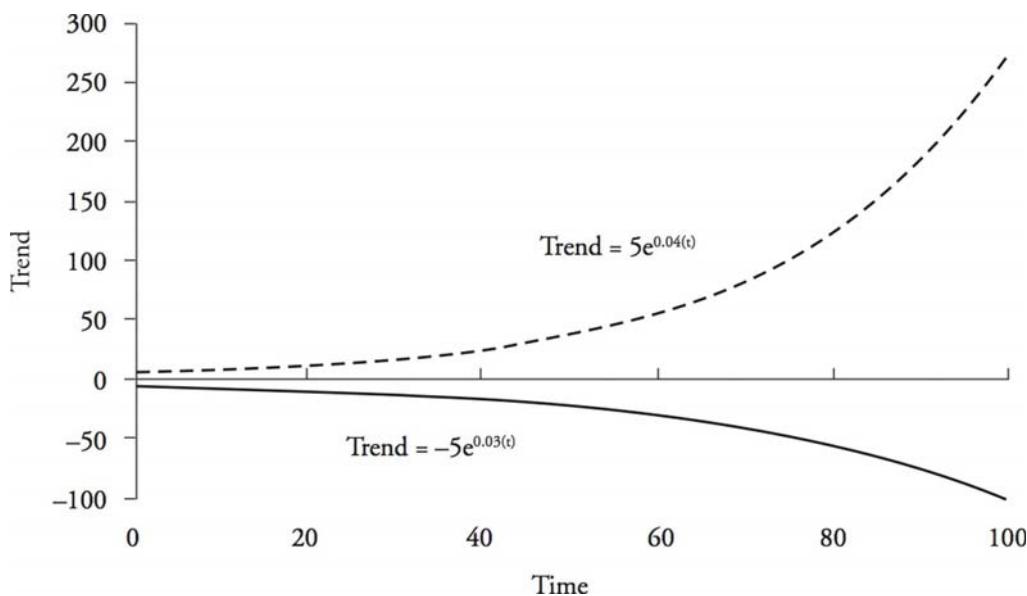
$y_t$  = the value of the dependent variable at time  $t$

$\beta_0$  = regression intercept term

$\beta_1$  = the constant rate of growth

$t$  = time:  $t = 1, 2, 3, \dots, T - 1, T$

As with quadratic trends, varying the signs and levels of the coefficients will create different patterns. The trend can increase or decrease at either an increasing or decreasing rate.

**Figure 3: Exponential Trends**

This nonlinear trend model defines  $y$ , the dependent variable, as an exponential function of time, the independent variable. Rather than try to fit the nonlinear data with a linear (straight line) regression, we take the natural log of both sides of the equation and arrive at the **log-linear trend** as follows:

$$\ln(y_t) = \ln(\beta_0) + \beta_1(t)$$

Now that the equation has been transformed from an exponential function to a linear function, we can use a linear regression technique to model the series. The use of the transformed data produces a linear trend line with a better fit for the data, which increases the predictive ability of the model.

## ESTIMATING AND FORECASTING TRENDS

### **LO 24.2: Describe trend models to estimate and forecast trends.**

**Ordinary least squares (OLS) regression** is used to estimate the coefficients in a trend line. It is calculated using the following prediction equation:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1(t)$$

where:

$\hat{y}_t$  = the predicted value of  $y$  (the dependent variable) at time  $t$

$\hat{\beta}_0$  = the estimated value of the intercept term

$\hat{\beta}_1$  = the estimated value of the slope coefficient

Recall that with trend models,  $t$  takes on the value of the time period. For example, in period 2, the equation becomes the following:

$$\hat{y}_2 = \hat{\beta}_0 + \hat{\beta}_1(2)$$

And, likewise, in period 3 the equation is as follows:

$$\hat{y}_3 = \hat{\beta}_0 + \hat{\beta}_1(3)$$

This means  $\hat{y}$  increases by the value of  $\hat{\beta}_1$  each period.

#### Example: Using a linear trend model

Suppose you are given a linear trend model with  $\hat{\beta}_0 = 1.7$  and  $\hat{\beta}_1 = 3.0$ .

Calculate  $\hat{y}_t$  for  $t = 1$  and  $t = 2$ .

**Answer:**

When  $t = 1$ ,  $\hat{y}_1 = 1.7 + 3.0(1) = 4.7$ .

When  $t = 2$ ,  $\hat{y}_2 = 1.7 + 3.0(2) = 7.7$ .

Notice that the difference between  $\hat{y}_1$  and  $\hat{y}_2$  is 3.0, the value of the trend coefficient  $\hat{\beta}_1$ .

#### Example: Trend analysis

Consider hypothetical time series data for manufacturing capacity utilization.

#### Manufacturing Capacity Utilization

Quarter	Time ( $t$ )	Manufacturing Capacity Utilization (in %)	Quarter	Time ( $t$ )	Manufacturing Capacity Utilization (in %)
2013.1	1	82.4	2014.4	8	80.9
2013.2	2	81.5	2015.1	9	81.3
2013.3	3	80.8	2015.2	10	81.9
2013.4	4	80.5	2015.3	11	81.7
2014.1	5	80.2	2015.4	12	80.3
2014.2	6	80.2	2016.1	13	77.9
2014.3	7	80.5	2016.2	14	76.4

Applying the OLS methodology to fit the linear trend model to the data produces the following results.

### Time Series Regression Results for Manufacturing Capacity Utilization

Regression model: $y_t = \beta_0 + \beta_1 t + \epsilon_t$			
R square	0.346	Standard Error	t-Statistic
Adjusted R square	0.292		
Standard error	1.334		
Observations	14		
	Coefficients	Standard Error	t-Statistic
Intercept	82.137	0.753	109.066
Manufacturing utilization	-0.223	0.088	-2.534

Based on this information, predict the projected capacity utilization for the time period involved in the study (i.e., in-sample estimates).

#### Answer:

As shown in the regression output, the estimated intercept and slope parameters for our manufacturing capacity utilization model are  $\hat{\beta}_0 = 82.137$  and  $\hat{\beta}_1 = -0.223$ , respectively. This means that the prediction equation for capacity utilization can be expressed as:

$$\hat{y}_t = 82.137 - 0.223t$$

With this equation, we can generate estimated values for capacity utilization,  $\hat{y}_t$ , for each of the 14 quarters in the time series. For example, using the model capacity utilization for the first quarter of 2013 is estimated at 81.914:

$$\hat{y}_t = 82.137 - 0.223(1) = 82.137 - 0.223 = 81.914$$

Note that the estimated value of capacity utilization in that quarter (using the model) is not exactly the same as the actual, measured capacity utilization for that quarter (82.4). The difference between the two is the error or residual term associated with that observation:

$$\text{residual (error)} = \text{actual value} - \text{predicted value} \approx 82.4 - 81.914 = 0.486$$

Note that since the actual, measured value is greater than the predicted value of  $y$  for 2013.1, the error term is positive. Had the actual, measured value been less than the predicted value, the error term would have been negative.

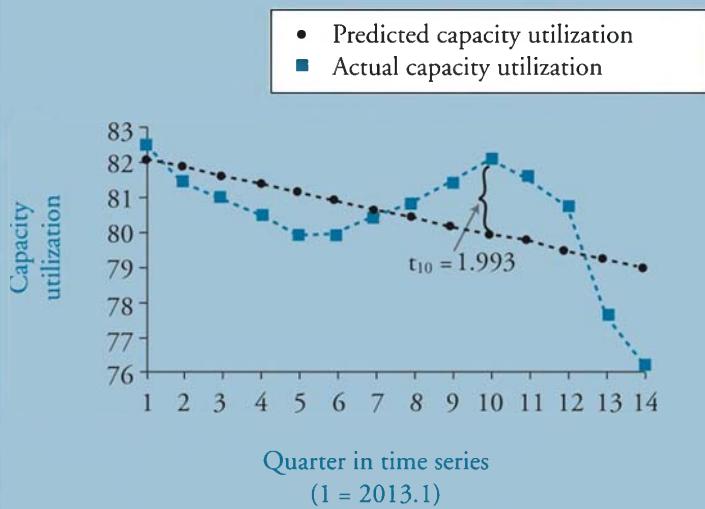
The projections (i.e., values generated by the model) for all quarters are compared to the actual values below.

### Projected vs. Actual Capacity Utilization

Quarter	Time	$\hat{y}_t$	$y_t$	Quarter	Time	$\hat{y}_t$	$y_t$
2013.1	1	81.914	82.4	2014.4	8	80.353	80.9
2013.2	2	81.691	81.5	2015.1	9	80.130	81.3
2013.3	3	81.468	80.8	2015.2	10	79.907	81.9
2013.4	4	81.245	80.5	2015.3	11	79.684	81.7
2014.1	5	81.022	80.2	2015.4	12	79.460	80.3
2014.2	6	80.799	80.2	2016.1	13	79.237	77.9
2014.3	7	80.576	80.5	2016.2	14	79.014	76.4

The following graph shows visually how the predicted values compare to the actual values, which were used to generate the regression equation. The **residuals**, or **error terms**, are represented by the distance between the predicted (straight) regression line and the actual data plotted in blue. For example, the residual for  $t = 10$  is  $81.9 - 79.907 = 1.993$ .

### Predicted vs. Actual Capacity Utilization



Since we utilized a linear regression model, the predicted values will by definition fall on a straight line. Since the raw data does not display a linear relationship, the model will probably not do a good job of predicting future values.

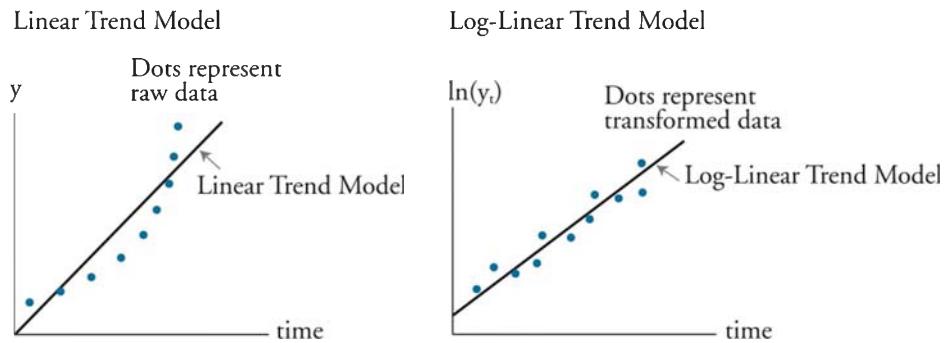
## SELECTING THE CORRECT TREND MODEL

To determine if a linear or log-linear (i.e., exponential) trend model should be used, an analyst should first plot the data. A linear trend model may be appropriate if the data points appear to be equally distributed above and below the regression line. Inflation rate data can often be modeled with a linear trend model.

If, on the other hand, the data plots with a nonlinear (curved) shape, then the residuals from a linear trend model will be persistently positive or negative for a period of time. In this case, the log-linear model may be more suitable. In other words, when the residuals from a linear trend model are serially correlated (i.e., autocorrelated), a log-linear trend model may be more appropriate. In other words, by taking the log of the  $y$  variable, a regression line can better fit the data. Financial data (e.g., stock indices and stock prices) and company sales data are often best modeled with log-linear models.

Figure 4 shows a time series that is best modeled with a log-linear trend model rather than a linear trend model.

**Figure 4: Linear vs. Log-Linear Trend Models**



The panel on the left is a plot of data that exhibits exponential growth along with a linear trend line. The panel on the right is a plot of the natural logs of the original data and a representative log-linear trend line. The log-linear model fits the transformed data better than the linear trend model and, therefore, yields more accurate forecasts. The bottom line is that when a variable grows at a constant rate, a log-linear model is most appropriate. When the variable increases over time by a constant amount, a linear trend model is most appropriate.

# CFA-FRM培训招生说明

## 一、VIP资料包括以下内容：（所有课程均有配套课件讲义）

**前导课程：**考试介绍、数量基础、金融英语、金融市场与产品的介绍，为学员建立考试框架，打牢基础；

**基础课程：**紧扣考纲要求，对考点全面解析，建立系统知识框架，帮助学员根据课程进度系统性学习；

**强化课程：**突出重点，化繁为简。强化对重要知识点的掌握，让学员完全把握知识点与掌握解题技巧；

**直播串讲：**以习题串讲的形式，针对每门课程进行对应的重点难点讲解，帮助考生更好的掌握对应知识；

**冲刺课程：**分析历年考试的重点和难点，通过2次模考与讲解，强化对重点的知识的把握，提高应试技巧；

**百题预测：**考点精准预测85%以上。以题带点，掌握出题思路与解题方法，最终通过考试取得证书；

**赠送资料：**官方原版教材 高清NOTES 习题 秘籍 考纲解读等电子资料；

## 二、VIP价格说明：（一次收费到当期课程完结，绝无二次收费）

2018年课程价格：488元。购买会赠送17年本级别全套课程

2017年课程价格：288元。

**备注：**此价格为一个级别全套课程价格，无其他任何隐形收费。

**限时优惠：**资料逐步涨价，第一阶段（截止5月1日）488元，第二阶段（截止考试结束）600元。

## 三、付款方式：

联系微信：jinronger01

（只支持微信支付，不支持其他任何支付方式！）

## 四、资料发放形式：

主要通过百度网盘提供下载，课程更新同样在百度网盘里进行。

为确保课程不外传，所有课程全程加密，每位VIP会员给两台设备授权观看。

课程支持windows/mac/ios/安卓等系统，支持手机/平板/电脑等设备。

## 五、2017年部分会员PASS报喜截图：（限于篇幅，随机抽取）

## MODEL SELECTION CRITERIA

---

**LO 24.3: Compare and evaluate model selection criteria, including mean squared error (MSE),  $s^2$ , the Akaike information criterion (AIC), and the Schwarz information criterion (SIC).**

---

### Mean Squared Error

Mean squared error (MSE) is a statistical measure computed as the sum of squared residuals divided by the total number of observations in the sample.

$$MSE = \frac{\sum_{t=1}^T e_t^2}{T}$$

where:

$T$  = total sample size

$e_t = y_t - \hat{y}_t$  (the residual for observation  $t$  or difference between the observed and expected observation)

$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1(t)$  (i.e., a regression model)

The MSE is based on *in-sample* data. The regression model with the smallest MSE is also the model with the smallest sum of squared residuals. The residuals are calculated as the difference between the actual value observed and the predicted value based on the regression model. Scaling the sum of squared residuals by  $1 / T$  does not change the ranking of the models based on squared residuals.

MSE is closely related to the coefficient of determination ( $R^2$ ). Notice in the  $R^2$  equation that the numerator is simply the sum of squared residuals (SSR), which is identical to the MSE numerator.

$$R^2 = 1 - \frac{\sum_{t=1}^T e_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

The denominator in the  $R^2$  calculation is the sum of the difference of observations from the mean. Notice that we subtract the second term from one in the  $R^2$  calculation. Thus, the regression model with the smallest MSE is also the one that has the largest  $R^2$ .

Model selection is one of the most important criteria in forecasting data. Unfortunately, selecting the best model based on the highest  $R^2$  or smallest MSE is not effective in producing good *out-of-sample* forecasting models. A better methodology to select the best forecasting model is to find the model with the smallest out-of-sample, one-step-ahead MSE.

## The $s^2$ Measure

The use of in-sample MSE to estimate out-of-sample MSE is not very effective because in-sample MSE cannot increase when more variables are included in the forecasting model. Thus, MSE will have a downward bias when predicting out-of-sample error variance. Selection criteria differ based on the penalty imposed when the number of parameter estimates is increased in the regression model. One way to reduce the bias associated with MSE is to impose a penalty on the degrees of freedom,  $k$ . The  $s^2$  measure is an unbiased estimate of the MSE because it corrects for degrees of freedom as follows:

$$s^2 = \frac{\sum_{t=1}^T e_t^2}{T - k}$$

As more variables are included in a regression equation, the model is at greater risk of over-fitting the in-sample data. This problem is also often referred to as **data mining**. The problem with data mining is that the regression model does a very good job of explaining the sample data but does a poor job of forecasting out-of-sample data. As more parameters are introduced to a regression model, it will explain the data better, but may be worse at forecasting out-of-sample data.

Therefore, it is important to adjust for the number of variables or parameters used in a regression model because increasing the number of parameters will not necessarily improve the forecasting model. The degrees of freedom penalty rises with more parameters, but the MSE could fall. Thus, the best model is selected based on the smallest unbiased MSE, or  $s^2$ .

The unbiased MSE estimate,  $s^2$ , will rank models in the same way as the adjusted  $R^2$  measure. Adjusted  $R^2$  using the  $s^2$  estimate can be computed as follows:

$$\bar{R}^2 = 1 - \frac{s^2}{\sum_{t=1}^T \frac{(y_t - \bar{y})^2}{T-1}}$$

Notice that the denominator in this equation is based only on the data used in the regression. Therefore, it will be a constant number and the model with the highest adjusted  $R^2$  will also have the smallest  $s^2$ . Thus, the  $s^2$  and adjusted  $R^2$  criteria will always rank forecasting models equivalently.

## Akaike and Schwarz Criterion

As mentioned, selection criteria are often compared based on a penalty factor. The unbiased MSE estimate,  $s^2$ , defined earlier, can be re-written (by multiplying  $T$  to the numerator and

denominator) to highlight the penalty for degrees of freedom. In the following equation, the first term ( $T / T - k$ ) can be thought of as the **penalty factor**.

$$s^2 = \left( \frac{T}{T - k} \right) \frac{\sum_{t=1}^T e_t^2}{T}$$

This notation is useful when comparing different selection criteria because it takes the form of a *penalty factor times the MSE*. The **Akaike information criterion** (AIC) and the **Schwarz information criterion** (SIC) use different penalty factors as follows:

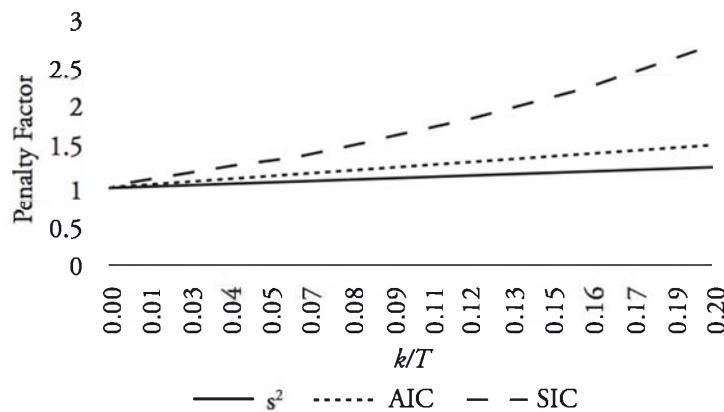
$$AIC = e^{\left( \frac{2k}{T} \right) \frac{\sum_{t=1}^T e_t^2}{T}}$$

$$SIC = T^{\left( \frac{k}{T} \right) \frac{\sum_{t=1}^T e_t^2}{T}}$$

Note that the penalty factors for  $s^2$ , AIC, and SIC are  $(T / T - k)$ ,  $e^{(2k / T)}$ , and  $T^{(k / T)}$ , respectively.

Suppose an analyst runs a forecasting model with a total sample size of 150. Figure 5 illustrates the change in penalty factors for the  $s^2$ , AIC, and SIC as the degrees of freedom to total sample size ( $k / T$ ) changes from 0 to 0.20. The  $s^2$  penalty factor is the flattest line with a slow increase in penalty as  $k / T$  increases. The AIC penalty factor increases at a slightly higher rate than the  $s^2$  penalty factor, and the SIC penalty factor increases exponentially at an increasing rate and, therefore, has the highest penalty factor.

**Figure 5: Penalty Factor for  $s^2$ , AIC, and SIC**



## EVALUATING CONSISTENCY

---

### LO 24.4: Explain the necessary conditions for a model selection criterion to demonstrate consistency.

---

**Consistency** is a key property that is used to compare different selection criteria. Two conditions are required for a model selection criteria to be considered consistent based on whether the *true* model is included among the regression models being considered.

- When the *true* model or *data generating process* (DGP) is one of the defined regression models, then the probability of selecting the *true* model approaches one as the sample size increases.
- When the *true* model is *not* one of the defined regression models being considered, then the probability of selecting the *best approximation* model approaches one as the sample size increases.

Because we live in a very complex world, almost all economic and financial models have assumptions that simplify this complex environment. Thus, the reality is that the second condition of consistency is more relevant. All of our models are most likely false so, therefore, we are seeking the best approximation.

So how do our selection criteria fair based on consistency? MSE does not penalize for degrees of freedom and therefore is not consistent. The unbiased MSE,  $s^2$ , adjusts MSE for degrees of freedom, but the adjustment is too small for consistency. Figure 5 illustrated that AIC has a larger penalty factor than  $s^2$ . However, with large sample sizes the AIC tends to select models that have too many variables or parameters. This suggests that the penalty factor for degrees of freedom is still not large enough. The most consistent selection criteria with the greatest penalty factor for degrees of freedom is the SIC.

While the SIC is considered the most consistent criteria, the AIC is still a useful measure. If we consider the fact that the true model may be much more complicated than the models under consideration, then the AIC measure should be examined. *Asymptotic efficiency* is the property that chooses a regression model with one-step-ahead forecast error variances closest to the variance of the true model. Interestingly, the AIC is asymptotically efficient and the SIC is not asymptotically efficient.

In conclusion, choosing the best forecasting model is an important task and we have discussed four key selection criteria. Adjusting for the degrees of freedom is extremely important and the SIC is the best selection criteria because it is consistent and also has the highest penalty factor. The AIC is also an important measure that is often considered in addition to SIC.

## KEY CONCEPTS

### LO 24.1

A linear trend is a time series pattern that can be graphed with a straight line:

$$y_t = \beta_0 + \beta_1(t)$$

A nonlinear trend is a time series pattern that can be graphed with a curve. Nonlinear trends can be modeled using either quadratic or exponential (i.e., log-linear) functions:

$$y_t = \beta_0 + \beta_1(t) + \beta_2(t)^2$$

$$y_t = \beta_0 e^{\beta_1(t)} \text{ or } \ln(y_t) = \ln(\beta_0) + \beta_1(t)$$

### LO 24.2

Most statistical software packages can apply ordinary least squares (OLS) regression to estimate the coefficients in a trend line. The regression output can then be used to forecast in-sample and out-of-sample data.

### LO 24.3

Mean squared error (MSE) is a statistical measure computed as the sum of squared residuals (SSR) divided by the number of observations in a regression model:

$$\text{MSE} = \frac{\sum_{t=1}^T e_t^2}{T}$$

The unbiased MSE,  $s^2$ , adjusts for the degrees of freedom,  $k$ , in the denominator as follows:

$$s^2 = \frac{\sum_{t=1}^T e_t^2}{T - k}$$

The penalty factors for  $s^2$ , Akaike information criterion (AIC), and Schwarz information criterion (SIC) are  $(T / T - k)$ ,  $e^{(2k / T)}$ , and  $T^{(k / T)}$ , respectively. SIC has the largest penalty factor.

#### **LO 24.4**

A selection criteria is considered to be consistent if the following two conditions are met:

- When the true model or data generating process (DGP) is one of the defined regression models under consideration, then the probability of selecting the true model approaches one as the sample size increases.
- When the true model is not one of the defined regression models being considered, then the probability of selecting the best approximation model approaches one as the sample size increases.

The SIC is the most consistent selection criteria.

## CONCEPT CHECKERS

1. An analyst has determined that monthly sport utility vehicle (SUV) sales in the United States have been increasing over the last 10 years, but the growth rate over that period has been relatively constant. Which model is most appropriate to predict future SUV sales?
  - A.  $SUVsales_t = \beta_0 + \beta_1(t)$ .
  - B.  $\ln SUVsales_t = \ln(\beta_0) + \beta_1(t)$ .
  - C.  $\ln SUVsales_t = \beta_0 + \beta_1(SUVsales_{t-1})$ .
  - D.  $SUVsales_t = \beta_0 + \beta_1(t) + \beta_2(t)^2$ .
2. Richard Frank, FRM, is running a regression model to forecast in-sample data. He is concerned about data mining and over-fitting the data. Which of the following criteria provides the highest penalty factor based on degrees of freedom?
  - A. Mean squared error (MSE).
  - B. Unbiased mean squared error ( $s^2$ ).
  - C. Akaike information criterion (AIC).
  - D. Schwarz information criterion (SIC).
3. Which of the following statements does not accurately describe the mean squared error (MSE) statistical measure?
  - A. The regression model with the smallest MSE is also the model with the smallest sum of squared residuals.
  - B. Scaling the sum of squared residuals by  $1 / T$  changes the ranking of the models based on squared residuals.
  - C. The residuals in the numerator of the MSE calculation are defined as the difference between the actual value observed and the predicted value based on the regression model.
  - D. The best regression model based on minimizing the MSE will also be the one that maximizes  $R^2$ .
4. Sally Morgan, a junior analyst, is identifying a forecasting model based on a number of industry factors, company factors, and leading market indicators. She decides to choose the model with the highest  $R^2$  measure because she knows this is a goodness-of-fit measure for selecting regression models. Morgan chooses a model with a very large number of parameters. How will Morgan's supervisor, Jessica Bolt, most likely respond to Morgan's choice of models? Bolt will:
  - A. agree with Morgan as  $R^2$  is the best goodness-of-fit measure available.
  - B. agree with Morgan as  $R^2$  is a common acceptable statistical measure and maximizing  $R^2$  is the same as minimizing MSE.
  - C. disagree with Morgan because MSE is a better measure than  $R^2$  for selecting forecasting models.
  - D. disagree with Morgan because  $R^2$  is a biased measure.

5. When selecting the best forecasting model among possible regression models, the property of consistency is desired. Which of the following statements most accurately describes a required condition for a model to be considered consistent?
- A. When the true model is one of the defined regression models under consideration, then the probability of selecting the best approximation model approaches one with a very large sample size.
  - B. When the true model is one of the defined regression models under consideration, then the probability of selecting the true model approaches one with a very small sample size.
  - C. When the true model is not one of the defined regression models being considered, then the probability of selecting the best approximation model approaches one as the sample size increases.
  - D. When the true model is not one of the defined regression models being considered, the choice of the model selected is irrelevant and cannot be determined.

## CONCEPT CHECKER ANSWERS

1. B A log-linear model is most appropriate for a time series that grows at a relatively constant growth rate.
2. D The Schwarz information criterion (SIC) has the highest penalty factor. The mean squared error (MSE) does not penalize the regression model based on the increased number of parameters,  $k$ . The penalty factors for  $s^2$ , AIC, and SIC are  $(T / T - k)$ ,  $e^{(2k / T)}$ , and  $T^{(k / T)}$ , respectively. Thus, SIC has the greatest penalty factor.
3. B Scaling the sum of squared residuals by  $1 / T$  in the MSE statistic does *not* change the ranking of the models based on squared residuals. The rankings will be the same.
4. D The model selected by Morgan is at greater risk of over-fitting the in-sample data. It is important to adjust for the number of variables or parameters used in a regression model. The best model should be selected based on the smallest unbiased MSE, or  $s^2$ .
5. C A selection criteria is considered to be consistent if the following two conditions are met:  
(1) when the true model is not one of the defined regression models being considered, then the probability of selecting the best approximation model approaches one as the sample size increases and (2) when the true model or data generating process (DGP) is one of the defined regression models under consideration, then the probability of selecting the true model approaches one as the sample size increases.

---

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

# MODELING AND FORECASTING SEASONALITY

---

Topic 25

## EXAM FOCUS

This topic expands on the concept of trend models by accounting for seasonality effects. Seasonality refers to the predictable changes that occur in a time series year to year. For the exam, be able to describe the sources of seasonal effects and the approaches for analyzing a time series that is affected by seasonality. Also, be able to explain how seasonal dummy variables can be used to model seasonality with regression analysis techniques. Finally, be able to describe how to incorporate various calendar effects to more accurately forecast a seasonal series.

---

## SOURCES OF SEASONALITY

---

### LO 25.1: Describe the sources of seasonality and how to deal with it in time series analysis.

---

Seasonality in a time series is a pattern that tends to repeat from year to year. One example is monthly sales data for a retailer. Because sales data normally varies according to the calendar, we might expect this month's sales ( $x_t$ ) to be related to sales for the same month last year ( $x_{t-12}$ ).

Specific examples of seasonality relate to increases that occur at only certain times of the year. For example, purchases of retail goods typically increase dramatically every year in the weeks leading up to Christmas. Similarly, sales of gasoline generally increase during the summer months when people take more vacations. Weather is another common example of a seasonal factor as production of agricultural commodities is heavily influenced by changing seasons and temperatures. For many industrialized countries, seasonality effects are significant: economic activity expands substantially in the fourth quarter and contracts in the first quarter.

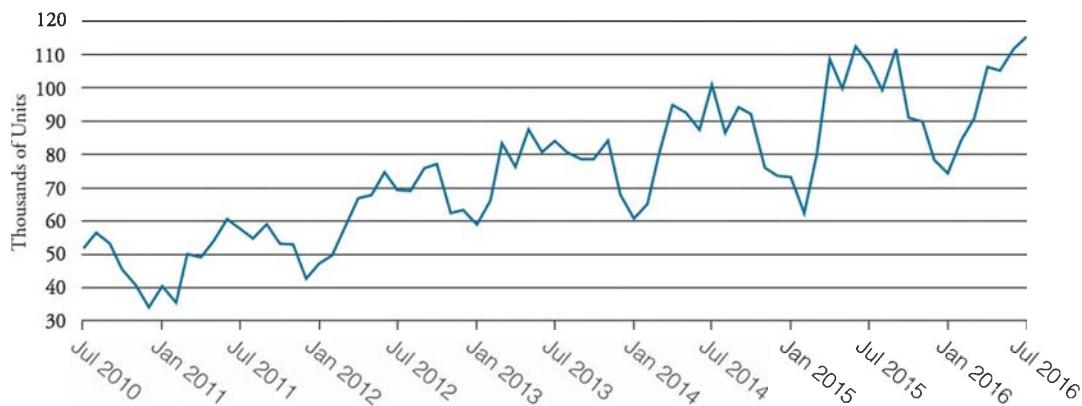
Annual changes can be approximate, as in the case of **stochastic seasonality**, or exact, as in the case of **deterministic seasonality**. Similar to the previous topic, where we focused on deterministic trends, the focus here will be exclusively on deterministic seasonality.

There are two approaches for modeling and forecasting a time series impacted by seasonality: (1) using a seasonally adjusted time series and (2) regression analysis with seasonal dummy variables.

A seasonally adjusted time series is created by removing the seasonal variation from the data. This type of adjustment is commonly made in macroeconomic forecasting where the goal is to only measure the *nonseasonal fluctuations* of a variable. However, the use of seasonal adjustments in business forecasting is usually inappropriate because seasonality often accounts for large variations in a time series. Financial forecasters should be interested in capturing *all* variation in a time series, not just the nonseasonal portions.

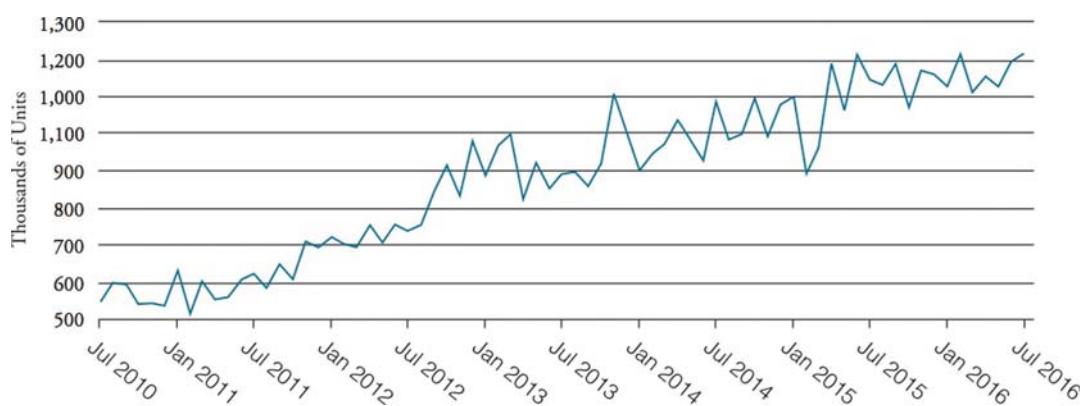
Figures 1 and 2 illustrate the difference between data that is not seasonally adjusted and data that is seasonally adjusted, using data for housing starts of privately owned homes between July 2010 and July 2016. As you can see from Figure 1, data that is not seasonally adjusted fluctuates greatly throughout the year. Housing starts typically rise in the spring, peak in the summer, and slump through the winter. Conversely, a seasonally adjusted time series, such as the one seen in Figure 2, eliminates variations due to seasonality. This adjustment makes it easier for an analyst to make month-to-month comparisons.

**Figure 1: Housing Starts—Not Seasonally Adjusted\***



\* Source: U.S. Bureau of the Census

**Figure 2: Housing Starts—Seasonally Adjusted Annual Rate\***



\* Source: U.S. Bureau of the Census

## MODELING SEASONALITY WITH REGRESSION ANALYSIS

---

### **LO 25.2: Explain how to use regression analysis to model seasonality.**

---

A regression that incorporates seasonal dummies can be an effective technique for modeling seasonality. In this process, **seasonal dummy variables** can take a value of either “1” or “0,” to represent the independent variable being “on” or “off.” For example, in a time series regression of monthly stock returns, we might incorporate a “January” dummy variable that would take on the value of “1” if a stock return occurred in January and “0” if it occurred in any other month. The January dummy variable helps us to see if stock returns in January were significantly different than stock returns in all other months of the year. Many “January effect” anomaly studies use this type of regression methodology.

A “pure” seasonal dummy model takes the following form:

$$y_t = \sum_{i=1}^s \gamma_i D_{i,t} + \varepsilon_t$$

In this model, the  $\gamma$  represent seasonal factors and the  $D$  represent the dummy variables. (If all of the  $\gamma_i$  turn out to be equal, the time series does not display seasonality and the seasonal dummy variables can be dropped.)

The estimated regression coefficient for dummy variables indicates the difference in the dependent variable for the category represented by the dummy variable versus the average value of the dependent variable for all classes other than the dummy variable class. For example, the slope coefficient for the January dummy variable would indicate whether, and by how much, security returns are different in January compared to other months.

An alternative to including a dummy variable in our model for each season is to include an intercept in the model and then  $s - 1$  dummy variables. The model then takes the following form:

$$y_t = \beta_0 + \sum_{i=1}^{s-1} \beta_i D_{i,t} + \varepsilon_t$$

An important consideration when performing multiple regression and modeling seasonality with dummy variables is the number of dummy variables to include in the model. As mentioned, if we include an intercept in our model and there are  $s$  seasons, we use  $s - 1$  dummy variables to avoid the problem of (perfect) multicollinearity. For example, to account for seasonality in monthly ( $s = 12$ ) data, we are likely to use not 12, but rather  $s - 1 = 11$  dummy variables in a model that incorporates an intercept.

## Interpreting a Dummy Variable Regression

Consider the following regression equation for explaining quarterly earnings per share (EPS) in terms of the quarter of their occurrence:

$$\text{EPS}_t = \beta_0 + \beta_1 D_{1,t} + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \varepsilon_t$$

where

$\text{EPS}_t$  = a quarterly observation of earnings per share

$D_{1,t}$  = 1 if period  $t$  is the first quarter of a year,  $D_{1,t} = 0$  otherwise

$D_{2,t}$  = 1 if period  $t$  is the second quarter of a year,  $D_{2,t} = 0$  otherwise

$D_{3,t}$  = 1 if period  $t$  is the third quarter of a year,  $D_{3,t} = 0$  otherwise

The intercept term,  $\beta_0$ , represents the average value of EPS for the fourth quarter. The slope coefficient on each dummy variable estimates the *difference* in EPS (on average) between the respective quarter (i.e., quarter 1, 2, or 3) and the omitted quarter (the fourth quarter, in this case). Think of the omitted class as the reference point.

For example, suppose we estimate the quarterly EPS regression model with 10 years of data (40 quarterly observations) and find that  $\beta_0 = 1.25$ ,  $\beta_1 = 0.75$ ,  $\beta_2 = -0.20$ , and  $\beta_3 = 0.10$ :

$$\widehat{\text{EPS}}_t = 1.25 + 0.75D_{1,t} - 0.20D_{2,t} + 0.10D_{3,t}$$

We can use this equation to determine the average EPS in each quarter over the past 10 years:

- average fourth-quarter EPS = 1.25
- average first-quarter EPS =  $1.25 + 0.75 = 2.00$
- average second-quarter EPS =  $1.25 - 0.20 = 1.05$
- average third-quarter EPS =  $1.25 + 0.10 = 1.35$

These are also the model's predictions of future EPS in each quarter of the following year.

For example, to use the model to predict EPS in the first quarter of the next year, set

$\hat{D}_{1,t} = 1$ ,  $\hat{D}_{2,t} = 0$ , and  $\hat{D}_{3,t} = 0$ . Then  $\widehat{\text{EPS}}_t = 1.25 + 0.75(1) - 0.20(0) + 0.10(0) = 2.00$ .

This simple model uses average EPS for a specific quarter over the past 10 years as the forecast of EPS in its respective quarter of the following year.

The concept of seasonal variation can also be extended to account for other types of calendar effects, such as **holiday variations** (HDV) and **trading-day variations** (TDV). For example, Easter is a holiday that is often modeled with a dummy variable as it affects many time series, such as sales, inventories, and hours worked. However, Easter can occur in either March or April, depending on the year, so a monthly model incorporating an Easter dummy variable would specify a 0 if the month did not contain Easter and a 1 if the month contains Easter in the given year. Regarding trading-day variation, regression models can be constructed to account for different numbers of trading days (or business days) each month. In this case, the value of the trading-day variable each month could be the exact number of trading days (generally between 19 and 23) for a given month.

## SEASONAL SERIES FORECASTING

---

### LO 25.3: Explain how to construct an h-step-ahead point forecast.

---

Forecasting a seasonal series is fairly straightforward. A pure seasonal dummy model can be constructed as follows:

$$y_t = \sum_{i=1}^s \gamma_i(D_{i,t}) + \epsilon_t$$

After adding a trend, the model can then take the following form:

$$y_t = \beta_1(t) + \sum_{i=1}^s \gamma_i(D_{i,t}) + \epsilon_t$$

Allowing for holiday variations (HDV) and trading-day variations (TDV) expands the forecasting model even further:

$$y_t = \beta_1(t) + \sum_{i=1}^s \gamma_i(D_{i,t}) + \sum_{i=1}^{v_1} \delta_i^{HDV}(HDV_{i,t}) + \sum_{i=1}^{v_2} \delta_i^{TDV}(TDV_{i,t}) + \epsilon_t$$

This complete model can now be used for *out-of-sample* forecasts at time  $T + h$  by constructing an h-step-ahead point forecast as follows:

$$y_{T+h} = \beta_1(T+h) + \sum_{i=1}^s \gamma_i(D_{i,T+h}) + \sum_{i=1}^{v_1} \delta_i^{HDV}(HDV_{i,T+h}) + \sum_{i=1}^{v_2} \delta_i^{TDV}(TDV_{i,T+h}) + \epsilon_{T+h}$$

## KEY CONCEPTS

---

### LO 25.1

Seasonality refers to the predictable changes that occur in a time series year to year. For example, the production of agricultural commodities is highly seasonal.

There are two approaches for modeling and forecasting a time series that is affected by seasonality: (1) using a seasonally adjusted time series and (2) regression analysis with seasonal dummy variables.

---

### LO 25.2

Modeling seasonality assigns seasonal dummy variables a value of either “0” or “1.” One consideration when modeling seasonality with dummy variables is the choice of the number of dummy variables to include in the model. To distinguish between  $s$  classes when we include an intercept term in the model, we use  $s - 1$  dummy variables. The intercept in the regression model accounts for the omitted season.

Seasonality can be extended to account for other types of calendar effects, such as holiday variations (which adjust for holidays like Easter that may occur in different months each year) and trading-day variations (which reflect the varying number of days each month).

---

### LO 25.3

An  $h$ -step-ahead point forecast that accounts for trend, seasonality, HDV, and TDV could be constructed as follows:

$$y_{T+h} = \beta_1(T + h) + \sum_{i=1}^s \gamma_i(D_{i,T+h}) + \sum_{i=1}^{v_1} \delta_i^{HDV}(HDV_{i,T+h}) + \sum_{i=1}^{v_2} \delta_i^{TDV}(TDV_{i,T+h}) + \epsilon_{T+h}$$

## CONCEPT CHECKERS

1. A forecaster is least likely to remove seasonality (and focus on forecasting nonseasonal fluctuations) in the case of a time series related to:
  - A. corporate earnings.
  - B. unemployment rates.
  - C. the consumer price index (CPI).
  - D. gross domestic product (GDP).
  
2. Jill Williams is an analyst in the retail industry. She is modeling a company's sales over time and has noticed a quarterly seasonal pattern. If Williams includes an intercept term in her model, how many dummy variables should she use to model the seasonality component?
  - A. 1.
  - B. 2.
  - C. 3.
  - D. 4.
  
3. Consider the following regression equation utilizing dummy variables for explaining quarterly SALES in terms of the quarter of their occurrence:

$$\text{SALES}_t = \beta_0 + \beta_1 D_{1,t} + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \varepsilon_t$$

where:

$\text{SALES}_t$  = a quarterly observation of EPS

$D_{1,t} = 1$  if period  $t$  is the first quarter,  $D_{1,t} = 0$  otherwise

$D_{2,t} = 1$  if period  $t$  is the second quarter,  $D_{2,t} = 0$  otherwise

$D_{3,t} = 1$  if period  $t$  is the third quarter,  $D_{3,t} = 0$  otherwise

- The intercept term  $\beta_0$  represents the average value of sales for the:
- A. first quarter.
  - B. second quarter.
  - C. third quarter.
  - D. fourth quarter.
- 
4. In a pure seasonal dummy model, if all seasonal factors (i.e., the  $\gamma$ ) in the model are the same, the conclusion is:
    - A. an absence of seasonality.
    - B. the need for seasonally adjusted data.
    - C. the need for additional dummy variables.
    - D. to retain all current seasonal dummy variables in the model.
  
  5. Which of the following scenarios would produce a forecasting model that exhibits perfect multicollinearity? A model that includes:
    - A. only one seasonal dummy that is equal to 1.
    - B. a dummy variable for each season, plus an intercept.
    - C. a holiday variation variable that accounts for an "Easter dummy variable."
    - D. a trading-day variation variable for modeling trading volume throughout the year.

## CONCEPT CHECKER ANSWERS

1. A It would be inappropriate to forecast a *seasonally adjusted* time series for corporate earnings: in this kind of business situation we want to forecast *all* variation in the time series, and not just the nonseasonal portion. A seasonal adjustment is accomplished by removing the seasonal variation and then modeling and forecasting a seasonally adjusted time series. This type of adjustment is commonly made in *macroeconomic* forecasting where the goal is to measure only the *nonseasonal* fluctuations of a variable.
2. C Whenever we want to distinguish between  $s$  seasons in a model that incorporates an intercept, we must use  $s - 1$  dummy variables. For example, if we have quarterly data,  $s = 4$ , and thus we would include  $s - 1 = 3$  seasonal dummy variables.
3. D The intercept term represents the average value of EPS for the fourth quarter. The slope coefficient on each dummy variable estimates the difference in EPS (on average) between the respective quarter (i.e., quarter 1, 2, or 3) and the omitted quarter (the fourth quarter, in this case).
4. A In a pure seasonal dummy model, the  $\gamma$  represent seasonal factors (i.e., the intercepts). If a time series does not exhibit seasonality, all  $\gamma_i$  would all be equal and the seasonal dummy variables can be dropped.
5. B Including the full set of dummy variables and an intercept term would produce a forecasting model that exhibits perfect multicollinearity.

# CHARACTERIZING CYCLES

---

Topic 26

## EXAM FOCUS

While earlier topics addressed trend and seasonality in a time series, here we begin to examine cycles that are not related to trend or seasonality. Cycles include any persistent relationship among past, present, and future data that can be found in a time series. This topic also provides the theoretical base for the moving average (MA), autoregressive (AR), and autoregressive moving average (ARMA) models we will introduce in the topic that follows. The main idea to grasp here is how and when past data is useful for forecasting, in particular what makes a time series covariance stationary and the properties of a white noise time series. Note that most of the LOs here require candidates to define, describe, and explain, not to recite and apply complicated formulas.

---

## COVARIANCE STATIONARY TIME SERIES

---

**LO 26.1: Define covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function, and autoregression.**

**LO 26.2: Describe the requirements for a series to be covariance stationary.**

---

A process such as a time series must have certain properties if we want to forecast its future values based on its past values. In particular, it needs the relationships among its present and past values to remain stable over time. We refer to such a time series as being covariance stationary.

To be covariance stationary, a time series must exhibit the following three properties:

1. Its mean must be stable over time.
2. Its variance must be finite and stable over time.
3. Its covariance structure must be stable over time.

*Covariance structure* refers to the covariances among the values of a time series at its various lags or displacements, which are a given number of periods apart at which we can observe its values. We use the lowercase Greek letter tau,  $\tau$ , to represent a displacement. For example,  $\tau = 1$  refers to a one-period displacement, comparing each value of a time series to its preceding value, and if  $\tau = 4$  we are comparing values four periods apart along the time series.

 *Professor's Note: We tend to use the words lags and displacements interchangeably because we are mostly concerned with time series. The forecasting techniques we will discuss can actually be applied to processes other than time series, with displacements measured in units other than periods of time. Later in this topic we will introduce a notation for lags that is specific to time series.*

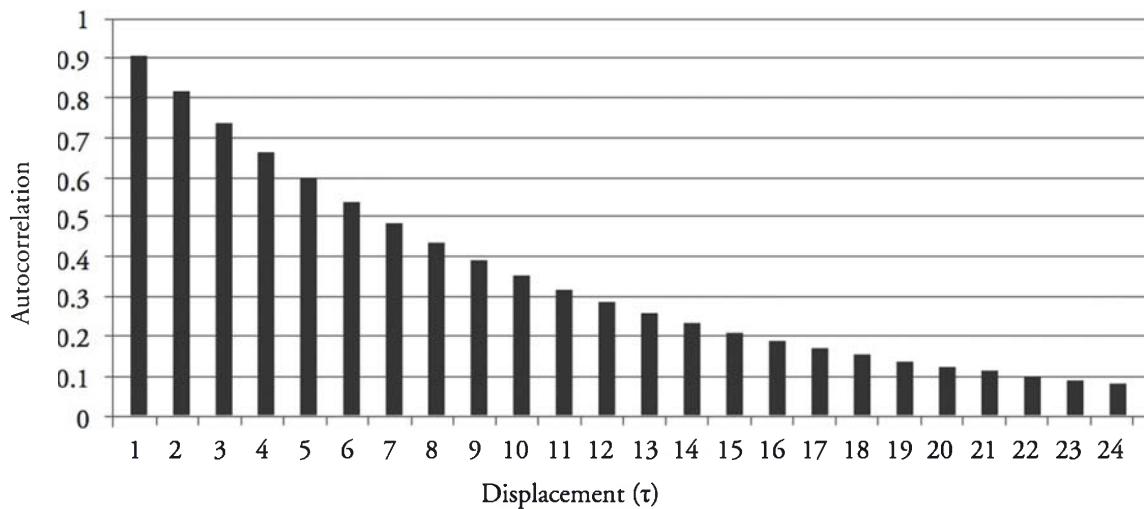
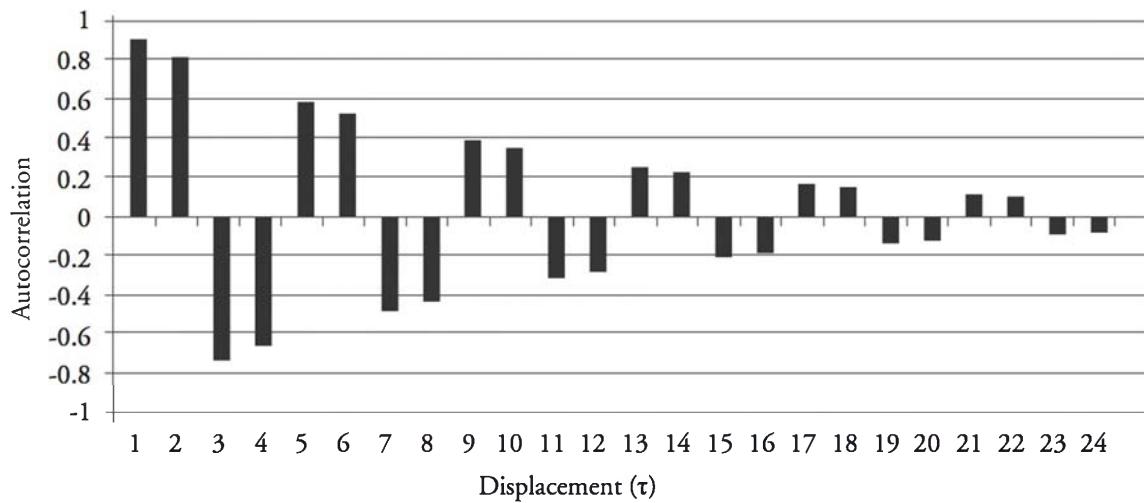
The covariance between the current value of a time series and its value  $\tau$  periods in the past is referred to as its **autocovariance** at displacement  $\tau$ . Its autocovariances for all  $\tau$  make up its **autocovariance function**. If a time series is covariance stationary, its autocovariance function is stable over time. That is, its autocovariance depends on the  $\tau$  we choose, but does not depend on the time over which we observe the series.

As we often do when working with covariances, we can convert them to correlations to better interpret the strength of the relationships. To convert an autocovariance function to an **autocorrelation function**, we divide the autocovariance at each  $\tau$  by the variance of the time series. This gives us an autocorrelation for each  $\tau$  that will be scaled between +1 and -1.

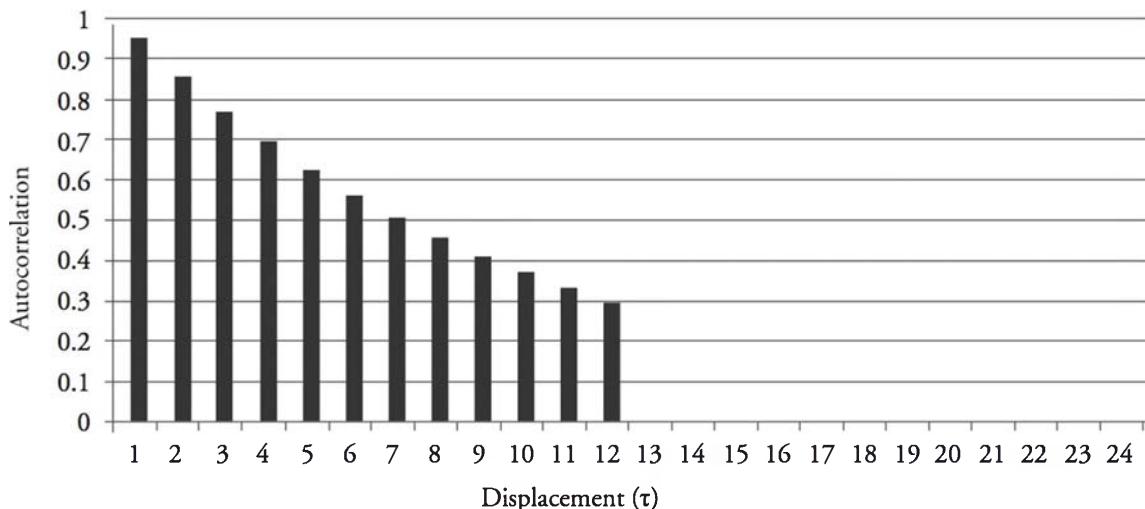
**Autoregression** is a linear regression of a time series against its own past values. The regression coefficient that results is referred to as the partial autocorrelation for that lag. Partial autocorrelations for all lags make up the **partial autocorrelation function** of the time series.

 *Professor's Note: These are "partial" in the sense that they are regressed one lag at a time. For example, if we regress a monthly time series against its year-ago values, we get a partial autocorrelation for  $\tau = 12$  that does not account for any effects from other lags. We would be unlikely to get the same result for  $\tau = 12$  if we ran a multiple regression that also included  $\tau = 1$ ,  $\tau = 2$ , and so forth.*

A useful way to analyze an autocorrelation function (or a partial autocorrelation function) is to display it on a graph. Figure 1 illustrates some of the typical patterns autocorrelation functions can have. One feature they all have in common is that autocorrelations approach zero as  $\tau$  gets large. This is always the case for a covariance stationary series.

**Figure 1: Autocorrelation Functions****a. Decreasing positive autocorrelations****b. Decreasing and oscillating autocorrelations**

c. Autocorrelations that drop off after some value of  $\tau$




---

**LO 26.3: Explain the implications of working with models that are not covariance stationary.**

---

If a time series is not covariance stationary, we cannot model it directly from its past values. However, a common modeling approach is to identify and isolate an underlying, covariance stationary aspect of a time series that can be modeled.

One way to do this is by dealing separately with properties like trend and seasonality, as we described in preceding topics, and modeling the time series after filtering out those properties. Another way is to apply a transformation to the series. To take the simplest example, even if a time series is not covariance stationary after removing its trend and seasonal components, its rate of change might be covariance stationary.



*Professor's Note: A time series can be transformed in a number of ways (first differences, logarithmic scaling, etc.). The idea is to find some covariance stationary property we can model, forecast that, and then transform our forecasts back into the units the time series is in.*

## WHITE NOISE

**LO 26.4: Define white noise, and describe independent white noise and normal (Gaussian) white noise.**

**LO 26.5: Explain the characteristics of the dynamic structure of white noise.**

A time series might exhibit no correlation among any of its lagged values. Such a time series is said to be **serially uncorrelated**.

A special type of serially uncorrelated series is one that has a mean of zero and a constant variance. This condition is referred to as **white noise**, or zero-mean white noise, and the time series is said to follow a white noise process.

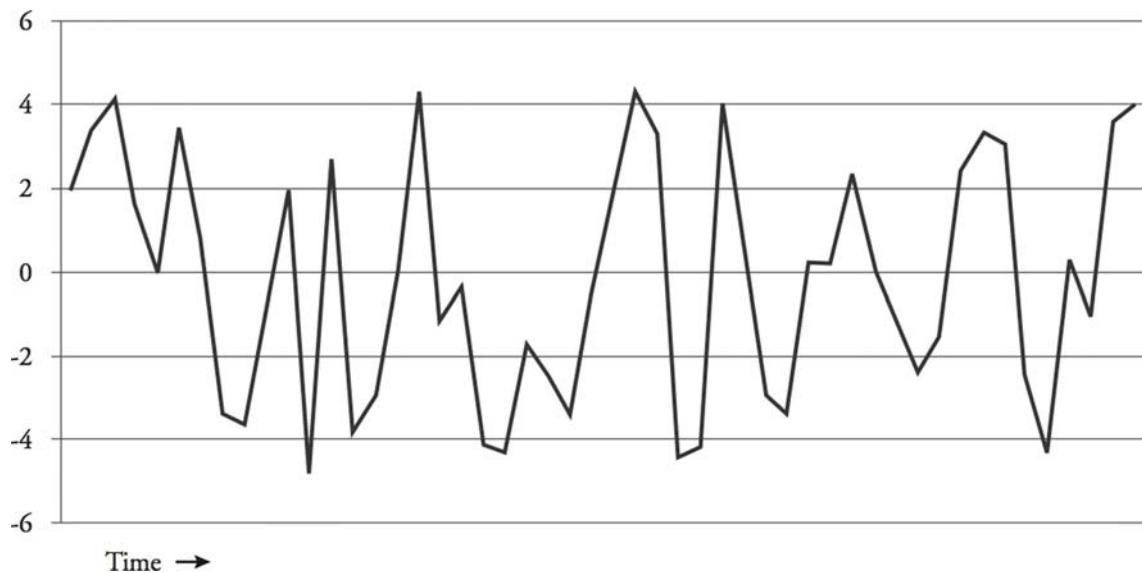
If the observations in a white noise process are independent, as well as uncorrelated, the process is referred to as **independent white noise**. If the process also follows a normal distribution, it is known as **normal white noise** or **Gaussian white noise**. Not all independent white noise processes are normally distributed, but all normal white noise processes are also independent white noise.



*Professor's Note: Anywhere you see "Gaussian," think "normally distributed."*

Graphically, a white noise process resembles Figure 2, with no identifiable patterns among the time periods.

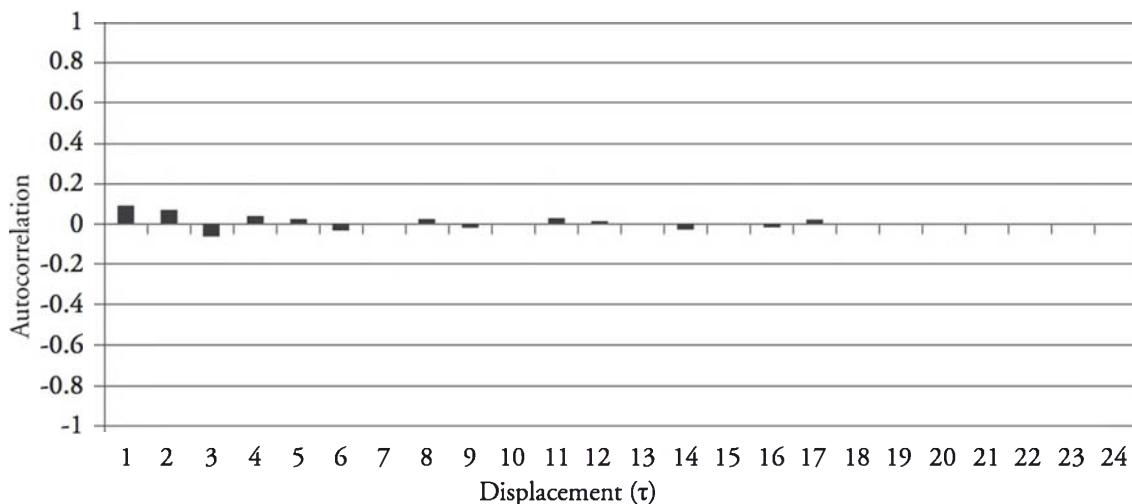
**Figure 2: A White Noise Process**



One important use of the white noise concept is analyzing a forecasting model. A model's forecast errors should follow a white noise process. If they do not, the errors themselves can be forecast based on their past values. This implies that the model is inaccurate in a predictable way and is therefore inadequate.

The autocorrelation or partial autocorrelation functions for a perfectly serially uncorrelated process would show nothing but zeroes for all its displacements. Figure 3 illustrates an autocorrelation function for a process that might be considered serially uncorrelated for practical purposes.

**Figure 3: Autocorrelation Function for an Approximately Serially Uncorrelated Process**



 *Professor's Note: The autocorrelation at  $\tau = 0$  must equal one, because this is the correlation of the series with itself. Some graphical displays of autocorrelation functions include this value, while others leave it out as we did in Figures 1 and 3.*

Earlier we stated that a white noise process has a mean of zero and a constant variance. More properly this refers to its *unconditional mean and variance*. A process may have a *conditional mean and variance* that are not necessarily constant. That is, the expected value of the next observation in the series might not be the mean of the time series, if the next observation is conditional on one or more of its earlier values. If such a relationship exists, we can use it for forecasting the time series.

For an *independent* white noise process, we can say the next value in the series has no conditional relationship to any of its past values. Therefore, its conditional mean is the same as its unconditional mean. In this case, we cannot forecast based on past values.

## LAG OPERATORS

---

### LO 26.6: Explain how a lag operator works.

---

A commonly used notation for time series modeling is the **lag operator**,  $L$ . If  $y_t$  is the value of a time series at time  $t$ , and  $y_{t-1}$  is its value one period earlier, we can express a lag operator as:

$$y_{t-1} = Ly_t$$

In the same way,  $y_{t-2}$  is the value of the time series two periods before  $y_t$  and one period before  $y_{t-1}$ . Stated using lag operators, we can say:

$$y_{t-2} = Ly_{t-1}$$

And because  $y_{t-1}$  is stated as  $Ly_t$ :

$$y_{t-2} = L(Ly_t) = L^2y_t$$

We can use this same notation for any degree of lag:

$$L^m y_t = y_{t-m}$$

Forecasting models often take the form of a **distributed lag** that assigns weights to the past values of a time series. For example, suppose we have the following model:

$$y_t + 0.7y_{t-1} + 0.4y_{t-2} + 0.2y_{t-3}$$

Using lag operators in this model, it would be expressed as:

$$(1 + 0.7L + 0.4L^2 + 0.2L^3)y_t$$

Note that despite the use of exponents in the lag operators, the model is still a linear process. Because it includes all the lags up to 3, this example may be referred to as a **lag operator polynomial** of degree 3. A lag operator polynomial can be constructed to any degree. We can even conceive of an **infinite distributed lag**, which is a lag operator polynomial with an infinite number of terms.

## WOLD'S THEOREM

---

**LO 26.7:** Describe Wold's theorem.

**LO 26.8:** Define a general linear process.

**LO 26.9:** Relate rational distributed lags to Wold's theorem.

---

Earlier we described the forecasting technique of removing the trend and seasonal components from a time series and isolating an underlying, covariance stationary process. **Wold's theorem**, or **Wold's representation**, proposes a way to model that underlying process. It holds that a covariance stationary process can be modeled as an infinite distributed lag of a white noise process. Such a model would take the following form:

$$\varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots = \sum_{i=0}^{\infty} b_i\varepsilon_{t-i}$$

Because this expression can be applied to any covariance stationary series, it is known as a **general linear process**.

In Wold's representation the  $\varepsilon$  terms are referred to as **innovations**. Innovations can be thought of as the errors that would result from a good forecast of the covariance stationary process. That is, they are a white noise process with an unconditional mean of zero.

Innovations are not necessarily independent. If innovations have a conditional relationship with past innovations, they can have a conditional mean that changes over time. This implies that past innovations may provide information that can be used for forecasting.

Wold's representation would not do us much good if it really required an infinite series of past innovations. However, it can be approximated with a ratio of **rational distributed lags**, which are distributed lags with a finite number of terms. Using this approach, an approximation would only have as many terms as the two rational distributed lags in the ratio. Some of the forecasting models we will introduce in the next topic are based on such rational approximations, including the ARMA model.

## ESTIMATING AUTOCORRELATIONS

---

**LO 26.10:** Calculate the sample mean and sample autocorrelation, and describe the Box-Pierce Q-statistic and the Ljung-Box Q-statistic.

**LO 26.11:** Describe sample partial autocorrelation.

---

When working with time series in practice, we do not know their population means or their true autocorrelation or partial autocorrelation functions. The best we can do is estimate them from the sample of data we have.

The sample mean,  $\bar{y}$ , is of course the arithmetic average of the observations. The **sample autocorrelation** for displacement  $\tau$  is estimated by the following formula:

$$\hat{\rho}_\tau = \frac{\sum_{t=\tau+1}^T [(y_t - \bar{y})(y_{t-\tau} - \bar{y})]}{\sum_{t=1}^T (y_t - \bar{y})^2}$$



*Professor's Note: The LO says "calculate the sample autocorrelation," but it's very unlikely that GARP will ask candidates to do this on a multiple-choice exam with nothing but a financial calculator.*

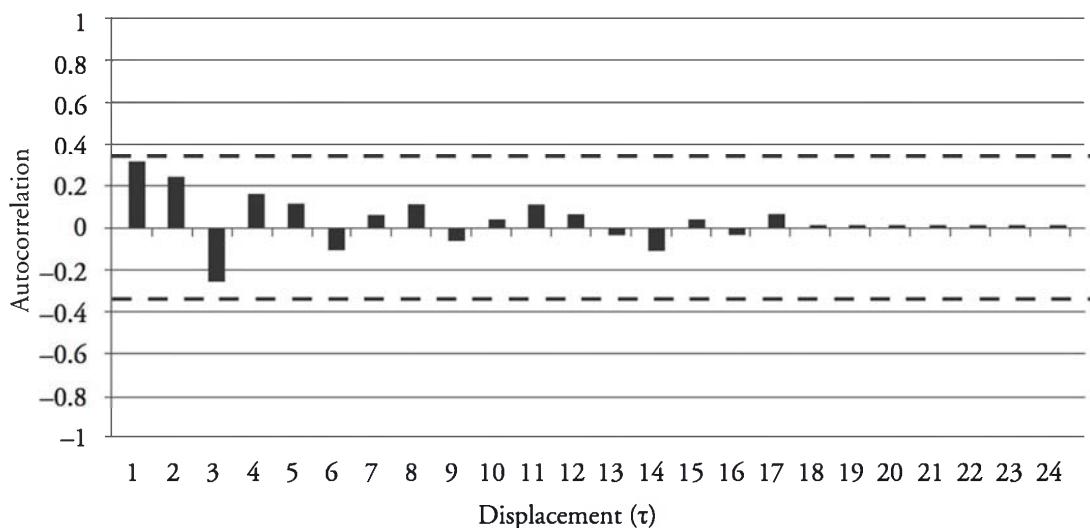
The set of sample autocorrelations for a time series is its **correlogram** or **sample autocorrelation function**.–

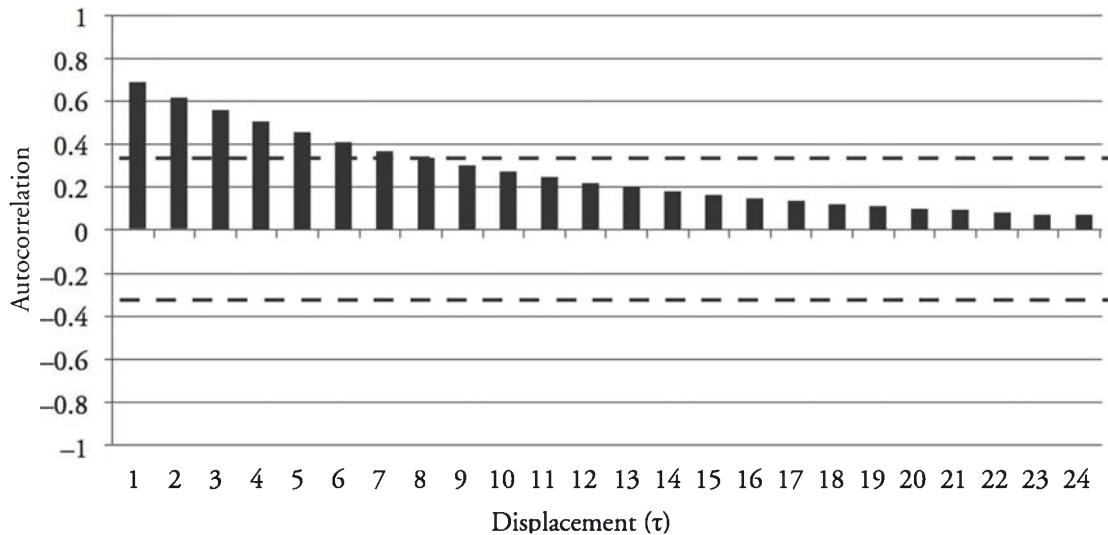
Earlier we defined **partial autocorrelations** as the results from a linear regression of a time series against its lagged values. Theoretically, this assumed an infinite set of observations. If we perform such a regression with a finite sample of time series data, what we actually get are **sample partial autocorrelations**.

If we have  $T$  observations of a time series, the standard deviation of its autocorrelations or partial correlations is  $1/\sqrt{T}$ . One of the ways to determine whether a time series can be considered white noise is by displaying its autocorrelation and partial autocorrelation functions with bands at  $\pm 2/\sqrt{T}$ , as in Figure 4. We would expect 95% of the sample autocorrelations and sample partial autocorrelations to fall within this interval if a time series is white noise.

**Figure 4: Autocorrelation Functions With Two-Standard-Error Bands**

a. Likely white noise



**b. Likely not white noise**

A more rigorous way of determining whether a time series is white noise is to test the hypothesis where autocorrelations are jointly equal to zero. A test statistic for this hypothesis is the **Box-Pierce Q-statistic**, which follows a Chi-squared distribution. A similar test statistic that may be more useful with small samples is the **Ljung-Box Q-statistic**.



*Professor's Note: We do not believe FRM candidates are required to calculate these test statistics or perform this hypothesis test for the exam. Knowing the null hypothesis and that these are Chi-squared tests should be sufficient.*

## KEY CONCEPTS

---

### LO 26.1

A time series is covariance stationary if the relationships among its present and past values remain stable over time.

The covariance between the current value of a time series and its value  $\tau$  periods in the past is referred to as its autocovariance at displacement  $\tau$ . Its autocovariances for all  $\tau$  make up its autocovariance function.

To convert an autocovariance function to an autocorrelation function, we divide the autocovariance at each  $\tau$  by the variance of the time series.

Autoregression is a linear regression of a time series against its own past values at a specific lag. The regression coefficient that results is referred to as the partial autocorrelation for that lag. Partial autocorrelations for all lags make up the partial autocorrelation function.

---

### LO 26.2

To be covariance stationary, a time series must exhibit the following three properties:

1. Its mean must be stable over time.
  2. Its variance must be finite and stable over time.
  3. Its covariance structure must be stable over time.
- 

### LO 26.3

If a time series is not covariance stationary, we cannot model it directly from its past values. However, we can often transform a series and find a covariance stationary underlying process—for example, by filtering out trend and seasonality.

---

### LO 26.4

White noise refers to a serially uncorrelated time series that has a mean of zero and a constant variance. A time series is serially uncorrelated if it exhibits no correlation among any of its lagged values.

If the observations in a white noise process are independent, as well as uncorrelated, the process is referred to as independent white noise. If the process also follows a normal distribution, it is known as normal or Gaussian white noise.

---

### LO 26.5

Events in a white noise process exhibit no correlation between the past and present. Thus, its autocorrelations and partial autocorrelations are zero for any displacement. Although its

unconditional mean is zero and its variance is constant, a process may have a conditional mean and variance that are not necessarily constant. If so, past values may be useful for forecasting.

### LO 26.6

The notation for a lag operator is  $L^m y_t = y_{t-m}$ , where  $y_t$  is the value of a time series at time  $t$ , and  $m$  is a number of periods before time  $t$ .

A distributed lag is a model that assigns weights to past values of a time series. A lag operator polynomial of degree  $m$  is a distributed lag that includes all lags from 1 to  $m$ .

### LO 26.7

Wold's theorem, or Wold's representation, proposes that a covariance stationary process can be modeled as an infinite distributed lag of a white noise process as follows:

$$\varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \dots = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$$

### LO 26.8

Wold's representation is an example of a general linear process because it can be applied to any covariance stationary series.

### LO 26.9

Wold's representation can be approximated with a ratio of rational distributed lags, which are distributed lags with a finite number of terms. Such an approximation has only as many terms as the two rational distributed lags in the ratio.

### LO 26.10

The sample mean of a time series is the arithmetic average of its observations. The sample autocorrelation for displacement  $\tau$  is estimated by:

$$\widehat{\rho}_\tau = \frac{\sum_{t=\tau+1}^T [(y_t - \bar{y})(y_{t-\tau} - \bar{y})]}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

The Box-Pierce Q-statistic is a test statistic for the hypothesis that the autocorrelations of a time series are jointly equal to zero. The Ljung-Box Q-statistic is similar but may be more useful with small samples. Both statistics follow a Chi-squared distribution.

**LO 26.11**

While partial autocorrelations are theoretically based on an infinite-length time series, sample partial autocorrelations can be estimated from a finite number of time series observations.

## CONCEPT CHECKERS

1. The conditions for a time series to exhibit covariance stationarity least likely include:
  - A. a stable mean.
  - B. a finite variance.
  - C. a finite number of observations.
  - D. autocovariances that do not depend on time.
2. Which of the following statements about white noise is most accurate?
  - A. All serially uncorrelated processes are white noise.
  - B. All Gaussian white noise processes are independent white noise.
  - C. All independent white noise processes are Gaussian white noise.
  - D. All serially uncorrelated Gaussian processes are independent white noise.
3. An analyst is working with statistical software to forecast a time series called “Series01.” The software produces an output that includes the following reference:

$L^4(\text{Series01})$

The analyst should most appropriately interpret this reference as indicating:

- A. the value of Series01 at time  $t - 4$ .
  - B. that Series01 is a nonlinear process.
  - C. the autocorrelation of Series01 at  $\tau = 4$ .
  - D. a lag operator polynomial of degree 4 for Series01.
4. For a time series with  $T$  observations, the bands displayed on a graph of its correlogram are most likely to be located at plus-or-minus:
    - A.  $1/T$
    - B.  $2/T$
    - C.  $1/\sqrt{T}$
    - D.  $2/\sqrt{T}$
  5. To test the hypothesis that the autocorrelations of a time series are jointly equal to zero based on a small sample, an analyst should most appropriately calculate:
    - A. a Ljung-Box Q-statistic.
    - B. a Box-Pierce Q-statistic.
    - C. either a Ljung-Box or a Box-Pierce Q-statistic.
    - D. neither a Ljung-Box nor a Box-Pierce Q-statistic.

## CONCEPT CHECKER ANSWERS

1. C In theory, a time series can be infinite in length and still be covariance stationary. To be covariance stationary, a time series must have a stable mean, a stable covariance structure (i.e., autocovariances depend only on displacement, not on time), and a finite variance.
2. B If a white noise process is Gaussian (i.e., normally distributed), it follows that the process is independent white noise. However, the reverse is not true; there can be independent white noise processes that are not normally distributed. Only those serially uncorrelated processes that have a zero mean and constant variance are white noise.
3. A The lag operator  $L^4(\text{Series01})$  refers to the value of Series01 at time  $t - 4$ .
4. D The bands displayed on a correlogram graph are typically two standard errors from zero, or  $\pm 2/\sqrt{T}$ .
5. A The Ljung-Box Q-statistic is appropriate for testing this hypothesis based on a small sample.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

# MODELING CYCLES: MA, AR, AND ARMA MODELS

Topic 27

## EXAM FOCUS

Moving average (MA) processes can be used to capture the relationship between a time series variable and its current and lagged random shocks. This is useful for researchers if an event is mostly described by random shocks. However, it becomes even more useful when it is transformed into an autoregressive representation. An autoregressive (AR) process attempts to capture how a time series variable's lagged observations of itself combine with random shocks to forecast a variable. Sometimes forecasters need a combination of these two concepts to improve the usefulness of a forecasting model, which results in an autoregressive moving average model (ARMA). For the exam, understand the properties of an MA(1) process and an AR(1) process and how they can be broaden to incorporate additional lag operators. Also, be able to describe an ARMA process and understand its applications.

## FIRST-ORDER MOVING AVERAGE PROCESS

**LO 27.1: Describe the properties of the first-order moving average (MA(1)) process, and distinguish between autoregressive representation and moving average representation.**

Conceptually, a moving average process is a linear regression of the current values of a time series against both the current and previous unobserved white noise error terms, which are random shocks. The first-order moving average [MA(1)] process has a mean of zero and a constant variance and can be defined as:

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

where:

$y_t$  = the time series variable being estimated

$\varepsilon_t$  = current random white noise shock

$\varepsilon_{t-1}$  = one-period lagged random white noise shock

$\theta$  = coefficient for the lagged random shock

The MA(1) process is considered to be first-order because it only has one lagged error term ( $\varepsilon_{t-1}$ ). This yields a very short-term memory because it only incorporates what happens one period ago. If we ignore the lagged error term for a moment and assume that  $\varepsilon_t > 0$ , then  $y_t > 0$ . This is equivalent to saying that a positive error term will yield a positive dependent variable ( $y_t$ ). When adding back the lagged error term, we are now saying that the dependent variable is impacted by not only the current error term, but also the previous

period's unobserved error term, which is amplified by a coefficient ( $\theta$ ). Consider an example using daily demand for ice cream ( $y_t$ ) to better understand how this works:

$$y_t = \varepsilon_t + 0.3\varepsilon_{t-1}$$

In this equation, the error term ( $\varepsilon_t$ ) is the daily change in temperature. Using only the current period's error term ( $\varepsilon_t$ ), if the daily change in temperature is positive, then we would estimate that daily demand for ice cream would also be positive. But, if the daily change yesterday ( $\varepsilon_{t-1}$ ) was also positive, then we would expect an amplified impact on our daily demand for ice cream by a factor of 0.3.

One key feature of moving average processes is called the *autocorrelation ( $\rho$ ) cutoff*. We would compute the autocorrelation using the following formula:

$$\rho_1 = \frac{\theta_1}{1 + \theta_1^2}; \text{ where } \rho_\tau = 0 \text{ for } \tau > 1$$

Using the previous example of estimating daily demand for ice cream with  $\theta = 0.3$ , we would compute the autocorrelation to be 0.2752 as follows:

$$0.2752 = \frac{0.3}{1 + 0.3^2}$$

For any value beyond the first lagged error term, the autocorrelation will be zero in an MA(1) process. This is important because it is one condition of being covariance stationary (i.e., mean = 0, variance =  $\sigma^2$ ), which is a condition of this process being a useful estimator.

It is also important to note that this **moving average representation** has both a current random shock ( $\varepsilon_t$ ) and a lagged unobservable shock ( $\varepsilon_{t-1}$ ) on the independent side of this equation. This presents a problem for forecasting in the real world because it does not incorporate observable shocks. The solution for this problem is known as an **autoregressive representation** where the MA(1) process formula is inverted so we have a lagged shock and a lagged value of the time series itself. The condition for inverting an MA(1) process is  $|\theta| < 1$ . The autoregressive representation, which is an algebraic rearrangement of the MA(1) process formula, is expressed in the following formula:

$$\varepsilon_t = y_t - \theta\varepsilon_{t-1}$$

This process of inversion enables the forecaster to express current observables in terms of past observables.

## MA(q) PROCESS

---

### LO 27.2: Describe the properties of a general finite-order process of order q (MA(q)) process.

---

The MA(1) process is a subset of a much larger picture. Forecasters can broaden their horizon to a finite-order moving average process of order  $q$ , which essentially adds lag operators out to the  $q^{\text{th}}$  observation and potentially improves on the MA(1) process. The MA(q) process is expressed in the following formula:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

where:

- $y_t$  = the time series variable being estimated
- $\varepsilon_t$  = current random white noise shock
- $\varepsilon_{t-1}$  = one-period lagged random white noise shock
- $\varepsilon_{t-q}$  =  $q^{\text{th}}$ -period lagged random white noise shock
- $\theta$  = coefficients for the lagged random shocks

The MA(q) process theoretically captures complex patterns in greater detail, which can potentially provide for more robust forecasting. This also lengthens the memory from one period to the  $q^{\text{th}}$  period. Returning to the previous example, using the demand for ice cream, a forecaster could use not only the current and previous day's changes in temperature to predict ice cream demand, but also the entire previous week's demand to enhance the informational value of the estimation.

Just as the MA(1) process exhibits autocorrelation cutoff after the first lagged error term, the MA(q) process experiences autocorrelation cutoff after the  $q^{\text{th}}$  lagged error term. Again, this is important because covariance stationarity is essential to the predictive ability of the model.

## FIRST-ORDER AUTOREGRESSIVE PROCESS

---

### LO 27.3: Describe the properties of the first-order autoregressive (AR(1)) process, and define and explain the Yule-Walker equation.

---

We have seen that when a moving average process is inverted it becomes an autoregressive representation, and is, therefore, more useful because it expresses the current observables in terms of past observables. An autoregressive process does not need to be inverted because it is already in the more favorable rearrangement, and is, therefore, capable of capturing a more robust relationship compared to the unadjusted moving average process. The first-order autoregressive [AR(1)] process must also have a mean of zero and a constant variance.

It is specified in the form of a variable regressed against itself in a lagged form. This relationship can be shown in the following formula:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

where:

$y_t$  = the time series variable being estimated

$y_{t-1}$  = one-period lagged observation of the variable being estimated

$\varepsilon_t$  = current random white noise shock

$\phi$  = coefficient for the lagged observation of the variable being estimated

Just like the moving average process, the predictive ability of this model hinges on it being covariance stationary. In order for an AR(1) process to be covariance stationary, the absolute value of the coefficient on the lagged operator must be less than one (i.e.,  $|\phi| < 1$ ).

Using our previous example of daily demand for ice cream, we would forecast our current period daily demand ( $y_t$ ) as a function of a coefficient ( $\phi$ ) multiplied by our lagged daily demand for ice cream ( $y_{t-1}$ ) and then add a random error shock ( $\varepsilon_t$ ). This process enables us to use a past observed variable to predict a current observed variable.

In order to estimate the autoregressive parameters, such as the coefficient ( $\phi$ ), forecasters need to accurately estimate the autocovariance of the data series. The **Yule-Walker equation** is used for this purpose. When using the Yule-Walker concept to solve for the autocorrelations of an AR(1) process, we use the following relationship:

$$\rho_t = \phi^t \text{ for } t = 0, 1, 2, \dots$$

The Yule-Walker equation is used to reinforce a very important distinction between autoregressive processes and moving average processes. Recall that moving average processes exhibit autocorrelation cutoff, which means the autocorrelations are essentially zero beyond the order of the process [an MA(1) process shows autocorrelation cutoff after time 1]. The significance of the Yule-Walker equation is that for autoregressive processes, the autocorrelation decays very gradually. Consider an AR(1) process that is specified using the following formula:

$$y_t = 0.65y_{t-1} + \varepsilon_t$$

The coefficient ( $\phi$ ) is equal to 0.65, and using the concept derived from the Yule-Walker equation, the first-period autocorrelation is 0.65 (i.e.,  $0.65^1$ ), the second-period autocorrelation is 0.4225 (i.e.,  $0.65^2$ ), and so on for the remaining autocorrelations.



*Professor's Note: While autocorrelation cutoff is a hallmark of moving average processes, a gradual decay in autocorrelations is a sure sign that a forecaster is dealing with an autoregressive process.*

It should also be noted that if the coefficient ( $\phi$ ) were to be a negative number, perhaps  $-0.65$ , then the decay would still occur but the graph would oscillate between negative and positive numbers. This is true because  $-0.65^3 = -0.2746$ ,  $-0.65^4 = 0.1785$ , and  $-0.65^5 = -0.1160$ . You would still notice the absolute value decaying, but the actual autocorrelations would alternate between positive and negative numbers over time.

## AR(p) PROCESS

---

### LO 27.4: Describe the properties of a general $p^{\text{th}}$ order autoregressive (AR(p)) process.

---

Just as the MA(1) process was described as a subset of the much broader MA(q) process, so is the relationship between the AR(1) process and the AR(p) process. The AR(p) process expands the AR(1) process out to the  $p^{\text{th}}$  observation as seen in the following formula:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

where:

$y_t$  = the time series variable being estimated

$y_{t-1}$  = one-period lagged observation of the variable being estimated

$y_{t-p}$  =  $p^{\text{th}}$ -period lagged observation of the variable being estimated

$\varepsilon_t$  = current random white noise shock

$\phi$  = coefficients for the lagged observations of the variable being estimated

The AR(p) process is also covariance stationary if  $|\phi| < 1$  and it exhibits the same decay in autocorrelations that was found in the AR(1) process. However, while an AR(1) process only evidences oscillation in its autocorrelations (switching from positive to negative) when the coefficient is negative, an AR(p) process will naturally oscillate as it has multiple coefficients interacting with each other.

## AUTOREGRESSIVE MOVING AVERAGE PROCESS

---

### LO 27.5: Define and describe the properties of the autoregressive moving average (ARMA) process.

---

So far, we have examined moving average processes and autoregressive processes assuming they interact independently of each other. While this may be the case, it is possible for a time series to show signs of both processes and theoretically capture a still richer relationship. For example, stock prices might show evidence of being influenced by both unobserved shocks (the moving average component) and their own lagged behavior (the

autoregressive component). This more complex relationship is called an **autoregressive moving average (ARMA) process** and is expressed by the following formula:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

where:

- $y_t$  = the time series variable being estimated
- $\phi$  = coefficient for the lagged observations of the variable being estimated
- $y_{t-1}$  = one-period lagged observation of the variable being estimated
- $\varepsilon_t$  = current random white noise shock
- $\theta$  = coefficient for the lagged random shocks
- $\varepsilon_{t-1}$  = one-period lagged random white noise shock

You can see that the ARMA formula merges the concepts of an AR process and an MA process. In order for the ARMA process to be covariance stationary, which is important for forecasting, we must still observe  $|\theta| < 1$ . Just as with the AR process, the autocorrelations in an ARMA process will also decay gradually for essentially the same reasons.

Consider an example regarding sales of an item ( $y_t$ ) and a random shock of advertising ( $\varepsilon_t$ ). We could attempt to forecast sales for this item as a function of the previous period's sales ( $y_{t-1}$ ), the current level of advertising ( $\varepsilon_t$ ), and the one-period lagged level of advertising ( $\varepsilon_{t-1}$ ). It makes intuitive sense that sales in the current period could be affected by both past sales and by random shocks, such as advertising. Another possible random shock for sales could be the seasonal effects of weather conditions.



*Professor's Note: Just as moving average models can be extrapolated to the  $q^{\text{th}}$  observation and autoregressive models can be taken out to the  $p^{\text{th}}$  observation, ARMA models can be used in the format of an ARMA( $p, q$ ) model. For example, an ARMA(3,1) model means 3 lagged operators in the AR portion of the formula and 1 lagged operator on the MA portion. This flexibility provides the highest possible set of combinations for time series forecasting of the three models discussed in this topic.*

## APPLICATION OF AR AND ARMA PROCESSES

---

### LO 27.6: Describe the application of AR and ARMA processes.

---

A forecaster might begin by plotting the autocorrelations for a data series and find that the autocorrelations decay gradually rather than cut off abruptly. In this case, the forecaster should rule out using a moving average process. If the autocorrelations instead decay gradually, he should consider specifying either an autoregressive (AR) process or an autoregressive moving average (ARMA) process. The forecaster should especially consider these alternatives if he notices periodic spikes in the autocorrelations as they are gradually decaying. For example, if every 12<sup>th</sup> autocorrelation jumps upward, this observation indicates a possible seasonality effect in the data and would heavily point toward using either an AR or ARMA model.

Another way of looking at model applications is to test various models using regression results. It is easiest to see the differences using data that follows some pattern of seasonality, such as employment data. In the real world, a moving average process would not specify a very robust model, and autocorrelations would decay gradually, so forecasters would be wise to consider both AR models and ARMA models for employment data.

We could begin with a base AR(2) model that adds in a constant value ( $\mu$ ) if all other values are zero. This is shown in the following generic formula:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Applying actual coefficients, our real AR(2) model might look something like:

$$y_t = 101.2413 + 1.4388 y_{t-1} - 0.4765 y_{t-2} + \varepsilon_t$$

We could also try to forecast our seasonally impacted employment data with an ARMA(3,1) model, which might look like the following formula:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \theta \varepsilon_{t-1} + \varepsilon_t$$

Applying actual coefficients our real ARMA(3,1) model might look something like:

$$y_t = 101.1378 + 0.5004 y_{t-1} + 0.8722 y_{t-2} - 0.4434 y_{t-3} + 0.9709 \varepsilon_{t-1} + \varepsilon_t$$

In practice, researchers would attempt to determine whether the AR(2) model or the ARMA(3,1) model provides a better prediction for the seasonally impacted data series.

## KEY CONCEPTS

---

### LO 27.1

The first-order moving average process enables forecasters to consider the likely current effect on a dependent variable of current and lagged white noise error terms. While this is a useful process, it is most useful when inverted as an autoregressive representation so that current observables can be explained in terms of past observables.

---

### LO 27.2

While the first-order moving average process does provide useful information for forecasting, the  $q$ th-order moving average process allows for a richer analysis because it incorporates significantly more lagged error terms all the way out to the order of  $q$ .

---

### LO 27.3

The first-order autoregressive process incorporates the benefits of an inverted MA(1) process. Specifically, the AR(1) process seeks to explain the dependent variable in terms of a lagged observation of itself and an error term. This is a better forecasting tool if the autocorrelations decay gradually rather than cut off immediately after the first observation with a first-order process.

---

### LO 27.4

The  $p$ th-order autoregressive process adds additional lagged observations of the dependent variable and enhances the informational value relative to an AR(1) process in much the same way that an MA( $q$ ) process adds a richer explanation to the MA(1) process.

---

### LO 27.5

The autoregressive moving average (ARMA) process has the potential to capture more robust relationships. The ARMA process incorporates the lagged error elements of the moving average process and the lagged observations of the dependent variable from the autoregressive process.

---

### LO 27.6

Both autoregressive (AR) and autoregressive moving average (ARMA) processes can be applied to time series data that show signs of seasonality. Seasonality is most apparent when the autocorrelations for a data series do not abruptly cut off, but rather decay gradually with periodic spikes.

## CONCEPT CHECKERS

1. In practice, the moving average representation of a first-order moving average [MA(1)] process presents a problem. Which of the following statements represents that problem and how can it be resolved? The problem is that a moving average representation of an MA(1) process:
  - A. does not incorporate observable shocks, so the solution is to use a moving average representation.
  - B. incorporates only observable shocks, so the solution is to use a moving average representation.
  - C. does not incorporate observable shocks, so the solution is to use an autoregressive representation.
  - D. incorporates only observable shocks, so the solution is to use an autoregressive representation.
2. Which of the following statements is a key differentiator between a moving average (MA) representation and an autoregressive (AR) process?
  - A. A moving average representation shows evidence of autocorrelation cutoff.
  - B. An autoregressive process shows evidence of autocorrelation cutoff.
  - C. An unadjusted moving average process shows evidence of gradual autocorrelation decay.
  - D. An autoregressive process is never covariance stationary.
3. The purpose of a  $q^{\text{th}}$ -order moving average process is to:
  - A. add exactly two additional lagged variables to the original specification.
  - B. add a second error term to an MA(1) process.
  - C. invert the moving average process to make the formula more useful.
  - D. add as many additional lagged variables as needed to more robustly estimate the data series.
4. Which of the following statements about an autoregressive moving average (ARMA) process is correct?
  - I. It involves autocorrelations that decay gradually.
  - II. It combines the lagged unobservable random shock of the MA process with the observed lagged time series of the AR process.
  - A. I only.
  - B. II only.
  - C. Both I and II.
  - D. Neither I nor II.
5. Which of the following statements is correct regarding the usefulness of an autoregressive (AR) process and an autoregressive moving average (ARMA) process when modeling seasonal data?
  - I. They both include lagged terms and, therefore, can better capture a relationship in motion.
  - II. They both specialize in capturing only the random movements in time series data.
  - A. I only
  - B. II only.
  - C. Both I and II.
  - D. Neither I nor II.

## CONCEPT CHECKER ANSWERS

1. C The problem with a moving average representation of an MA(1) process is that it attempts to estimate a variable in terms of unobservable white noise random shocks. If the formula is inverted into an autoregressive representation, then it becomes more useful for estimation because an observable item is now being used.
2. A A key difference between a moving average (MA) representation and an autoregressive (AR) process is that the MA process shows autocorrelation cutoff while an AR process shows a gradual decay in autocorrelations.
3. D The whole point of using more independent variables in a  $q^{\text{th}}$ -order moving average process is to capture a better estimation of the dependent variable. More lagged operators often provide a more robust estimation.
4. C The autoregressive moving average (ARMA) process is important because its autocorrelations decay gradually and because it captures a more robust picture of a variable being estimated by including both lagged random shocks and lagged observations of the variable being estimated. The ARMA model merges the lagged random shocks from the MA process and the lagged time series variables from the AR process.
5. A Both autoregressive (AR) models and autoregressive moving average (ARMA) models are good at forecasting with seasonal patterns because they both involve lagged observable variables, which are best for capturing a relationship in motion. It is the moving average representation that is best at capturing only random movements.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

# VOLATILITY

## Topic 28

### EXAM FOCUS

Traditionally, volatility has been synonymous with risk. Thus, the accurate estimation of volatility is crucial to understanding potential risk exposure. This topic pertains to methods that employ historical data when generating estimates of volatility. Simplistic models tend to generate estimates assuming volatility remains constant over short time periods. Conversely, complex models account for variations over time. For the exam, be able to estimate volatility using both the exponentially weighted moving average (EWMA) and the generalized autoregressive conditional heteroskedasticity [GARCH(1,1)] models.

### VOLATILITY, VARIANCE, AND IMPLIED VOLATILITY

#### LO 28.1: Define and distinguish between volatility, variance rate, and implied volatility.

The volatility of a variable,  $\sigma$ , is represented as the standard deviation of that variable's continuously compounded return. With option pricing, volatility is typically expressed as the standard deviation of return over a one-year period. This differs from risk management, where volatility is typically expressed as the standard deviation of return over a one-day period.

The traditional measure of volatility first requires a measure of change in asset value from period to period. The calculation of a continuously compounded return over successive days is as follows:

$$u_i = \ln\left(\frac{S_i}{S_{i-1}}\right)$$

where:

$S_i$  = asset price at time  $i$

This is similar to the proportional change in an asset, which is calculated as follows:

$$u_i = \frac{S_i - S_{i-1}}{S_{i-1}}$$

From a risk management perspective, the daily volatility of an asset usually refers to the standard deviation of the daily proportional change in asset value.

By assuming daily returns are independent with the same level of variation, daily volatility can be extended over a number of days,  $T$ , by multiplying the standard deviation of the return by the square root of  $T$ . This is known as the *square root of time rule*. For example, if the daily volatility is 1.5%, the standard deviation of the return (compounded continuously) over a 10-day period would be computed as  $1.5\% \times \sqrt{10} = 4.74\%$ . Note that when converting daily volatility to annual volatility, the usual practice is to use the square root of 252 days, which is the number of business days in a year, as opposed to the number of calendar days in a year.

Risk managers may also compute a variable's **variance rate**, which is simply the square of volatility (i.e., standard deviation squared:  $\sigma^2$ ). In contrast to volatility, which increases with the square root of time, the variance of an asset's return will increase in a linear fashion over time. For example, if the daily volatility is 1.5%, the variance rate is  $1.5\%^2 = 0.0225\%$ . Thus, over a 10-day period, the variance will be 0.225% (i.e.,  $0.0225\% \times 10$ ).

In addition to variance and standard deviation, which are computed using historical data, risk managers may also derive implied volatilities. The **implied volatility** of an option is computed from an option pricing model, such as the Black-Scholes-Merton (BSM) model. The volatility of an asset is not directly observed in the BSM model, so we compute implied volatility as the volatility level that will result when equating an option's market price to its model price.



*Professor's Note: Computing option prices using the BSM model will be demonstrated in Book 4.*

The most widely used index for publishing implied volatility is the Chicago Board Options Exchange (CBOE) Volatility Index (ticker symbol: VIX). The VIX demonstrates implied volatility on a wide variety of 30-day calls and puts on the S&P 500 Index. Note that trading in futures and options on the VIX is a bet on volatility only. Since its inception, the VIX has mainly traded between 10 and 20 (which corresponds to volatility of 10%–20% on the S&P 500 Index options), but it reached a peak of close to 80 in October 2008, after the collapse of Lehman Brothers. The VIX is often referred to as the fear index by market participants because it reflects current market uncertainties.

## THE POWER LAW

---

### LO 28.2: Describe the power law.

---

It is typically assumed that the change in asset prices is normally distributed. This makes it convenient to apply standard deviation when determining confidence intervals for an asset's price. For example, by assuming an asset price of \$50 and a volatility of 4.47%, we can compute a one-standard-deviation move as  $50 \times 0.0447 = 2.24$ . With this information, we can define the 95% confidence interval as  $50 \pm 1.96 \times 2.24$ .

In practice, however, the distribution of asset price changes is more likely to exhibit fatter tails than the normal distribution. Thus, heavy-tailed distributions can be used to better capture the possibility of extreme price movements (e.g., a five-standard-deviation move). An alternative approach to assuming a normal distribution is to apply the power law.

The power law states that when  $X$  is large, the value of a variable  $V$  has the following property:

$$P(V > X) = K \times X^{-\alpha}$$

where:

$V$  = the variable

$X$  = large value of  $V$

$K$  and  $\alpha$  = constants

#### Example: The power law

Assume that data on asset price changes determines the constants in the power law equation to be the following:  $K = 10$  and  $\alpha = 5$ . Calculate the probability that this variable will be greater than a value of 3 and a value of 5.

Answer:

$$P(V > 3) = 10 \times 3^{-5} = 0.0412 \text{ or } 4.12\%$$

$$P(V > 5) = 10 \times 5^{-5} = 0.0032 \text{ or } 0.32\%$$

By taking the logarithm of both sides in the power law equation, we can perform regression analysis to determine the power law constants,  $K$  and  $\alpha$ :

$$\ln[P(V > X)] = \ln(K) - \alpha \ln(X)$$

In this case, the dependent variable,  $\ln[P(V > X)]$ , can be plotted against the independent variable,  $\ln(X)$ . Furthermore, if we assume that  $X$  represents the number of standard deviations that a given variable will change, we can determine the probability that  $V$  will exceed a certain number of standard deviations. For example, if regression analysis indicates that  $K = 8$  and  $\alpha = 5$ , the probability that the variable will exceed four standard deviations will be equal to  $8 \times 4^{-5} = 0.0078$  or 0.78%. The power law suggests that extreme movements have a very low probability of occurring, but this probability is still higher than what is indicated by the normal distribution.

## ESTIMATING VOLATILITY

---

**LO 28.3: Explain how various weighting schemes can be used in estimating volatility.**

---

By collecting continuously compounded return data,  $u_i$ , over a number of days (as shown in LO 28.1), we can compute the mean return of the individual returns as follows:

$$\bar{u} = \frac{1}{m} \sum_{i=1}^m u_{n-i}$$

where:

$m$  = number of observations leading up to the present period

If we assume that the mean return is zero, which would be true when the mean is small compared to the variability, we obtain the maximum likelihood estimator of variance:

$$\sigma_n^2 = \frac{1}{m} \sum_{i=1}^m u_{n-i}^2$$

In simplest terms, historical data is used to generate returns in an asset-pricing series. This historical return information is then used to generate a volatility parameter, which can be used to infer expected realizations of risk. However, the straightforward approaches just presented weight each observation equally in that more distant past returns have the same influence on estimated volatility as observations that are more recent. If the goal is to estimate the current level of volatility, we may want to weight recent data more heavily. There are various weighting schemes, which can all essentially be represented as:

$$\sigma_n^2 = \sum_{i=1}^m \alpha_i u_{n-i}^2$$

where:

$\alpha_i$  = weight on the return  $i$  days ago

The weights (the  $\alpha$ 's) must sum to one, and if the objective is to generate a greater influence on recent observations, then the  $\alpha$ 's will decline in value for older observations.

One extension to this weighting scheme is to assume a long-run variance level in addition to the weighted squared return observations. The most frequently used model is an **autoregressive conditional heteroskedasticity model**, ARCH(m), which can be represented by:

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^m \alpha_i u_{n-i}^2 \text{ with } \gamma + \sum \alpha_i = 1 \text{ so that}$$

$$\sigma_n^2 = \omega + \sum_{i=1}^m \alpha_i u_{n-i}^2$$

where:

$\omega = \gamma V_L$  (long-run variance weighted by the parameter  $\gamma$ )

Therefore, the volatility estimate is a function of a long-run variance level and a series of squared return observations, whose influence declines the older the observation is in the time series of the data.

## THE EXPONENTIALLY WEIGHTED MOVING AVERAGE MODEL

**LO 28.4:** Apply the exponentially weighted moving average (EWMA) model to estimate volatility.

**LO 28.8:** Explain the weights in the EWMA and GARCH(1,1) models.

The exponentially weighted moving average (EWMA) model is a specific case of the general weighting model presented in the previous section. The main difference is that the weights are assumed to decline exponentially back through time. This assumption results in a specific relationship for variance in the model:

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1 - \lambda)u_{n-1}^2$$

where:

$\lambda$  = weight on previous volatility estimate ( $\lambda$  between zero and one)

The simplest interpretation of the EWMA model is that the day- $n$  volatility estimate is calculated as a function of the volatility calculated as of day  $n - 1$  and the most recent squared return. Depending on the weighting term  $\lambda$ , which ranges between zero and one, the previous volatility and most recent squared returns will have differential impacts. High values of  $\lambda$  will minimize the effect of daily percentage returns, whereas low values of  $\lambda$  will tend to increase the effect of daily percentage returns on the current volatility estimate.

### Example: EWMA model

The decay factor in an exponentially weighted moving average model is estimated to be 0.94 for daily data. Daily volatility is estimated to be 1%, and today's stock market return is 2%. What is the new estimate of volatility using the EWMA model?

**Answer:**

$$\sigma_n^2 = 0.94 \times 0.01^2 + (1 - 0.94) \times 0.02^2 = 0.000118$$

$$\sigma_n = \sqrt{0.000118} = 1.086\%$$

One benefit of the EWMA is that it requires few data points. Specifically, all we need to calculate the variance is the current estimate of the variance and the most recent squared return. The current estimate of variance will then feed into the next period's estimate, as will this period's squared return. Technically, the only "new" piece of information for the volatility calculation will be that attributed to the squared return.

## THE GARCH(1,1) MODEL

---

**LO 28.5: Describe the generalized autoregressive conditional heteroskedasticity (GARCH (p,q)) model for estimating volatility and its properties.**

**LO 28.6: Calculate volatility using the GARCH(1,1) model.**

---

One of the most popular methods of estimating volatility is the **generalized autoregressive conditional heteroskedastic** (GARCH)(1,1) model. A GARCH(1,1) model not only incorporates the most recent estimates of variance and squared return, but also a variable that accounts for a long-run average level of variance.



*Professor's Note: In the GARCH(p,q) notation, the p stands for the number of lagged terms on historical returns squared, and the q stands for the number of lagged terms on historical volatility.*

The best way to describe a GARCH(1,1) model is to take a look at the formula representing its determination of variance, which can be shown as:

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

where:

$\alpha$  = weighting on the previous period's return

$\beta$  = weighting on the previous volatility estimate

$\omega$  = weighted long-run variance =  $\gamma V_L$

$V_L$  = long-run average variance =  $\frac{\omega}{1 - \alpha - \beta}$

$\alpha + \beta + \gamma = 1$

$\alpha + \beta < 1$  for stability so that  $\gamma$  is not negative

The EWMA is nothing other than a special case of a GARCH(1,1) volatility process, with  $\omega = 0$ ,  $\alpha = 1 - \lambda$ , and  $\beta = \lambda$ . Similar to the EWMA model,  $\beta$  represents the exponential decay rate of information. The GARCH(1,1) model adds to the information generated by the EWMA model in that it also assigns a weighting to the average long-run variance estimate. An additional characteristic of a GARCH(1,1) estimate is the implicit assumption that variance tends to revert to a long-term average level. Recognition of a mean-reverting characteristic in volatility is an important feature when pricing derivative securities such as options.

**Example: GARCH(1,1) model**

The parameters of a generalized autoregressive conditional heteroskedastic (GARCH)(1,1) model are  $\omega = 0.000003$ ,  $\alpha = 0.04$ , and  $\beta = 0.92$ . If daily volatility is estimated to be 1%, and today's stock market return is 2%, what is the new estimate of volatility using the GARCH(1,1) model, and what is the implied long-run volatility level?

**Answer:**

$$\sigma_n^2 = 0.000003 + 0.04 \times 0.02^2 + 0.92 \times 0.01^2 = 0.000111$$

$$\sigma_n = \sqrt{0.000111} = 1.054\%$$

$$\text{long-run average variance} = \frac{\omega}{(1 - \alpha - \beta)} = \frac{0.000003}{(1 - 0.04 - 0.92)} = 0.000075$$

$$\bar{\sigma} = \sqrt{0.000075} = 0.866\% = \text{long-run volatility}$$

**Mean Reversion****LO 28.7: Explain mean reversion and how it is captured in the GARCH(1,1) model.**

Empirical data indicates that volatility exhibits a mean-reverting characteristic. Given that stylized fact, a GARCH model tends to display a better theoretical justification than the EWMA model. The method for estimating the GARCH parameters (or weights), however, often generates outcomes that are not consistent with the model's assumptions. Specifically, the sum of the weights of  $\alpha$  and  $\beta$  are sometimes greater than one, which causes instability in the volatility estimation. In this case, the analyst must resort to using an EWMA model.

The sum of  $\alpha + \beta$  is called the **persistence**, and if the model is to be stationary over time (with reversion to the mean), the sum must be less than one. The persistence describes the rate at which the volatility will revert to its long-term value following a large movement. The higher the persistence (given that it is less than one), the longer it will take to revert to the mean following a shock or large movement. A persistence of one means that there is no reversion, and with each change in volatility, a new level is attained.

**ESTIMATION AND PERFORMANCE OF GARCH MODELS**

As was previously mentioned, one way to estimate volatility (e.g., variance) is to use a **maximum likelihood estimator**. Maximum likelihood estimators select values of model parameters that maximize the likelihood that the observed data will occur in a sample. Any variable of interest can be estimated via the maximum likelihood method, which requires formulating an expression or function for the underlying probability distribution of the data and then searching for the parameters that maximize the value generated by the expression.

One important consideration relates to which distribution is chosen when calculating probability. The most popular is the normal distribution, but normally distributed data are not often found in financial markets.

GARCH models are estimated using maximum likelihood techniques. The estimation process begins with a guess of the model's parameters. Then a calculation of the likelihood function based on those parameter estimates is made. The parameters are then slightly adjusted until the likelihood function fails to increase, at which time the estimation process assumes it has maximized the function and stops. The values of the parameters at the point of maximum value in the likelihood function are then used to estimate GARCH model volatility.

---

**LO 28.9: Explain how GARCH models perform in volatility forecasting.****LO 28.10: Describe the volatility term structure and the impact of volatility changes.**

---

One of the useful features of GARCH models is that they do a very good job at modeling volatility clustering when periods of high volatility tend to be followed by other periods of high volatility and periods of low volatility tend to be followed by subsequent periods of low volatility. Thus, there is autocorrelation in  $u_i^2$ . If GARCH models do a good job of explaining volatility changes, there should be very little autocorrelation in  $u_i^2 / \sigma_i^2$ . GARCH models appear to do a very good job of explaining volatility.

The question then arises, if GARCH models do a good job at explaining past volatility, how well do they forecast future volatility? The simple answer to this question is that GARCH models do a fine job at forecasting volatility from a volatility term structure perspective (e.g., estimates of volatility given time to expiration for options). Even though the actual volatility term structure figures are somewhat different from those forecasted by GARCH models, GARCH-generated volatility data does an excellent job in predicting how the volatility term structure responds to changes in volatility. This modeling tool is quite frequently used by financial institutions when estimating exposure to various option positions.

## KEY CONCEPTS

---

### LO 28.1

The volatility of a variable is the standard deviation of that variable's continuously compounded return. The variance rate of a variable is the square of its standard deviation. Variance and standard deviation are computed using historical data. Risk managers may also compute implied volatility, which is the volatility that forces a model price (i.e., option pricing model) to equal the market price.

---

### LO 28.2

The power law is an alternative approach to using probabilities from a normal distribution. It states that when X is large, the value of a variable V has the following property, where K and  $\alpha$  are constants:

$$P(V > X) = K \times X^{-\alpha}$$

---

### LO 28.3

Historical price data is used to generate return estimates, which are then used to estimate volatility. Traditional volatility estimation methods weight past information equally across time. Weighting schemes can be used to weight recent information more heavily than distant data.

---

### LO 28.4

The EWMA model generates volatility estimates based on weightings of the last estimate of volatility and the latest current price change information. The objective is to account for previous volatility estimates, as well as to account for the latest return information.

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2$$

where:

$\lambda$  = weight on previous volatility estimate ( $\lambda$  between zero and one)

---

### LO 28.5

One of the most popular methods of estimating volatility is the generalized autoregressive conditional heteroskedastic (GARCH)(p,q) model. In a GARCH(p,q) model, the  $p$  stands for the number of lagged terms on historical returns squared, and the  $q$  stands for the number of lagged terms on historical volatility.

**LO 28.6**

GARCH(1,1) models not only incorporate the most recent estimates of volatility and return, but also incorporate a long-run average level of variance.

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

where:

$\alpha$  = weighting on the previous period's return

$\beta$  = weighting on the previous volatility estimate

$\omega$  = weighted long-run variance =  $\gamma V_L$

$V_L$  = long-run average variance =  $\frac{\omega}{1 - \alpha - \beta}$

$\alpha + \beta + \gamma = 1$

$\alpha + \beta < 1$  for stability so that  $\gamma$  is not negative

GARCH(1,1) estimates of volatility have a better theoretical justification than the EWMA model. In the event that model parameter estimates indicate instability, however, EWMA volatility estimates may be used.

---

**LO 28.7**

In a GARCH(1,1) model, the sum of  $\alpha + \beta$  is called the persistence. The persistence describes the rate at which the volatility will revert to its long-term value. A persistence equal to one means there is no mean reversion.

**LO 28.8**

The EWMA is nothing other than a special case of a GARCH(1,1) volatility process, with  $\omega = 0$ ,  $\alpha = 1 - \lambda$ , and  $\beta = \lambda$ . Similar to the EWMA model,  $\beta$  in the GARCH(1,1) equation represents the exponential decay rate of information. The GARCH(1,1) model adds to the information generated by the EWMA model in that it also assigns a weighting to the average long-run variance estimate.

---

**LO 28.9**

GARCH models do a very good job at modeling volatility clustering when periods of high volatility tend to be followed by other periods of high volatility and periods of low volatility tend to be followed by subsequent periods of low volatility.

---

**LO 28.10**

When forecasting future volatility, GARCH-generated volatility data does an excellent job in predicting the volatility term structure (i.e., differing volatilities for options given differing maturities). This modeling tool is quite frequently used by financial institutions when estimating exposure to various option positions.

## CONCEPT CHECKERS

1. An analyst is attempting to compute a confidence interval for asset Z prices. Assume a daily volatility of 1% and a current asset price of 100. What is the 95% confidence interval for asset price at the end of five days, assuming price changes are normally distributed?
  - A.  $100 \pm 1.96$ .
  - B.  $100 \pm 2.24$ .
  - C.  $100 \pm 4.39$ .
  - D.  $100 \pm 9.80$ .
2. The parameters of a generalized autoregressive conditional heteroskedastic (GARCH)(1,1) model are  $\omega = 0.00003$ ,  $\alpha = 0.04$ , and  $\beta = 0.92$ . If daily volatility is estimated to be 1.5%, and today's stock market return is 0.8%, what is the new estimate of the standard deviation?
  - A. 1.68%.
  - B. 1.55%.
  - C. 1.45%.
  - D. 2.74%.
3. The  $\lambda$  of an exponentially weighted moving average (EWMA) model is estimated to be 0.9. Daily standard deviation is estimated to be 1.5%, and today's stock market return is 0.8%. What is the new estimate of the standard deviation?
  - A. 1.68%.
  - B. 1.55%.
  - C. 1.45%.
  - D. 2.74%.
4. The parameters of a GARCH(1,1) model are  $\omega = 0.00003$ ,  $\alpha = 0.04$ , and  $\beta = 0.92$ . These figures imply a long-run daily standard deviation of:
  - A. 1.68%.
  - B. 1.55%.
  - C. 1.45%.
  - D. 2.74%.
5. GARCH(1,1) models can only be used to estimate volatility in the case where:
  - A.  $\alpha + \beta > 0$ .
  - B.  $\alpha + \beta < 1$ .
  - C.  $\alpha > \beta$ .
  - D.  $\alpha < \beta$ .

## CONCEPT CHECKER ANSWERS

1. C First, convert daily volatility to weekly volatility using the square root to time:  $1\% \times \sqrt{5} = 2.24\%$ . Next, compute the one-standard-deviation move:  $100 \times 0.0224 = 2.24$ . Finally, derive the confidence interval:  $100 \pm 1.96 \times 2.24 = 100 \pm 4.39$ .

2. B  $\sigma_n^2 = 0.00003 + (0.008)^2 \times 0.04 + (0.015)^2 \times 0.92 = 0.00023956$

$$\sigma_n = \sqrt{0.00023956} = 0.0155 = 1.55\%$$

3. C  $\sigma_n^2 = 0.9 \times (0.015)^2 + (1 - 0.9) \times (0.008)^2 = 0.0002089$

$$\sigma_n = \sqrt{0.0002089} = 0.0145 = 1.45\%$$

4. D The long-run variance rate can be estimated by dividing the  $\omega$  of a GARCH(1,1) model by  $1 - \alpha - \beta$ . This yields  $0.00003 / (1 - 0.04 - 0.92) = 0.00075$ ; long-run standard deviation =  $\sqrt{0.00075} = 0.0274 = 2.74\%$ .

5. B Stable GARCH(1,1) models require  $\alpha + \beta < 1$ ; otherwise the model is unstable.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

# CORRELATIONS AND COPULAS

## Topic 29

### EXAM FOCUS

This topic examines correlation and covariance calculations and how covariance is used in exponentially weighted moving average (EWMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models. The later part of this topic defines copulas and distinguishes between several different types of copulas. For the exam, be able to calculate covariance using EWMA and GARCH(1,1) models. Also, understand how copulas are used to estimate correlations between variables. Finally, be able to explain how marginal distributions are mapped to known distributions to form copulas.

### CORRELATION AND COVARIANCE

#### LO 29.1: Define correlation and covariance and differentiate between correlation and dependence.

Correlation and covariance refer to the co-movements of assets over time and measure the strength between the linear relationships of two variables. Correlation and covariance essentially measure the same relationship; however, correlation is standardized so the value is always between  $-1$  and  $1$ . This standardized measure is more convenient in risk analysis applications than covariance, which can have values between  $-\infty$  and  $\infty$ . Correlation is mathematically determined by dividing the covariance between two random variables,  $\text{cov}(X, Y)$ , by the product of their standard deviations,  $\sigma_X \sigma_Y$ .

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Multiplying each side of this equation by  $\sigma_X \sigma_Y$  provides the formula for calculating covariance:

$$\text{cov}(X, Y) = \rho_{X,Y} \times \sigma_X \sigma_Y$$

In practice, it is necessary to first calculate the covariance between two random variables using the following equation and then solve for the standardized correlation.

$$\text{cov}(X, Y) = E[(X - E(X)) \times (Y - E(Y))] = E(X, Y) - E(X) \times E(Y)$$

In this covariance equation,  $E(X)$  and  $E(Y)$  are the means or expected values of random variables  $X$  and  $Y$ , respectively.  $E(X,Y)$  is the expected value of the product of random variables  $X$  and  $Y$ .

Variables are defined as independent variables if the knowledge of one variable does not impact the probability distribution for another variable. In other words, the conditional probability of  $V_2$  given information regarding the probability distribution of  $V_1$  is equal to the unconditional probability of  $V_2$  as expressed in the following equation:

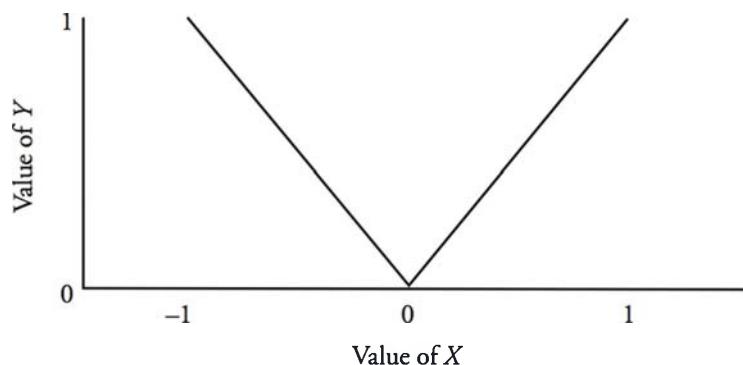
$$P(V_2 | V_1 = x) = P(V_2)$$

A correlation of zero between two variables *does not* imply that there is no dependence between the two variables. It simply implies that there is no linear relationship between the two variables, but the value of one variable can still have a nonlinear relationship with the other variable.

As an example, suppose variable  $X$  has three expected values of  $-1$ ,  $0$ , and  $1$  with an equal probability of occurrence, and variable  $Y$  has a value of  $1$  when variable  $X$  has a value of either  $-1$  or  $1$ . When variable  $X$  has a value of  $0$ , then variable  $Y$  has a value of  $0$ . This V-shaped relationship is illustrated in Figure 1.

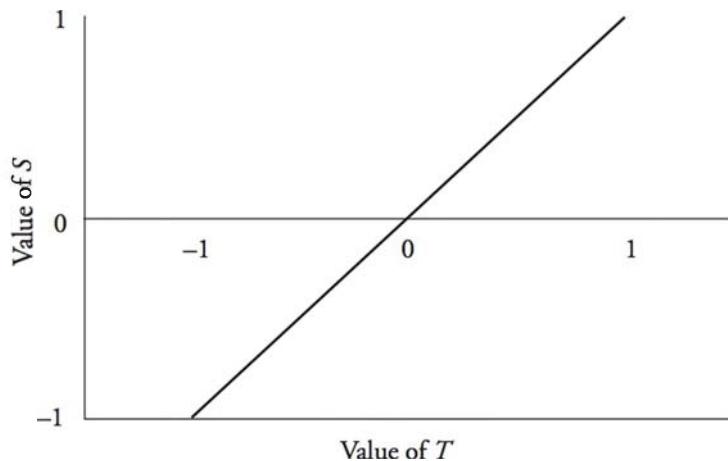
**Figure 1: Relationship Between  $X$  and  $Y$**

---



Also suppose that variables  $S$  and  $T$  are perfectly positively correlated and that variable  $S$  has three expected values of  $-1$ ,  $0$ , and  $1$  with an equal probability of occurrence. When variable  $S$  has a value of  $-1$ ,  $0$ , or  $1$ , then variable  $T$  has a value of  $-1$ ,  $0$ , and  $1$ , respectively. This relationship is illustrated in Figure 2.

Figure 2: Relationship Between  $S$  and  $T$



With this information, we can now determine the correlation coefficient and dependency of these two pairs of variables. In this example, the coefficient of correlation between variables  $X$  and  $Y$  is zero, and the coefficient of correlation between variables  $S$  and  $T$  is one.

The coefficient of correlation is a statistical measure of linear dependency. If we know the value of  $X$ , it will change our expectations of the value or probability distribution of  $Y$ . Likewise, if we know the value of  $Y$ , it will change our expectations of the probability distribution of  $X$ . Clearly, there is a dependency between  $X$  and  $Y$ , as well as a dependency between  $S$  and  $T$ . A practical example of the V-shaped dependency in Figure 1 is with respect to financial derivatives that may have more value with large market movements in either direction.

## COVARIANCE USING EWMA AND GARCH MODELS

### LO 29.2: Calculate covariance using the EWMA and GARCH(1,1) models.

#### EWMA Model

Covariance is a statistical measure that is calculated over historical time periods. Conventional wisdom suggests that more recent observations should carry more weight because they more accurately reflect the current market environment. The following equation calculates a new covariance on day  $n$  using an **exponentially weighted moving average (EWMA) model**. This model is designed to vary the weight given to more recent observations (by adjusting  $\lambda$ ).

$$\text{cov}_n = \lambda \text{cov}_{n-1} + (1 - \lambda)X_{n-1}Y_{n-1}$$

where:

$\lambda$  = the weight for the most recent covariance on day  $n - 1$

$X_{n-1}$  = the percentage change for variable  $X$  on day  $n - 1$

$Y_{n-1}$  = the percentage change for variable  $Y$  on day  $n - 1$

**Example: Calculating covariance using the EWMA model**

Assume an analyst uses the EWMA model with  $\lambda = 0.90$  to update correlation and covariance rates. The correlation estimate for two variables  $X$  and  $Y$  on day  $n - 1$  is 0.7. In addition, the estimated standard deviations on day  $n - 1$  for variables  $X$  and  $Y$  are 1.5% and 2%, respectively. Also, the percentage change on day  $n - 1$  for variables  $X$  and  $Y$  are 2% and 1%, respectively. What is the updated estimate of the covariance rate and correlation between  $X$  and  $Y$  on day  $n$ ?

**Answer:**

The estimated covariance rate between variables  $X$  and  $Y$  on day  $n - 1$  can be calculated as:

$$\text{cov}(X, Y) = \rho_{X,Y} \times \sigma_X \sigma_Y = 0.7 \times 0.015 \times 0.02 = 0.00021$$

With this value, the EWMA model can update the covariance rate for day  $n$ .

$$\text{cov}_n = 0.9 \times 0.00021 + 0.1 \times 0.02 \times 0.01 = 0.000189 + 0.00002 = 0.000209$$

Note that the covariance of an asset with itself is equal to the variance of the asset ( $\text{cov}(X, X) = \sigma_X^2$ ). Thus, the EWMA equation can also be used to estimate the new variances for variables  $X$  and  $Y$ . The modified equation for updating the variance of  $X$  becomes:

$$\sigma_{X,n}^2 = \lambda \sigma_{X,n-1}^2 + (1 - \lambda) X_{n-1}^2$$

$$\sigma_{X,n}^2 = 0.9 \times 0.015^2 + 0.1 \times 0.02^2 = 0.0002025 + 0.00004 = 0.0002425$$

Similarly, the updated variance for variable  $Y$  is calculated as follows:

$$\sigma_{Y,n}^2 = 0.9 \times 0.02^2 + 0.1 \times 0.01^2 = 0.00036 + 0.00001 = 0.00037$$

The new standard deviation estimates for  $X$  and  $Y$  are found by taking the square root of their respective variances. The new volatility measure of  $X$  is:

$$\sigma_{X,n} = \sqrt{0.0002425} = 0.0155724$$

The new volatility measure of  $Y$  is:

$$\sigma_{Y,n} = \sqrt{0.00037} = 0.0192354$$

Therefore, the new correlation on day  $n$  can be found by dividing the updated covariance ( $\text{cov}_n$ ) by the updated standard deviations for  $X$  and  $Y$ :

$$\frac{0.000209}{0.0155724 \times 0.0192354} = 0.6977$$

### GARCH(1,1) Model

An alternative method for updating the covariance rate for two variables  $X$  and  $Y$  uses the **generalized autoregressive conditional heteroskedasticity (GARCH) model**. The GARCH(1,1) model for updating covariance rates is defined as follows:

$$\text{cov}_n = \omega + \alpha X_{n-1} Y_{n-1} + \beta \text{cov}_{n-1}$$

GARCH(1,1) applies a weight of  $\alpha$  to the most recent observation on covariance ( $X_{n-1} Y_{n-1}$ ) and a weight of  $\beta$  to the most recent covariance estimate ( $\text{cov}_{n-1}$ ). In addition, a weight of  $\omega$  is given to the long-term average covariance rate.



*Professor's Note: Recall that the EWMA is a special case of GARCH(1,1), where  $\omega = 0$ ,  $\alpha = 1 - \lambda$ , and  $\beta = \lambda$ .*

#### Example: Calculating covariance using the GARCH(1,1) model

Assume an analyst uses daily data to estimate a GARCH(1,1) model as follows:

$$\text{cov}_n = 0.000002 + 0.14X_{n-1}Y_{n-1} + 0.76\text{cov}_{n-1}$$

This implies  $\alpha = 0.14$ ,  $\beta = 0.76$ , and  $\omega = 0.000002$ . The analyst also determines that the estimate of covariance on day  $n - 1$  is 0.000324 and the most recent returns on  $X$  and  $Y$  are both 0.02. What is the updated estimate of covariance?

#### Answer:

The updated estimate of covariance on day  $n$  is 0.0304%, which is calculated as:

$$\begin{aligned}\text{cov}_n &= 0.000002 + (0.14 \times 0.02^2) + (0.76 \times 0.000324) \\ &= 0.000002 + 0.000056 + 0.000246 = 0.000304\end{aligned}$$

## EVALUATING CONSISTENCY FOR COVARIANCES

### LO 29.3: Apply the consistency condition to covariance.

A variance-covariance matrix can be constructed using the calculated estimates of variance and covariance rates for a set of variables. The diagonal of the matrix represents the variance rates where  $i = j$ . The covariance rates are all other elements of the matrix where  $i \neq j$ .

A matrix is known as *positive-semidefinite* if it is internally consistent. The following expression defines the necessary condition for an  $N \times N$  variance-covariance matrix,  $\Omega$ , to be internally consistent for all  $N \times 1$  vectors  $\omega$ , where  $\omega^T$  is the transpose of vector  $\omega$ :

$$\omega^T \Omega \omega \geq 0$$

Variance and covariance rates are calculated using the same EWMA or GARCH model parameters to ensure that a positive-semidefinite model is constructed. For example, if a EWMA model uses  $\lambda = 0.95$  for estimating variances, the same EWMA and  $\lambda$  should be used to estimate covariance rates.

When small changes are made to a small positive-semidefinite matrix such as a  $3 \times 3$  matrix, the matrix will most likely remain positive-semidefinite. However, small changes to a large positive-semidefinite matrix such as  $1,000 \times 1,000$  will most likely cause the matrix to no longer be positive-semidefinite.

An example of a variance-covariance matrix that is not internally consistent is shown as follows:

$$\begin{pmatrix} 1 & 0 & 0.8 \\ 0 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}$$

Notice that the variances (i.e., diagonal of the matrix) are all equal to one. Therefore, the correlation for each pair of variables must equal the covariance for each pair of variables. This is true because the standard deviations are all equal to one. Thus, correlation is calculated as the covariance divided by one.

Also, notice that there is no correlation between the first and second variables. However, there is a strong correlation between the first and third variables as well as the second and third variables. This is very unusual to have one pair with no correlation while the other two pairs have high correlations. If we transpose a vector such that  $\omega^T = (1, 1, -1)$ , we would find that this variance-covariance matrix is not internally consistent since  $\omega^T \Omega \omega \geq 0$  is not satisfied.

Another method for testing for consistency is to evaluate the following expression:

$$\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} \leq 1$$

We can substitute data from the variance-covariance matrix into this expression because all covariances are also correlation coefficients. When computing the formula, we would determine that the left side of the expression is actually greater than the right side, indicating that the matrix is not internally consistent.

$$\begin{aligned}0^2 + 0.8^2 + 0.8^2 - 2 \times 0 \times 0.8 \times 0.8 &= 1.28 \\1.28 &> 1\end{aligned}$$

## GENERATING SAMPLES

---

### LO 29.4: Describe the procedure of generating samples from a bivariate normal distribution.

---

Suppose there is a bivariate normal distribution with two variables,  $X$  and  $Y$ . Variable  $X$  is known and the value of variable  $Y$  is conditional on the value of variable  $X$ . If variables  $X$  and  $Y$  have a bivariate normal distribution, then the expected value of variable  $Y$  is normally distributed with a mean of:

$$\mu_Y + \rho_{XY} \times \sigma_Y \times \frac{X - \mu_X}{\sigma_X}$$

and a standard deviation of:

$$\sigma_Y \sqrt{1 - \rho_{XY}^2}$$

The means,  $\mu_X$  and  $\mu_Y$ , of variables  $X$  and  $Y$  are both unconditional means. The standard deviations of variables  $X$  and  $Y$  are both unconditional standard deviations. Also note that the expected value of  $Y$  is linearly dependent on the conditional value of  $X$ .

The following procedure is used to generate two sample sets of variables from a bivariate normal distribution.

*Step 1:* Independent samples  $Z_X$  and  $Z_Y$  are obtained from a univariate standardized normal distribution. Microsoft Excel® and other software programming languages have routines for sampling random observations from a normal distribution. For example, this is done in Excel with the formula = NORMSINV(RAND()).

*Step 2:* Samples  $\varepsilon_X$  and  $\varepsilon_Y$  are then generated. The first sample of  $X$  variables is the same as the random sample from a univariate standardized normal distribution,  $\varepsilon_X = Z_X$ .

*Step 3:* The conditional sample of  $Y$  variables is determined as follows:

$$\varepsilon_Y = \rho_{XY} Z_X + Z_Y \sqrt{1 - \rho_{XY}^2}$$

where:

$\rho_{XY}$  = correlation between variables  $X$  and  $Y$  in the bivariate normal distribution

## FACTOR MODELS

---

### LO 29.5: Describe properties of correlations between normally distributed variables when using a one-factor model.

---

A factor model can be used to define correlations between normally distributed variables. The following equation is a one-factor model where each  $U_i$  has a component dependent on one common factor ( $F$ ) in addition to another component ( $Z_i$ ) that is uncorrelated with other variables.

$$U_i = \alpha_i F + \sqrt{1 - \alpha_i^2} Z_i$$

Between normally distributed variables, one-factor models are structured as follows:

- Every  $U_i$  has a standard normal distribution (mean = 0, standard deviation = 1).
- The constant  $\alpha_i$  is between -1 and 1.
- $F$  and  $Z_i$  have standard normal distributions and are uncorrelated with each other.
- Every  $Z_i$  is uncorrelated with each other.
- All correlations between  $U_i$  and  $U_j$  result from their dependence on a common factor,  $F$ .

There are two major advantages of the structure of one-factor models. First, the covariance matrix for a one-factor model is positive-semidefinite. Second, the number of correlations between variables is greatly reduced. Without assuming a one-factor model, the correlations of each variable must be computed. If there are  $N$  variables, this would require  $[N \times (N - 1)] / 2$  calculations. However, the one-factor model only requires  $N$  estimates for correlations, where each of the  $N$  variables is correlated with one factor,  $F$ . The most well-known one factor model in finance is the *capital asset pricing model* (CAPM). Under the CAPM, each asset return has a systematic component (measured by beta) that is correlated with the market portfolio return. Each asset return also has a nonsystematic (or idiosyncratic) component that is independent of the return on other stocks and the market.

## COPULAS

---

### LO 29.6: Define copula and describe the key properties of copulas and copula correlation.

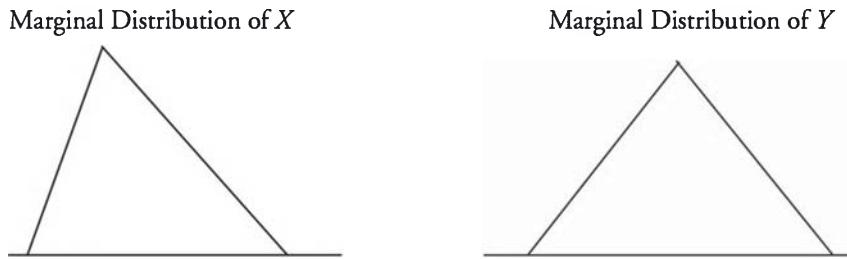
---

Suppose we have two **marginal distributions** of expected values for variables  $X$  and  $Y$ . The marginal distribution of variable  $X$  is its distribution with no knowledge of variable  $Y$ . The marginal distribution of variable  $Y$  is its distribution with no knowledge of variable  $X$ . If both distributions are normal, then we can assume the joint distribution of the variables is bivariate normal. However, if the marginal distributions are not normal, then a copula is necessary to define the correlation between these two variables.

A **copula** creates a joint probability distribution between two or more variables while maintaining their individual marginal distributions. This is accomplished by mapping the marginal distributions to a new known distribution. For example, a Gaussian copula (discussed in LO 29.8) maps the marginal distribution of each variable to the standard normal distribution, which, by definition, has a mean of zero and a standard deviation of one. The mapping of each variable to the new distribution is done based on percentiles.

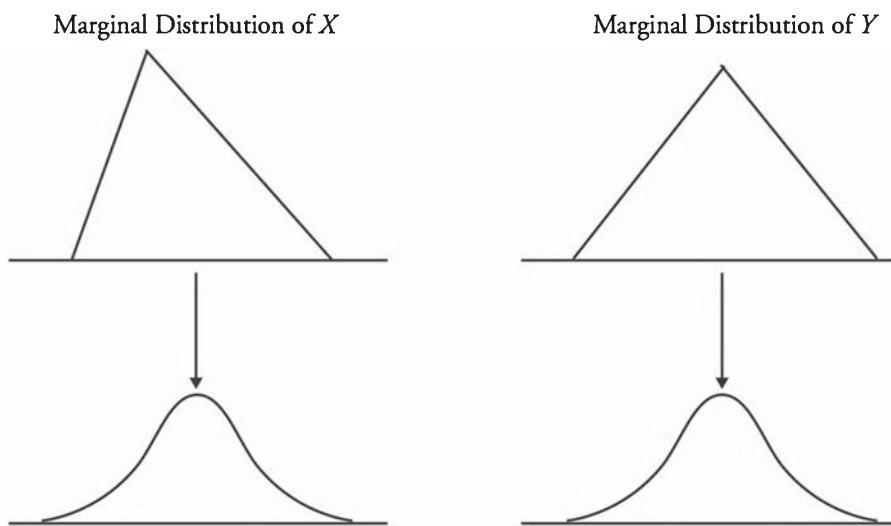
Suppose we have two triangular marginal distributions for two variables  $X$  and  $Y$  as illustrated in Figure 3.

**Figure 3: Marginal Distributions**



These two triangular marginal distributions for  $X$  and  $Y$  are preserved by mapping them to a known joint distribution. Figure 4 illustrates how a copula correlation is created.

**Figure 4: Mapping Variables to Standard Normal Distributions**



The key property of a copula correlation model is the *preservation of the original marginal distributions while defining a correlation between them*. A correlation copula is created by converting two distributions that may be unusual or have unique shapes and mapping them to known distributions with well-defined properties, such as the normal distribution. As mentioned, this is done by mapping on a percentile-to-percentile basis.

For example, the 5<sup>th</sup> percentile observation for the variable  $X$  marginal distribution is mapped to the 5<sup>th</sup> percentile point on the  $U_X$  standard normal distribution. The 5<sup>th</sup> percentile will have a value of  $-1.645$ . This is repeated for each observation on a percentile-to-percentile basis. The value that represents the 95<sup>th</sup> percentile of the  $X$  marginal distribution will have a value mapped to the 95<sup>th</sup> percentile of the  $U_X$  standard normal distribution and will have a value of  $+1.645$ . Likewise, every observation on the variable  $Y$  distribution is mapped to the corresponding percentile on the  $U_Y$  standard normal distribution. The new distribution is now a multivariate normal distribution.

Both  $U_X$  and  $U_Y$  are now normal distributions. If we make the assumption that the two distributions are joint bivariate normal distributions, then a correlation structure can be defined between the two variables. The triangular structures are not well-behaved structures. Therefore, it is difficult to define a relationship between the two variables. However, the normal distribution is a well-behaved distribution. Therefore, using a copula is a way to indirectly define a correlation structure between two variables when it is not possible to directly define correlation.

As mentioned, the correlation between  $U_X$  and  $U_Y$  is referred to as the copula correlation. The conditional mean of  $U_Y$  is linearly dependent on  $U_X$ , and the conditional standard deviation of  $U_Y$  is constant because the two distributions are bivariate normal.

For example, suppose the correlation between  $U_X$  and  $U_Y$  is 0.5. A partial table of the joint probability distribution between variables  $X$  and  $Y$  when the values of  $X$  and  $Y$  are 0.1, 0.2, and 0.3 is illustrated in Figure 5.

**Figure 5: Partial Cumulative Joint Probability Distribution**

		Variable Y		
		0.1	0.2	0.3
Variable X	0.1	0.006	0.017	0.028
	0.2	0.013	0.043	0.081
		0.017	0.061	0.124

Now assume that the variable  $X$  under the original distribution had a value of 0.1 at the 5<sup>th</sup> percentile with a corresponding  $U_X$  value of -1.645. Also assume that the variable  $Y$  under the original distribution had a value of 0.1 with a corresponding value of -2.05. The joint probability that  $U_X < -1.645$  and  $U_Y < -2.05$  can be determined as 0.006 based on the row and column in Figure 5 that corresponds to a 0.1 value for both variables  $X$  and  $Y$ .

## TYPES OF COPULAS

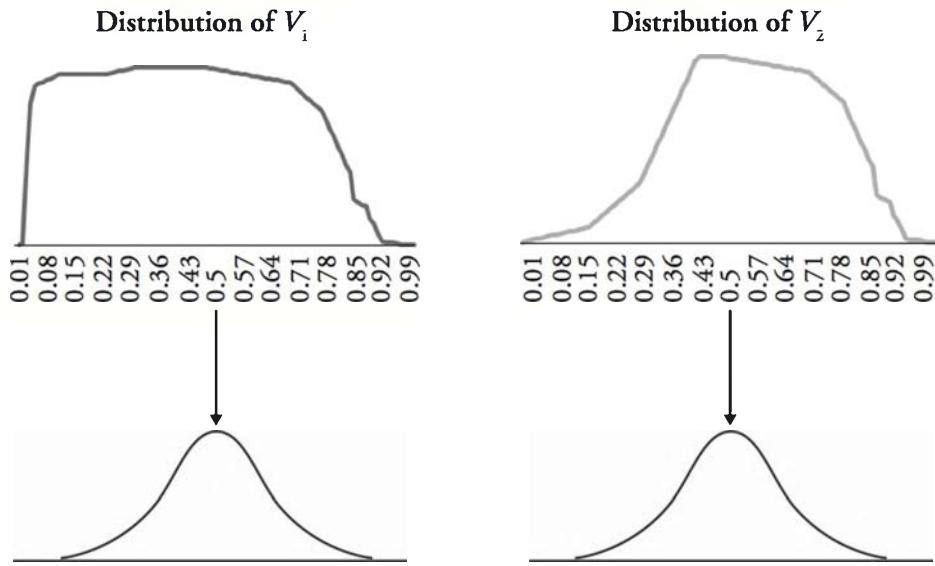
---

### LO 29.8: Describe the Gaussian copula, Student's t-copula, multivariate copula, and one-factor copula.

---

A Gaussian copula maps the marginal distribution of each variable to the standard normal distribution. The mapping of each variable to the new distribution is done based on percentiles. Figure 6 illustrates that  $V_1$  and  $V_2$  have unique marginal distributions. The observations of each distribution is mapped to the standard normal distribution on a percentile-to-percentile basis to create a Gaussian copula as follows.

Figure 6: Mapping Gaussian Copula to Standard Normal Distribution



Other types of copulas are created by mapping to other well-known distributions. The Student's  $t$ -copula is similar to the Gaussian copula. However, variables are mapped to distributions of  $U_1$  and  $U_2$  that have a bivariate Student's  $t$ -distribution rather than a normal distribution.

The following procedure is used to create a Student's  $t$ -copula assuming a bivariate Student's  $t$ -distribution with  $f$  degrees of freedom and correlation  $\rho$ .

*Step 1:* Obtain values of  $\chi$  by sampling from the inverse chi-squared distribution with  $f$  degrees of freedom.

*Step 2:* Obtain values by sampling from a bivariate normal distribution with correlation  $\rho$ .

*Step 3:* Multiply  $\sqrt{f/\chi}$  by the normally distributed samples.

A multivariate copula is used to define a correlation structure for more than two variables. Suppose the marginal distributions are known for  $N$  variables:  $V_1, V_2, \dots, V_N$ . Distribution  $V_i$  for each  $i$  variable is mapped to a standard normal distribution,  $U_i$ . Thus, the correlation structure for all variables is now based on a multivariate normal distribution.

Factor copula models are often used to define the correlation structure in multivariate copula models. The nature of the dependence between the variables is impacted by the choice of the  $U_i$  distribution. The following equation defines a **one-factor copula model** where  $F$  and  $Z_i$  are standard normal distributions:

$$U_i = \alpha_i F + \sqrt{1 - \alpha_i^2} Z_i$$

The  $U_i$  distribution has a multivariate Student's  $t$ -distribution if  $Z_i$  and  $F$  are assumed to have a normal distribution and a Student's  $t$ -distribution, respectively. The choice of  $U_i$  determines the dependency of the  $U$  variables, which also defines the covariance copula for the  $V$  variables.

A practical example of how a one-factor copula model is used is in calculating the value at risk (VaR) for loan portfolios. A risk manager assumes a one-factor copula model maps the default probability distributions for different loans. The percentiles of the one-factor distribution are then used to determine the number of defaults for a large portfolio.

## TAIL DEPENDENCE

---

### **LO 29.7: Explain tail dependence.**

---

There is greater **tail dependence** in a bivariate Student's  $t$ -distribution than a bivariate normal distribution. In other words, it is more common for two variables to have the same tail values at the same time using the bivariate Student's  $t$ -distribution. During a financial crisis or some other extreme market condition, it is common for assets to be highly correlated and exhibit large losses at the same time. This suggests that the Student's  $t$ -copula is better than a Gaussian copula in describing the correlation structure of assets that historically have extreme outliers in the distribution tails at the same time.

## KEY CONCEPTS

### LO 29.1

Correlation and covariance measure the strength between the linear relationship of two variables as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

A correlation of zero between two variables does not imply that there is no dependence between the two variables.

### LO 29.2

The formula for calculating a new covariance on day  $n$  using an exponentially weighted moving average (EWMA) model is:

$$\text{cov}_n = \lambda \text{cov}_{n-1} + (1 - \lambda) X_{n-1} Y_{n-1}$$

GARCH(1,1) applies a weight of  $\alpha$  to the most recent observation on covariance ( $X_{n-1} Y_{n-1}$ ), a weight of  $\beta$  to the most recent covariance estimate ( $\text{cov}_{n-1}$ ), and a weight of  $\omega$  to the long-term average covariance rate as follows:

$$\text{cov}_n = \omega + \alpha X_{n-1} Y_{n-1} + \beta \text{cov}_{n-1}$$

### LO 29.3

A matrix is positive-semidefinite if it is internally consistent. The following expression defines the necessary condition for an  $N \times N$  variance-covariance matrix,  $\Omega$ , to be internally consistent for all  $N \times 1$  vectors  $\omega$ , where  $\omega^T$  is the transpose of vector  $\omega$ :

$$\omega^T \Omega \omega \geq 0$$

### LO 29.4

Independent samples of two variables  $Z_X$  and  $Z_Y$  can be generated from a univariate standardized normal distribution. The conditional sample of  $Y$  variables for a bivariate normal distribution is then generated as:

$$\varepsilon_Y = \rho_{XY} Z_X + Z_Y \sqrt{1 - \rho_{XY}^2}$$

**LO 29.5**

The covariance matrix for a one-factor model is positive-semidefinite. Also, the one-factor model only requires  $N$  estimates for correlations, where each of the  $N$  variables is correlated with one factor,  $F$ .

---

**LO 29.6**

A copula creates a joint probability distribution between two or more variables while maintaining their individual marginal distributions.

---

**LO 29.7**

The Student's  $t$ -copula is better than a Gaussian copula in describing the correlation structure of assets that historically have extreme outliers in tails at the same time.

---

**LO 29.8**

A Gaussian copula maps the marginal distribution of each variable to the standard normal distribution. The Student's  $t$ -copula maps variables to distributions of  $U_1$  and  $U_2$  that have a bivariate Student's  $t$ -distribution. The multivariate copula defines a correlation structure for three or more variables. The choice of  $U_i$  determines the dependency of the  $U$  variables in a one-factor copula model, which also defines the covariance copula for the  $V$  variables.

## CONCEPT CHECKERS

1. Suppose an analyst uses the EWMA model with  $\lambda = 0.95$  to update correlation and covariance rates. The observed percentage change on day  $n - 1$  for variables  $X$  and  $Y$  are 2.0% and 1.0%, respectively. The correlation estimate based on historical data for two variables  $X$  and  $Y$  on day  $n - 1$  is 0.52. In addition, the estimated standard deviations on day  $n - 1$  for variables  $X$  and  $Y$  are 1.4% and 1.8%, respectively. What is the new estimate of the correlation between  $X$  and  $Y$  on day  $n$ ?
  - A. 0.14.
  - B. 0.42.
  - C. 0.53.
  - D. 0.68.
2. An equity analyst is concerned about satisfying the consistency condition for estimating new covariance rates. Which of the following procedures will most likely result in a positive-semidefinite matrix?
  - A. The analyst uses an EWMA model with  $\lambda = 0.95$  to update variances and a GARCH(1,1) model with  $\lambda = 0.95$  to update the covariance rates for a  $1,000 \times 1,000$  variance-covariance matrix.
  - B. The analyst uses an EWMA model with  $\lambda = 0.90$  to update variances and an EWMA model with  $\lambda = 0.90$  to update the covariance rates for a  $3 \times 3$  variance-covariance matrix.
  - C. The analyst uses a GARCH(1,1) model with  $\lambda = 0.95$  to update variances and a GARCH(1,1) model with  $\lambda = 0.90$  to update the covariance rates for a  $1,000 \times 1,000$  variance-covariance matrix.
  - D. The analyst uses an EWMA model with  $\lambda = 0.90$  to update variances and a GARCH(1,1) model with  $\lambda = 0.90$  to update the covariance rates for a  $3 \times 3$  variance-covariance matrix.
3. Suppose two samples,  $Z_X$  and  $Z_Y$ , are generated from a bivariate normal distribution. If variable  $Y$  is conditional on variable  $X$ , which of the following statements regarding these two samples is incorrect?
  - A. The expected value of  $Y$  has a nonlinear relationship with all values of  $X$ .
  - B. The mean and standard deviations for sample  $Z_X$  are unconditional.
  - C. The value of variable  $Y$  is normally distributed.
  - D. The conditional sample of  $Y$  variables is determined by:

$$\varepsilon_Y = \rho_{XY} Z_X + Z_Y \sqrt{1 - \rho_{XY}^2} .$$

## Topic 29

### Cross Reference to GARP Assigned Reading – Hull, Chapter 11

4. Which of the following statements is most reflective of a characteristic of one-factor models between multivariate normally distributed variables? The one-factor model is shown as follows:

$$U_i = \alpha_i F + \sqrt{1 - \alpha_i^2} Z_i$$

- A. Each  $U_i$  has a component dependent on one common factor ( $F$ ) in addition to another component ( $Z_i$ ) that is uncorrelated with other variables.
  - B.  $F$  and  $Z_i$  must both have Student's  $t$ -distributions.
  - C. The covariance matrix for a one-factor model is not positive-semidefinite.
  - D. The number of calculations for estimating correlations is equal to  $[N \times (N - 1)] / 2$ .
5. Suppose a risk manager wishes to create a correlation copula to estimate the risk of loan defaults during a financial crisis. Which type of copula will most accurately measure tail risk?
- A. Gaussian copula.
  - B. Student's  $t$ -copula.
  - C. Gaussian one-factor copula.
  - D. Standard normal copula.

## CONCEPT CHECKER ANSWERS

1. C First, calculate the estimated covariance rate between variables  $X$  and  $Y$  on day  $n - 1$  as:

$$\text{cov}(X, Y) = \rho_{X,Y} \times \sigma_X \sigma_Y = 0.52 \times 0.014 \times 0.018 = 0.00013$$

The EWMA model is then used to update the covariance rate for day  $n$ :

$$\text{cov}_n = 0.95 \times 0.00013 + 0.05 \times 0.02 \times 0.01 = 0.0001235 + 0.00001 = 0.0001335$$

The updated variance of  $X$  is:

$$\sigma_{X,n}^2 = 0.95 \times 0.014^2 + 0.05 \times 0.02^2 = 0.0001862 + 0.00002 = 0.0002062$$

The new volatility measure of  $X$  is then:

$$\sigma_{X,n} = \sqrt{0.0002062} = 0.0143597$$

The updated variance for variable  $Y$  is:

$$\sigma_{Y,n}^2 = 0.95 \times 0.018^2 + 0.05 \times 0.01^2 = 0.0003078 + 0.000005 = 0.0003128$$

The new volatility measure of  $Y$  is then:

$$\sigma_{Y,n} = \sqrt{0.0003128} = 0.01768615$$

The new correlation is found by dividing the new  $\text{cov}_n$  by the new standard deviations for  $X$  and  $Y$  as follows:

$$\frac{0.0001335}{0.0143597 \times 0.0176862} = 0.5257$$

2. B A matrix is positive-semidefinite if it is internally consistent. Variance and covariance rates must be calculated using the same EWMA or GARCH model and parameters to ensure that a positive-semidefinite model is constructed. For example, if an EWMA model is used with  $\lambda = 0.90$  for estimating variances, the same EWMA model and  $\lambda$  should be used to estimate covariance rates.
3. A Both samples are normally distributed. The expected value of variable  $Y$  is normally distributed with a mean of:

$$\mu_Y + \rho_{XY} \times \sigma_Y \times \frac{X - \mu_X}{\sigma_X}$$

and a standard deviation of:

$$\sigma_Y \sqrt{1 - \rho_{XY}^2}$$

The expected value of  $Y$  is therefore linearly dependent on the conditional value of  $X$ .

4. A Each  $U_i$  has a component dependent on one common factor ( $F$ ) in addition to another component ( $Z_i$ ) that is uncorrelated with other variables.  $F$  and  $Z_i$  have standard normal distributions and are uncorrelated with each other. The covariance matrix for a one-factor model is positive-semidefinite and the one-factor model only requires  $N$  estimates for correlations, where each of the  $N$  variables is correlated with one factor,  $F$ .

## Topic 29

### Cross Reference to GARP Assigned Reading – Hull, Chapter 11

5. B There is greater *tail dependence* in a bivariate Student's *t*-distribution than a bivariate normal distribution. This suggests that the Student's *t*-copula is better than a Gaussian copula in describing the correlation structure of assets that historically have extreme outliers in tails at the same time.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

# SIMULATION METHODS

## Topic 30

### EXAM FOCUS

Simulation methods model uncertainty by generating random inputs that are assumed to follow an appropriate probability distribution. This topic discusses the basic steps for conducting a Monte Carlo simulation and compares this simulation method to the bootstrapping technique. For the exam, be able to explain ways to reduce Monte Carlo sampling error, including the use of antithetic and control variates. Also, understand the pseudo-random number generation method and the benefits of reusing sets of random number draws in Monte Carlo experiments. Finally, be able to describe the advantages and disadvantages of the bootstrapping technique in comparison to the traditional Monte Carlo approach.

### MONTE CARLO SIMULATION

#### LO 30.1: Describe the basic steps to conduct a Monte Carlo simulation.

Monte Carlo simulations are often used to model complex problems or to estimate variables when there are small sample sizes. A few practical finance applications of Monte Carlo simulations are: pricing exotic options, estimating the impact to financial markets of changes in macroeconomic variables, and examining capital requirements under stress-test scenarios.

There are four basic steps required to conduct a Monte Carlo simulation.

- Step 1:* Specify the data generating process (DGP)
- Step 2:* Estimate an unknown variable or parameter
- Step 3:* Save the estimate from step 2
- Step 4:* Go back to step 1 and repeat this process  $N$  times

The first step of conducting a simulation requires generating random inputs that are assumed to follow a specific probability distribution. The DGP could be a simple time series model or a more complex full structural model that requires multiple DGPs.

The second step of the simulation generates scenarios or trials based on randomly generated inputs drawn from a pre-specified probability distribution. The most common probability distribution used is the standard normal distribution. However, Student's  $t$  distribution is often used if the user believes it is a better fit for the data. A well-defined simulation model requires the generation of variables that follow appropriate probability distributions.

The last two steps in the simulation process allow for data analysis related to the properties of the probability distributions of the output variables. In other words, rather than making

just one output estimate for a problem, the model generates a probability distribution of estimates. This provides the user with a better understanding of the range of possible outcomes. The quantity  $N$  in step four is the number of times the simulation is repeated. This is referred to as the number of replications or iterations and is typically 1,000 to 10,000 times depending on how costly it is to generate the sample size.

For example, suppose we are managing an investment portfolio and desire to estimate the ending capital in the portfolio in one year,  $C_1$ . The initial capital investment,  $C_0$ , is \$100 invested in the Standard & Poor's 500 index (S&P 500). The return is a random variable that depends on how the market performs over the next year.

If we assume the return over the next year is equal to a historical mean return, we can calculate one point estimate of the ending capital based on the equation:  $C_1 = C_0(1 + r)$ . The return over the next period is a random variable, and a simulation model estimates multiple scenarios to represent future returns based on a probability distribution of possible outcomes. The output variable is an estimate of an ending amount of capital that is also a random variable. The simulation model allows us to visualize the output and analyze the probability distribution of the ending capital amounts generated by the model.

## REDUCING MONTE CARLO SAMPLING ERROR

---

### LO 30.2: Describe ways to reduce Monte Carlo sampling error.

---

The sampling variation for a Monte Carlo simulation is quantified as the standard error estimate. The standard error of the true expected value is computed as  $s / \sqrt{N}$ , where  $s$  is the standard deviation of the output variables and  $N$  is the number of scenarios or replications in the simulation. Based on this equation, it intuitively follows that in order to reduce the standard error estimate by a factor of 10, the analyst must increase  $N$  by a factor of 100. (Because the square root of 100 is 10, if we increase the sample size 100 times it will reduce the standard error estimate by dividing by 10.)

Suppose we continue the illustration from the previous example and run a simulation to estimate the ending capital amount for an initial investment portfolio of \$100. The number of replications is initially 100 (i.e.,  $N = 100$ ), resulting in a mean ending capital of \$110 and a standard deviation of \$14.798. For this example, the standard error estimate is computed as \$1.4798 (i.e., \$14.798 / 10). Now, suppose we want to increase the accuracy by reducing the standard error estimate. How can we increase the accuracy of the simulation?

The accuracy of simulations depends on the standard deviation and the number of scenarios run. We cannot control the standard deviation, but we can control the number of replications. Assume we rerun the previous simulation with 400 replications that result in the same mean ending capital of \$110, and the standard deviation remains at \$14.798. The standard error estimate for the simulation with 400 replications is then \$0.7399 (i.e., \$14.798 / 20). With four times the number of scenarios ( $4 \times N$ , or 400, in this example) the standard error estimate is cut in half to \$0.7399. In other words, quadrupling the number of scenarios will improve the accuracy twofold.

However, increasing the number of generated scenarios can become costly for more complex multi-period simulations. Variance reduction techniques offer an alternative way to reduce the sampling error of a Monte Carlo simulation. The two most commonly used techniques for reducing the standard error estimate are antithetic variates and control variates.

## ANTITHETIC VARIATES

---

### LO 30.3: Explain how to use antithetic variate technique to reduce Monte Carlo sampling error.

---

One reason sampling error occurs is because there are often a wide range of possible outcomes for a particular experiment or problem. Thus, in order to replicate the entire range of possible outcomes the sampling sets must be recreated numerous times. However, increasing the number of samples drawn may be too costly and time consuming. As an alternative approach, the **antithetic variate technique** can reduce Monte Carlo sampling error by rerunning the simulation using a *complement* set of the original set of random variables.

If the original set of random draws is denoted  $u_t$  for each replication, then the simulation is rerun with the complement set of random numbers denoted  $-u_t$ . By definition, the use of antithetic variates results in a lower covariance and variance, because the two sets are perfectly negatively correlated [i.e.,  $\text{corr}(u_t, -u_t) = -1$ ]. The following example illustrates how the standard error for a Monte Carlo simulation is reduced by using the antithetic variate technique.

First, consider a simulation of two sets that does not use the antithetic variate technique. Suppose the average parameter estimate is determined by two Monte Carlo simulations using different random sample sets. The average output parameter value,  $\bar{x}$ , for the two simulations using different random sample replications is simply calculated as:

$$\bar{x} = (x_1 + x_2) / 2$$

Where  $x_1$  and  $x_2$  are the average output parameter values for simulation sets 1 and 2, respectively.

Next, we can calculate the variance of the average of the two sets as follows:

$$\text{var}(\bar{x}) = \frac{\text{var}(x_1) + \text{var}(x_2) + 2 \text{cov}(x_1, x_2)}{4}$$

Without using antithetic variates, the two sets of Monte Carlo replications are independent. Thus, the covariance will be zero and the variance of  $\bar{x}$  is simply reduced to the following:

$$\text{var}(\bar{x}) = \frac{\text{var}(x_1) + \text{var}(x_2)}{4}$$

The use of antithetic variates results in a negative covariance between the original random draws and their complements (i.e., antithetic variates). Thus, the use of antithetic variates

causes the error terms to be independent for the two sets, which results in a negative covariance term in the variance equation. This negative relationship means that the Monte Carlo sampling error must always be smaller using this approach.

## CONTROL VARIATES

---

### LO 30.4: Explain how to use control variates to reduce Monte Carlo sampling error and when it is effective.

---

The **control variate technique** is a widely used method to reduce the sampling error in Monte Carlo simulations. A control variate involves replacing a variable  $x$  (under simulation) that has unknown properties with a similar variable  $y$  that has known properties.

Suppose two separate simulations are conducted on variable  $x$  with unknown properties and control variable  $y$  with known properties using the same set of random numbers. Also assume that the Monte Carlo simulation estimated variables for  $x$  and  $y$  are denoted as  $\hat{x}$  and  $\hat{y}$ , respectively. The original estimate for  $x$  can be redefined as  $x^*$  as follows:

$$x^* = y + (\hat{x} - \hat{y})$$

The new  $x^*$  variable estimate will have a smaller sampling error than the original  $x$  variable if the control statistic and statistic of interest are highly correlated. The Monte Carlo results for the new  $x^*$  variable are assumed to have similar properties to the known  $y$  control variable.

The following mathematical equations help illustrate the condition that is necessary to reduce the sampling error using control variates. Consider taking the variance of both sides of the equation that defines the new variable such that:

$$\text{var}(x^*) = \text{var}[y + (\hat{x} - \hat{y})]$$

The control variable  $y$  does not have a sampling error because it has known properties. Thus, the  $\text{var}(y)$  equals zero. Now, the variance of the remaining two variables can be rewritten as follows:

$$\text{var}(x^*) = \text{var}(\hat{x}) + \text{var}(\hat{y}) - 2 \text{cov}(\hat{x}, \hat{y})$$

The control variate method will only reduce the sampling error in Monte Carlo simulations if  $\text{var}(x^*)$  is less than  $\text{var}(\hat{x})$ . Another way of expressing this condition is as follows:

$$\text{var}(\hat{y}) - 2 \text{cov}(\hat{x}, \hat{y}) < 0$$

This relationship can be simplified as follows:

$$\text{cov}(\hat{x}, \hat{y}) > \frac{\text{var}(\hat{y})}{2}$$

The covariance can be converted to correlation by dividing both sides of the previous inequality by the product of the standard deviations as follows:

$$\text{corr}(\hat{x}, \hat{y}) > \frac{1}{2} \sqrt{\frac{\text{var}(\hat{y})}{\text{var}(\hat{x})}}$$

A practical financial example of applying control variates is the use of Monte Carlo simulations in pricing Asian options (which will be discussed in Book 4). An Asian option is priced based on the average value of the underlying asset over the lifespan of the option. The use of a similar derivative, such as a European option, with known statistical properties can be used as a control variate. The price of the European option,  $P_{BS}$ , is determined by the Black-Scholes-Merton option pricing model. Next, simulated prices are determined for the Asian option and the European option and denoted  $P_A$  and  $P_{BS}^*$ , respectively. The new estimate of the Asian option price,  $P_A^*$ , could then be determined based on the following equation:

$$P_A^* = (P_A - P_{BS}) + P_{BS}^*$$

## REUSING SETS OF RANDOM NUMBERS

---

### LO 30.5: Describe the benefits of reusing sets of random number draws across Monte Carlo experiments and how to reuse them.

---

Reusing sets of random number draws across Monte Carlo experiments reduces the estimate variability across experiments by using the same set of random numbers for each simulation. Normally, a user would not desire to reuse the same random draws. However, in certain situations this technique is useful. Two examples of reusing sets of random numbers are for testing the power of the Dickey-Fuller test (used to determine whether a time series is covariance stationary) or for different experiments with options using time series data.

*Dickey-Fuller (DF) test.* Suppose an analyst wants to examine the DF test for sample sizes of 1,000 to test whether or not a particular market follows a random walk or contains a drift element. The analyst could reuse the same set of standard normal random variables for each simulation run while testing with different DF parameters. Using the same set of random numbers for each Monte Carlo experiment reduces the sampling variation across experiments. In this case, the sampling variability is reduced, but the accuracy of the actual estimates is not increased.

*Different experiments.* Another example where reusing sample data is useful is in testing differences among options. For example, suppose an analyst is examining option prices that are similar in all aspects except for time to maturity. The analyst could simulate a long time series of random draws and then split this longer time series into shorter time frames. A six-month time series of data could be subdivided into three sets of two-month maturity options or six sets of one-month maturity options. Using the same random number data set reduces the variability of simulated option prices across maturities.

## BOOTSTRAPPING METHOD

---

### LO 30.6: Describe the bootstrapping method and its advantage over Monte Carlo simulation.

---

Another way to generate random numbers is the bootstrapping method. The bootstrapping approach draws random return data from a sample of historical data. Under traditional Monte Carlo simulation, data sets are created by selecting random variables drawn from a pre-determined probability distribution. The bootstrapping method uses actual historical data instead of random data from a probability distribution. In addition, bootstrapping repeatedly draws data from a historical data set and replaces the data so it can be drawn again.

For example, suppose an analyst uses the bootstrapping method to estimate parameter  $\theta$ . The analyst begins by obtaining sample historical data over a specific time period. This historical data is denoted:

$$y = y_1, y_2, \dots, y_T$$

The statistical properties of parameter  $\hat{\theta}_T$  are then estimated based on the bootstrapping sample data. The analyst creates  $N$  samples of  $T$  variables with replacement from the original  $y$  data sample. The parameter estimate  $\hat{\theta}$  is calculated for every sample to create  $N$  estimates. In other words, the samples that are drawn are not totally random, but are drawn from a pre-determined historical sample set  $y$ . The statistical properties of this sample of  $\hat{\theta}$  estimates are then analyzed.

An obvious advantage of the bootstrapping approach is that no assumptions are made regarding the true distribution of the parameter estimate that is being examined. This implies that it can include extreme events that have occurred in the past (e.g., during a financial crisis). Inclusion of outliers will produce a distribution that has fatter tails than the normal distribution, which allows for a more realistic view of actual return data. Thus, the bootstrapping methodology generates a collection of data sets with approximately the same distribution properties as the original data. However, any dependency of variables or autocorrelations in the original data set will no longer be present, because variables are not drawn in the same sequence as the original data set.

The following example describes how bootstrapping is used with a regression model. Assume that the bootstrapping approach is used to re-sample data with respect to the following standard regression model:

$$y = u + X\beta$$

The first step of the bootstrapping approach is to generate a sample size  $T$  of the historical data by drawing samples with replacement that take all related data corresponding to each observation  $y_i$ . In other words, for the 21st data observation,  $y_{21}$ , the approach takes this estimate along with all values of the explanatory variables for the 21st observation.

Next the coefficient matrix,  $\hat{\beta}^*$ , is estimated for this bootstrap sample. This process is then repeated a total of  $N$  times. Every time data is resampled, a sample size of  $T$  is generated from the original sample data with replacement and a coefficient matrix is estimated. This results in a set of  $N$  coefficient vectors that will all be unique, and a distribution of estimates is created for each coefficient.

This bootstrapping approach has a methodological problem resulting from sampling from regressors rather than using a fixed estimate in repeated samples. To correct for this problem, the approach can be slightly modified where re-sampling occurs with the residuals. Thus, the first step would be to sample actual data, estimate the value  $\hat{y}$  and calculate the residuals,  $\hat{u}$ . The coefficient vector is then created using a modified dependent variable that is the sum of the fitted values and the bootstrap residuals  $\hat{u}^*$  as follows:

$$\mathbf{y}^* = \hat{\mathbf{y}} + \hat{\mathbf{u}}^*$$

#### **LO 30.8: Describe situations where the bootstrapping method is ineffective.**

Two situations that cause the bootstrapping method to be ineffective are *outliers* in the data and *non-independent data*.

If outliers exist in the data, the inferences drawn from parameter estimates may not be accurate depending on how many times the outliers are included in the bootstrapped sample. Because replacement is used in the bootstrap method, outliers could be drawn more often, causing the bootstrap distribution to have fatter tails. Alternatively, not drawing the outlier in the bootstrapped sample may lead to the opposite conclusions regarding the parameter estimate statistical properties. Recall that a major advantage of the bootstrapping approach over traditional approaches is that it does not require any assumptions of the probability distribution of the sampled data. Thus, the best way to mitigate this issue is to have a large number of replications.

If autocorrelation exists in the original sample data, then the original historical data are not independent of one another. A technique known as a *moving block bootstrap* is used to overcome the problem of autocorrelation. Blocks of data are examined at one time in order to preserve the original data dependency.

#### **RANDOM NUMBER GENERATION**

#### **LO 30.7: Describe the pseudo-random number generation method and how a good simulation design alleviates the effects the choice of the seed has on the properties of the generated series.**

A good random number generator has the ability to reproduce a random sequence and analyze characteristics of random numbers. Simulation software programs are able to reproduce the same sequence of iterations by starting sequences with a seed random number. The algorithms used to generate these random sequences are referred to as **pseudo-random number generators**. These number generators are advantageous because risk managers can improve models by reducing the estimate variance or debugging computer

codes if the same sequence of random numbers is reproduced when programming the model.

A very common pseudo-random number generator is one that generates random number sequences uniformly distributed between 0 and 1. Each number has an equal probability of being drawn from this uniform (0,1) distribution. Numbers can be drawn from a discrete or continuous distribution. The term *pseudo* implies that these computer-generated numbers are *not truly random*, because they are actually generated from a formula. For example, suppose random numbers are generated from a continuous uniform (0,1) distribution based on the following formula:

$$y_{i+1} = (ay_i + c) \text{ modulo } m, i = 0, 1, 2, \dots, T$$

In this formula,  $T$  is the total number of random numbers drawn,  $y_0$  is the initial value of  $y$ , which is referred to as the *seed*,  $a$  is a constant multiplier, and  $c$  is an incremental value. The statement “modulo  $m$ ” in this formula refers to modulo operator, which is a clocklike process where the generator returns to 1 when the value  $m$  is reached.

In order to run a simulation, the user must first define the initial seed value,  $y_0$ . The choice of seed value will influence the properties of the random number distribution that is generated. The effect is strongest for the early draws in a series, but eventually the impact fades away. Therefore, the best way to control for this problem is to generate a very large number of observations and then discard the earliest observations.

For example, if a user requires 800 observations, then 1,000 random numbers are generated and the first 200 are eliminated from the sample. This ensures that the statistical properties of the sample reflect those of true random numbers that are not based on a pre-specified formula. Eventually random number sequences will repeat. Therefore, a good random number generator uses sequences with long cycles that require numerous iterations before a sequence is repeated.

## DISADVANTAGES OF SIMULATION APPROACHES

---

### LO 30.9: Describe disadvantages of the simulation approach to financial problem solving.

---

Disadvantages of the simulation approach to financial problem solving include:

- High computation costs
- Results are imprecise
- Results are difficult to replicate
- Results are experiment-specific

Some problems may require a large number of replications to obtain more accurate results. If estimated parameters are complex, the computations may take an extremely long time to run. Computer processor times have improved exponentially. However, the complexity of markets and issues that are examined have also become increasingly complex, leading to *high computation costs*.

*Imprecise results* may be present even with a very large number of simulation iterations when the assumptions of model inputs or the data generating process are unrealistic. A common mis-specified model assumption is related to the underlying probability distribution of inputs. For example, option prices are typically fat-tailed, but a model could erroneously draw option prices from a normal distribution. This would lead to inaccurate results regardless of the number of replications.

In practice, users seldom use a defined seed for the start of random draws in simulations. Without the use of an initial seed, it is *not possible to replicate results* from previous experiments. The best way to overcome this problem and reduce the variation of results is to use a very large number of replications. Thus, it is common to use at least 10,000 replications in Monte Carlo simulations if it is computationally cost-effective.

*Simulation results are experiment-specific* because financial problems are analyzed based on a specific data generating process and set of equations. If alternate assumptions are made in the equations or data generating process, the results may differ substantially.

## KEY CONCEPTS

---

### LO 30.1

The basic steps of a Monte Carlo simulation are: (1) specify the data generating process (DGP), (2) estimate an unknown variable, (3) save the estimate from step 2, and (4) go back to step 1 and repeat this process  $N$  times.

---

### LO 30.2

The standard error estimate of a Monte Carlo simulation,  $s / \sqrt{N}$ , can be reduced by a factor of 10 by increasing  $N$  by a factor of 100.

---

### LO 30.3

The antithetic variate technique reduces Monte Carlo sampling error by rerunning the simulation using a complement set of the original set of random variables.

---

### LO 30.4

The control variate technique replaces a variable  $x$  that has unknown properties in a Monte Carlo simulation with a similar variable  $y$  that has known properties. The new  $x^*$  variable estimate will have a smaller sampling error than the original  $x$  variable if the control statistic and statistic of interest are highly correlated.

---

### LO 30.5

Reusing sets of random number draws across Monte Carlo experiments reduces the estimate variability across experiments.

---

### LO 30.6

Bootstrapping simulations repeatedly draw data from historical data sets and replace the data so it can be re-drawn. The bootstrapping technique requires no assumptions with respect to the true distribution of the parameter estimates.

---

### LO 30.7

Pseudo-random numbers are not truly random, because they are actually generated from a formula. The choice of the initial seed value influences the properties of the random number distribution that is generated. Thus, when using a seed value, increasing the number of replications and eliminating early estimates from the sample can mitigate any biases.

---

### LO 30.8

The bootstrapping method is ineffective when there are outliers in the data or when the data is non-independent.

**LO 30.9**

Disadvantages of the simulation approach to financial problem solving include: high computation costs, imprecise results, difficulty with replicating results, and experiment-specific results.

## CONCEPT CHECKERS

1. Suppose an analyst is concerned about Monte Carlo sampling error. Based on an initial Monte Carlo simulation with 100 replications, the results indicated a standard deviation of 12.64. The simulation was rerun with 900 replications and the standard deviation remained at 12.64. What are the standard error estimates for the simulations with 100 replications and 900 replications, respectively?
- |                |                |
|----------------|----------------|
| <u>N = 100</u> | <u>N = 900</u> |
| A. 0.126       | 0.014          |
| B. 0.126       | 0.140          |
| C. 1.264       | 0.421          |
| D. 1.264       | 0.214          |
2. A concern for Monte Carlo simulations is the size of the sampling error. One way to reduce the sampling error is to use the antithetic variate technique. Which of the following statements best describe this technique?
- A. The simulation is rerun using a complement set of the original set of random variables.
  - B. The number of replications is increased significantly to reduce sampling error.
  - C. Sample data is replaced after every replication to ensure it has an equal probability of being redrawn.
  - D. The data generating process is approximated by redefining the unknown variable with a variable that has known properties.
3. Suppose an analyst is testing the robustness of the Dickey-Fuller test by changing the drift parameter for several different experiments. Reusing sets of random number draws across Monte Carlo experiments will most likely result in:
- A. increasing the accuracy of the drift estimates for each experiment.
  - B. increasing the sampling variance across experiments.
  - C. reducing the accuracy of the drift estimates for each experiment.
  - D. reducing the sampling variance across experiments.
4. Suppose a pseudo-random number generator is used that generates random number sequences uniformly and continuously distributed between 0 and 1. An analyst begins by defining the initial seed value for the number generator process. The analyst knows that the choice of seed value will influence the properties of the generated random number distribution. The best way to reduce this problem is by using a:
- A. large number of replications and discarding the outliers.
  - B. large number of replications and discarding the earliest draws.
  - C. small seed or initial value.
  - D. large seed or initial value.
5. Monte Carlo simulation is a widely used technique in solving economic and financial problems. Which of the following statements is not a limitation of the Monte Carlo technique when solving problems of this nature?
- A. High computational costs arise with complex problems.
  - B. Simulation results are experiment-specific because financial problems are analyzed based on a specific data generating process and set of equations.
  - C. Results of most Monte Carlo experiments are difficult to replicate.
  - D. If the input variables have fat tails, Monte Carlo simulations are not relevant because it always draws random variables from a normally distributed population.

## CONCEPT CHECKER ANSWERS

1. C The standard error is determined by dividing the standard deviation by the square root of the number of replications  $s / \sqrt{N}$ . The standard error estimate for the first simulation of 100 replications is 1.264 (i.e.,  $12.64 / 10$ ). With 900 replications, the standard error estimate is reduced to 0.4213 (i.e.,  $12.64 / 30$ ).
2. A The antithetic variate technique reduces Monte Carlo sampling error by rerunning the simulation using a complement set of the original set of random variables.
3. D Using the same set of random numbers for each Monte Carlo experiment reduces the sampling variation across experiments. Although the sampling variability is reduced, the accuracy of the actual estimates in each case is not influenced.
4. B The best way to control for this problem is to generate a very large number of observations and then discard the earliest observations. This ensures that the statistical properties of the sample reflect those of true random numbers that are not based on a pre-specified formula.
5. D A disadvantage of Monte Carlo simulation is that imprecise results may be present when the assumptions of model inputs or data generating process are unrealistic. The distribution of input variables does not need to be the normal distribution. The problem arises when a variable in the real world is fat-tailed, but a model could erroneously draw option prices from a normal distribution.

---

# SELF-TEST: QUANTITATIVE ANALYSIS

---

## 10 Questions: 24 Minutes

1. Given the following probability data for the return on the market and the return on Best Oil, calculate the covariance of returns between Best Oil and the market.

**Probability Matrix**

	$R_{Best} = 20\%$	$R_{Best} = 10\%$	$R_{Best} = 5\%$
$R_{Mkt} = 15\%$	40%	0	0
$R_{Mkt} = 10\%$	0	20%	0
$R_{Mkt} = 0\%$	0	0	40%

- A. 44.0.  
B. 12.0.  
C. 2.8.  
D. 22.5.
2. Rob Conniff has encountered a difficult section on a multiple-choice exam. There are five questions in this section and each question has three equally likely answer choices. Which of the following amounts is closest to the probability that he will get three or more questions correct by randomly guessing?  
A. 4.5%.  
B. 16.5%.  
C. 21.0%.  
D. 79.0%.
3. You are forecasting the sales of a building materials supplier by assessing the expansion plans of its largest customer, a homebuilder. You estimate the probability that the customer will increase its orders for building materials to 25%. If the customer does increase its orders, you estimate the probability that the homebuilder will start a new development at 70%. If the customer does not increase its orders from this supplier, you estimate only a 20% chance that it will start the new development. Later, you find out that the homebuilder will start the new development. In light of this new information, what is your new (updated) probability that the builder will increase its orders from this supplier?  
A. 17.50%.  
B. 32.55%.  
C. 53.85%.  
D. 60.00%.

4. In performing hypothesis testing as a quantitative analyst, you have recently encountered some unsatisfactory results. You consult your boss and he suggests that you consider increasing the significance level in your testing activities. Which of the following outcomes would most likely occur with such an increase?
- Increased probability of making a Type I error.
  - Increased probability of making a Type I or II error.
  - Decreased probability of making a Type I error.
  - Decreased probability of making a Type I or II error.

**Use the following information to answer Question 5.**

An analyst is given the data in the following table for a regression of the annual sales for Company XYZ, a maker of paper products, on paper product industry sales.

Parameters	Coefficient	Standard Error of the Coefficient
Intercept	-94.88	32.97
Slope (industry sales)	0.2796	0.0363

The correlation between company and industry sales is 0.9757. The regression was based on five observations.

5. Which of the following is closest to the value and reports the most likely interpretation of the  $R^2$  for this regression? The  $R^2$  is:
- 0.048, indicating that the variability of industry sales explains about 4.8% of the variability of company sales.
  - 0.048, indicating that the variability of company sales explains about 4.8% of the variability of industry sales.
  - 0.952, indicating that the variability of industry sales explains about 95.2% of the variability of company sales.
  - 0.952, indicating that the variability of company sales explains about 95.2% of the variability of industry sales.

**Use the following information to answer Questions 6 through 8.**

Theresa Miller is attempting to forecast sales for Alton Industries based on a multiple regression model. The model Miller estimates is:

$$\text{sales} = b_0 + (b_1 \times \text{DOL}) + (b_2 \times \text{IP}) + (b_3 \times \text{GDP}) + \varepsilon_t$$

where:

sales = change in sales adjusted for inflation

DOL = change in the real value of the \$ (rates measured in €/\$)

IP = change in industrial production adjusted for inflation (millions of \$)

GDP = change in inflation-adjusted GDP (millions of \$)

All changes in variables are in percentage terms.

Miller runs the regression using monthly data for the prior 180 months. The model estimates (with coefficient standard errors in parentheses) are:

$$\text{sales} = 10.2 + (5.6 \times \text{DOL}) + (6.3 \times \text{IP}) + (9.2 \times \text{GDP})$$

(5.4)	(3.5)	(4.2)	(5.3)
-------	-------	-------	-------

The sum of squared residuals (SSR) is 145.6 and the total sum of squares (TSS) is 357.2.

**Figure 1: Partial Student's *t*-distribution (one-tailed probabilities)**

df	p = 0.10	p = 0.05	p = 0.025	p = 0.01	p = 0.005
170	1.287	1.654	1.974	2.348	2.605
176	1.286	1.654	1.974	2.348	2.604
180	1.286	1.653	1.973	2.347	2.603

**Figure 2: Partial *F*-Table critical values for right-hand tail area equal to 0.05**

	df 1 = 1	df 1 = 3	df 1 = 5
df 2 = 170	3.90	2.66	2.27
df 2 = 176	3.89	2.66	2.27
df 2 = 180	3.89	2.65	2.26

**Figure 3: Partial *F*-Table critical values for right-hand tail area equal to 0.025**

	df 1 = 1	df 1 = 3	df 1 = 5
df 2 = 170	5.11	3.19	2.64
df 2 = 176	5.11	3.19	2.64
df 2 = 180	5.11	3.19	2.64

6. The unadjusted  $R^2$  and the standard error of the regression (SER) are closest to:
- |                      |            |
|----------------------|------------|
| <u>R<sup>2</sup></u> | <u>SER</u> |
|----------------------|------------|
- |          |       |
|----------|-------|
| A. 59.2% | 1.425 |
| B. 59.2% | 0.910 |
| C. 40.8% | 0.910 |
| D. 40.8% | 1.425 |

7. The appropriate decision with regard to the  $F$ -statistic for testing the null hypothesis that all of the independent variables are simultaneously equal to zero at the 5% significance level is to:
- reject the null hypothesis because the  $F$ -statistic is larger than the critical  $F$ -value of 3.19.
  - fail to reject the null hypothesis because the  $F$ -statistic is smaller than the critical  $F$ -value of 3.19.
  - reject the null hypothesis because the  $F$ -statistic is larger than the critical  $F$ -value of 2.66.
  - fail to reject the null hypothesis because the  $F$ -statistic is smaller than the critical  $F$ -value of 2.66.
8. What is the width of the 99% confidence interval for GDP, and is zero in that 99% confidence interval?
- | <u>Width of 99% CI</u> | <u>Zero in interval</u> |
|------------------------|-------------------------|
| A. 13.8                | Yes                     |
| B. 3.8                 | No                      |
| C. 27.6                | Yes                     |
| D. 27.6                | No                      |
9. The GTEC Corporation uses an exponentially weighted moving average (EWMA) model with a decay factor of 0.75 to model the daily volatility of a stock. The current estimate of daily volatility 1.8%. The closing price of the stock was \$38 yesterday and \$35 today. Using continuously compounded returns, what is the updated estimate of volatility?
- 5.39%.
  - 4.39%.
  - 3.39%.
  - 2.39%.
10. A risk manager estimates the daily variance using a GARCH(1,1) model on daily returns ( $r_t$ ):

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1}$$

The model parameter values are:

$$\alpha_0 = 0.0000008$$

$$\alpha_1 = 0.050$$

$$\beta = 0.93$$

Using the model, what is the long-run annualized volatility estimate (assuming 252 trading days in a year and that volatility increases by the square root of time)?

- 0.52%.
- 0.63%.
- 9.89%.
- 10.04%.

---

# SELF-TEST ANSWERS: QUANTITATIVE ANALYSIS

---

1. A  $E(R_{\text{Best}}) = 0.4(20\%) + 0.2(10\%) + 0.4(5\%) = 12\%$

$$E(R_{\text{Mkt}}) = 0.4(15\%) + 0.2(10\%) + 0.4(0\%) = 8\%$$

$$\text{Cov}(R_{\text{Best}}, R_{\text{Mkt}}) = 0.4(20\% - 12\%)(15\% - 8\%)$$

$$+ 0.2(10\% - 12\%)(10\% - 8\%)$$

$$+ 0.4(5\% - 12\%)(0\% - 8\%)$$

$$= 0.4(8)(7) + 0.2(-2)(2) + 0.4(-7)(-8) = 44$$

The units of covariance (like variance) are percent squared here. We used whole number percents in the calculations and got 44; if we had used decimals, we would have gotten 0.0044.

(See Topic 16)

- 2 C The number of questions correct would follow a binomial distribution. Probability of success is 1/3 and the number of trials is 5. The probability of getting three or more questions correct is the sum of the following:

$$P(3) = 10 \times (1/3)^3 \times (2/3)^2 = 0.1646$$

$$P(4) = 5 \times (1/3)^4 \times (2/3)^1 = 0.0412$$

$$P(5) = 1 \times (1/3)^5 \times (2/3)^0 = 0.0041$$

$$0.1646 + 0.0412 + 0.0041 = 21.0\%$$

(See Topic 17)

3. C The prior probability that the builder will increase its orders is 25%.

$$P(\text{increase}) = 0.25$$

$$P(\text{no increase}) = 0.75$$

There are four possible outcomes:

- Builder increases its orders and starts new development.
- Builder increases its orders and does not start new development.
- Builder does not increase its orders and starts new development.
- Builder does not increase its orders and does not start new development.

The probabilities of each outcome are as follows:

- $P(\text{increase and development}) = (0.25)(0.70) = 0.175$ .
- $P(\text{increase and no development}) = (0.25)(0.30) = 0.075$ .
- $P(\text{no increase and development}) = (0.75)(0.20) = 0.15$ .
- $P(\text{no increase and no development}) = (0.75)(0.80) = 0.60$ .

We want to update the probability of an increase in orders, given the new information that the builder is starting the development. We can apply Bayes' formula:

$$P(\text{increase} \mid \text{development}) = \frac{P(\text{development} \mid \text{increase}) \times P(\text{increase})}{P(\text{development})}$$

From our assumptions,  $P(\text{development} \mid \text{increase}) = 0.70$ , and  $P(\text{increase}) = 0.25$ , so the numerator is  $(0.70)(0.25) = 0.175$ .

$P(\text{development})$  is the sum of  $P(\text{development and increase})$  and  $P(\text{development and no increase})$ .

$$P(\text{development}) = 0.175 + 0.15 = 0.325$$

$$\text{Thus, } P(\text{increase} \mid \text{development}) = \frac{(0.7) \times (0.25)}{0.175 + 0.15} = \frac{0.175}{0.325} = 0.5385, \text{ or } 53.85\%$$

(See Topic 18)

4. A An increase in the significance level (from 1% to 5%, for example) means that a researcher is more likely to reject the null hypothesis since the critical value will be lower. Therefore, there is a greater probability of making a Type I error (rejecting the null hypothesis when it is actually true).

(See Topic 19)

5. C The  $R^2$  is computed as the correlation squared:  $(0.9757)^2 = 0.952$ .

The interpretation of this  $R^2$  is that 95.2% of the variation in Company XYZ's sales is explained by the variation in industry sales. Answer D is incorrect because it is the independent variable (industry sales) that explains the variation in the dependent variable (company sales). This interpretation is based on the economic reasoning used in constructing the regression model.

(See Topic 20)

$$6. B \quad \text{SER} = \sqrt{\frac{145.6}{180 - 3 - 1}} = 0.910$$

$$\text{unadjusted } R^2 = \frac{357.2 - 145.6}{357.2} = 0.592$$

(See Topic 22)

7. C  $\text{ESS} = 357.2 - 145.6 = 211.6$ ,  $F\text{-statistic} = (211.6 / 3) / (145.6 / 176) = 85.3$ . The critical value for a one-tailed 5%  $F$ -test with 3 and 176 degrees of freedom is 2.66. Because the  $F$ -statistic is greater than the critical  $F$ -value, the null hypothesis that all of the independent variables are simultaneously equal to zero should be rejected.

(See Topic 23)

8. C The confidence interval is  $9.2 \pm (5.3 \times 2.604)$ , where 2.604 is the two-tailed 1%  $t$ -statistic with 176 degrees of freedom (which is the same as a one-tailed 0.5%  $t$ -statistic with 176 degrees of freedom). The interval is -4.6 to 23.0, which has a width of 27.6 and zero is in that interval.

(See Topic 23)

**Book 2****Self-Test Answers: Quantitative Analysis**

9. B Updated volatility estimate =  $[\lambda \times (\text{volatility}_{t-1})^2 + (1 - \lambda) \times (\text{current return})^2]^{0.5}$

Current return =  $\ln(\text{price today} / \text{price yesterday})$

$\ln(35/38) = -8.223\%$

Updated volatility estimate =  $[0.75 \times (0.018)^2 + 0.25 \times (-0.08223)^2]^{0.5}$

=  $[0.000243 + 0.001690443]^{0.5}$

= 4.39%

(See Topic 28)

10. D Remember that when questions ask for volatility, they are referring to the standard deviation.

We first calculate the daily variance, which then needs to be adjusted to an annualized variance and finally we can take the square root to find the annualized volatility (standard deviation).

Long-run daily variance =  $\alpha_0 / (1 - \alpha_1 - \beta)$

=  $0.0000008 / (1 - 0.05 - 0.93) = 0.00004$

Long-run daily standard deviation =  $\sqrt{\text{variance}} = \sqrt{0.00004} = 0.6325\%$

Annualized standard deviation = daily standard deviation  $\times \sqrt{\text{time}}$

=  $0.6325\% \times \sqrt{252} = 10.04\%$

(See Topic 28)

---

# FORMULAS

---

## Quantitative Analysis

---

### Topic 15

joint probability:  $P(AB) = P(A | B) \times P(B)$

conditional probability:  $P(A | B) = \frac{P(AB)}{P(B)}$

independent events:  $P(A | B) = P(A)$

---

### Topic 16

expected value:  $E(X) = \sum P(x_i)x_i$

variance:  $\text{Var}(X) = E[(X - \mu)^2]$

covariance:  $\text{Cov}(R_i, R_j) = E\{[R_i - E(R_i)][R_j - E(R_j)]\}$

correlation:  $\text{Corr}(R_i, R_j) = \frac{\text{Cov}(R_i, R_j)}{\sigma(R_i)\sigma(R_j)}$

portfolio variance:  $\text{Var}(R_p) = w_A^2\sigma^2(R_A) + w_B^2\sigma^2(R_B) + 2w_Aw_B\sigma(R_A)\sigma(R_B)\rho(R_A, R_B)$

skewness =  $\frac{E[(R - \mu)^3]}{\sigma^3}$

kurtosis =  $\frac{E[(R - \mu)^4]}{\sigma^4}$

---

### Topic 17

Poisson distribution:  $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

binomial probability function: (number of ways to choose x from n)  $p^x(1 - p)^{n-x}$

expected value of a binomial random variable:  $E(X) = np$

variance of a binomial random variable:  $np(1 - p) = npq$

uniform distribution range:  $P(x_1 \leq X \leq x_2) = (x_2 - x_1)/(b - a)$

mean of uniform distribution:  $E(x) = \frac{a + b}{2}$

variance of uniform distribution:  $Var(x) = \frac{(b - a)^2}{12}$

---

### Topic 18

Bayes' theorem:  $P(A | B) = \frac{P(B | A) \times P(A)}{P(B)}$

---

### Topic 19

population mean:  $\mu = \frac{\sum_{i=1}^N X_i}{N}$

sample mean:  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$

population variance:  $\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$

population standard deviation:  $\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$

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

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

sample covariance: covariance =  $\sum_{i=1}^n \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$

sample correlation coefficient:  $r_{XY} = \frac{\text{Cov}(X, Y)}{(s_X)(s_Y)}$

$z = \frac{\text{observation} - \text{population mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$

sampling error of the mean = sample mean – population mean =  $\bar{x} - \mu$

standard error of the sample mean:  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

chi-squared test statistic:  $\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$

$F$ -test =  $\frac{s_1^2}{s_2^2}$

test statistic =  $\frac{\text{sample statistic} - \text{hypothesized value}}{\text{standard error of the sample statistic}}$

confidence interval:

$$\left| \frac{\text{sample statistic} - (\text{critical value})}{\text{error}} \right| < \frac{\text{population parameter}}{\text{standard error}} < \left| \frac{\text{sample statistic} + (\text{critical value})}{\text{error}} \right|$$

$t$ -statistic:  $t_{n-1} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$

$z$ -statistic =  $\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

## Topic 20

sample regression function:  $Y_i = b_0 + b_1 \times X_i + e_i$

residual:  $e_i = Y_i - (b_0 + b_1 \times X_i)$

$$\text{regression slope coefficient: } b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

$$\text{regression intercept: } b_0 = \bar{Y} - b_1 \bar{X}$$

where:

$\bar{Y}$  = mean of Y

$\bar{X}$  = mean of X

$$\text{sum of squared residuals (SSR)} = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$\text{total sum of squares} = \text{explained sum of squares} + \text{sum of squared residuals}$$

$$\begin{aligned} \sum (Y_i - \bar{Y})^2 &= \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2 \\ \text{TSS} &= \text{ESS} + \text{SSR} \end{aligned}$$

coefficient of determination:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{\text{SSR}}{\text{TSS}} = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2}$$


---

### Topic 22

$$\text{standard error of the regression: } \text{SER} = \sqrt{s_e^2} = \sqrt{\frac{\text{SSR}}{n - k - 1}}$$

$$F\text{-statistic} = (\text{ESS} / \text{df}) / (\text{SSR} / \text{df})$$

$$\text{adjusted } R^2 = 1 - (1 - R^2) \times \frac{n - 1}{n - k - 1}$$


---

### Topic 23

$$\text{homoskedasticity-only } F\text{-statistic: } F = \frac{(R_{ur}^2 - R_t^2)/m}{(1 - R_{ur}^2)/(n - k_{ur} - 1)}$$

## Topic 24

linear trend model:  $y_t = \beta_0 + \beta_1(t)$

quadratic trend model:  $y_t = \beta_0 + \beta_1(t) + \beta_2(t)^2$

exponential trend model:  $y_t = \beta_0 e^{\beta_1(t)}$

mean squared error (MSE):  $MSE = \frac{\sum_{t=1}^T e_t^2}{T}$

unbiased mean squared error ( $s^2$ ):  $s^2 = \left( \frac{T}{T-k} \right) \frac{\sum_{t=1}^T e_t^2}{T}$

Akaike information criterion (AIC):  $AIC = e^{\left( \frac{2k}{T} \right)} \frac{\sum_{t=1}^T e_t^2}{T}$

Schwarz information criterion (SIC):  $SIC = T^{\left( \frac{k}{T} \right)} \frac{\sum_{t=1}^T e_t^2}{T}$

## Topic 25

pure seasonal dummy model:  $y_t = \sum_{i=1}^s \gamma_i (D_{i,t}) + \varepsilon_t$

trend model with seasonality:  $y_t = \beta_1(t) + \sum_{i=1}^s \gamma_i (D_{i,t}) + \varepsilon_t$

## Topic 26

lag operators:  $L^m y_t = y_{t-m}$

sample autocorrelation:  $\hat{\rho}_\tau = \frac{\sum_{t=\tau+1}^T [(y_t - \bar{y})(y_{t-\tau} - \bar{y})]}{\sum_{t=1}^T (y_t - \bar{y})^2}$

---

## Topic 27

first-order moving average [MA(1)] process:

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

where:

- $y_t$  = the time series variable being estimated
- $\varepsilon_t$  = current random white noise shock
- $\varepsilon_{t-1}$  = one-period lagged random white noise shock
- $\theta$  = coefficient for the lagged random shock

MA(q) process:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

where:

- $y_t$  = the time series variable being estimated
- $\varepsilon_t$  = current random white noise shock
- $\varepsilon_{t-1}$  = one-period lagged random white noise shock
- $\varepsilon_{t-q}$  =  $q^{\text{th}}$ -period lagged random white noise shock
- $\theta$  = coefficients for the lagged random shocks

first-order autoregressive [AR(1)] process:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

where:

- $y_t$  = the time series variable being estimated
- $y_{t-1}$  = one-period lagged observation of the variable being estimated
- $\varepsilon_t$  = current random white noise shock
- $\phi$  = coefficient for the lagged observation of the variable being estimated

Yule-Walker equation:  $\rho_t = \phi^t$  for  $t = 0, 1, 2, \dots$

AR(p) process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

where:

- $y_t$  = the time series variable being estimated
- $y_{t-1}$  = one-period lagged observation of the variable being estimated
- $y_{t-p}$  =  $p^{\text{th}}$ -period lagged observation of the variable being estimated
- $\varepsilon_t$  = current random white noise shock
- $\phi$  = coefficients for the lagged observations of the variable being estimated

autoregressive moving average (ARMA) process:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

where:

$y_t$  = the time series variable being estimated

$\phi$  = coefficient for the lagged observations of the variable being estimated

$y_{t-1}$  = one-period lagged observation of the variable being estimated

$\varepsilon_t$  = current random white noise shock

$\theta$  = coefficient for the lagged random shocks

$\varepsilon_{t-1}$  = one-period lagged random white noise shock

## Topic 28

the power law:  $P(V > X) = K \times X^{-\alpha}$

continuously compounded return:  $u_i = \ln\left(\frac{S_i}{S_{i-1}}\right)$

exponentially weighted moving average (EWMA) model (volatility):

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2$$

where:

$\lambda$  = weight on previous volatility estimate ( $\lambda$  between zero and one)

GARCH(1,1) model (volatility):

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

where:

$\alpha$  = weighting on the previous period's return

$\beta$  = weighting on the previous volatility estimate

$\omega$  = weighted long-run variance =  $\gamma V_L$

$$V_L = \text{long-run average variance} = \frac{\omega}{1 - \alpha - \beta}$$

$$\alpha + \beta + \gamma = 1$$

$\alpha + \beta < 1$  for stability so that  $\gamma$  is not negative

### Topic 29

exponentially weighted moving average (EWMA) model (covariance):

$$\text{cov}_n = \lambda \text{cov}_{n-1} + (1 - \lambda) X_{n-1} Y_{n-1}$$

where:

$\lambda$  = the weight for the most recent covariance on day  $n - 1$

$X_{n-1}$  = the percentage change for variable  $X$  on day  $n - 1$

$Y_{n-1}$  = the percentage change for variable  $Y$  on day  $n - 1$

GARCH(1,1) model (covariance):  $\text{cov}_n = \omega + \alpha X_{n-1} Y_{n-1} + \beta \text{cov}_{n-1}$

covariance consistency condition:  $\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} \leq 1$

factor model:  $U_i = \alpha_i F + \sqrt{1 - \alpha_i^2} Z_i$

# USING THE CUMULATIVE Z-TABLE

## Probability Example

Assume that the annual earnings per share (EPS) for a large sample of firms is normally distributed with a mean of \$5.00 and a standard deviation of \$1.50. What is the approximate probability of an observed EPS value falling between \$3.00 and \$7.25?

If  $\text{EPS} = x = \$7.25$ , then  $z = (x - \mu)/\sigma = (\$7.25 - \$5.00)/\$1.50 = +1.50$

If  $\text{EPS} = x = \$3.00$ , then  $z = (x - \mu)/\sigma = (\$3.00 - \$5.00)/\$1.50 = -1.33$

*For z-value of 1.50:* Use the row headed 1.5 and the column headed 0 to find the value 0.9332. This represents the area under the curve to the left of the critical value 1.50.

*For z-value of -1.33:* Use the row headed 1.3 and the column headed 3 to find the value 0.9082. This represents the area under the curve to the left of the critical value +1.33. The area to the left of -1.33 is  $1 - 0.9082 = 0.0918$ .

The area between these critical values is  $0.9332 - 0.0918 = 0.8414$ , or 84.14%.

## Hypothesis Testing—One-Tailed Test Example

A sample of a stock's returns on 36 non-consecutive days results in a mean return of 2.0%. Assume the population standard deviation is 20.0%. Can we say with 95% confidence that the mean return is greater than 0%?

$$H_0: \mu \leq 0.0\%, H_A: \mu > 0.0\%. \text{ The test statistic } = z\text{-statistic} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \\ = (2.0 - 0.0) / (20.0 / 6) = 0.60.$$

The significance level =  $1.0 - 0.95 = 0.05$ , or 5%.

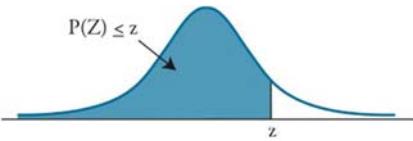
Since this is a one-tailed test with an alpha of 0.05, we need to find the value 0.95 in the cumulative z-table. The closest value is 0.9505, with a corresponding critical z-value of 1.65. Since the test statistic is less than the critical value, we fail to reject  $H_0$ .

## Hypothesis Testing—Two-Tailed Test Example

Using the same assumptions as before, suppose that the analyst now wants to determine if he can say with 99% confidence that the stock's return is not equal to 0.0%.

$$H_0: \mu = 0.0\%, H_A: \mu \neq 0.0\%. \text{ The test statistic (z-value)} = (2.0 - 0.0) / (20.0 / 6) = 0.60. \\ \text{The significance level} = 1.0 - 0.99 = 0.01, \text{ or } 1\%.$$

Since this is a two-tailed test with an alpha of 0.01, there is a 0.005 rejection region in both tails. Thus, we need to find the value 0.995 ( $1.0 - 0.005$ ) in the table. The closest value is 0.9951, which corresponds to a critical z-value of 2.58. Since the test statistic is less than the critical value, we fail to reject  $H_0$  and conclude that the stock's return equals 0.0%.

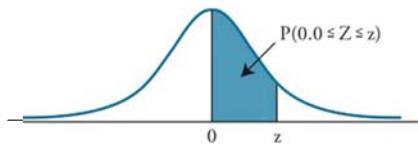


## CUMULATIVE Z-TABLE

$$P(Z \leq z) = N(z) \text{ for } z \geq 0$$

$$P(Z \leq -z) = 1 - N(z)$$

<b>z</b>	<b>0</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0</b>	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
<b>0.1</b>	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
<b>0.2</b>	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
<b>0.3</b>	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
<b>0.4</b>	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
<b>0.5</b>	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
<b>0.6</b>	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
<b>0.7</b>	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
<b>0.8</b>	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
<b>0.9</b>	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
<b>1</b>	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
<b>1.1</b>	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
<b>1.2</b>	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
<b>1.3</b>	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
<b>1.4</b>	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
<b>1.5</b>	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
<b>1.6</b>	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
<b>1.7</b>	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
<b>1.8</b>	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
<b>1.9</b>	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
<b>2</b>	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
<b>2.1</b>	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
<b>2.2</b>	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
<b>2.3</b>	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
<b>2.4</b>	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
<b>2.5</b>	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
<b>2.6</b>	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
<b>2.7</b>	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
<b>2.8</b>	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
<b>2.9</b>	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
<b>3</b>	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990



## ALTERNATIVE Z-TABLE

$$P(Z \leq z) = N(z) \text{ for } z \geq 0$$

$$P(Z \leq -z) = 1 - N(z)$$

<b>z</b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0.0</b>	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
<b>0.1</b>	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
<b>0.2</b>	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
<b>0.3</b>	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
<b>0.4</b>	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
<b>0.5</b>	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
<b>0.6</b>	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
<b>0.7</b>	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
<b>0.8</b>	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
<b>0.9</b>	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3356	0.3389
<b>1.0</b>	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
<b>1.1</b>	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
<b>1.2</b>	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
<b>1.3</b>	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
<b>1.4</b>	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
<b>1.5</b>	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
<b>1.6</b>	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
<b>1.7</b>	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
<b>1.8</b>	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
<b>1.9</b>	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
<b>2.0</b>	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
<b>2.1</b>	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
<b>2.2</b>	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
<b>2.3</b>	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
<b>2.4</b>	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
<b>2.5</b>	0.4939	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
<b>2.6</b>	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
<b>2.7</b>	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
<b>2.8</b>	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
<b>2.9</b>	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
<b>3.0</b>	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

# STUDENT'S T-DISTRIBUTION

Level of Significance for One-Tailed Test						
df	0.100	0.050	0.025	0.01	0.005	0.0005
Level of Significance for Two-Tailed Test						
df	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.294
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
$\infty$	1.282	1.645	1.960	2.326	2.576	3.291

# F-TABLE AT 5%

Critical values of the *F*-distribution at a 5% level of significance

Degrees of freedom for the numerator along top row

Degrees of freedom for the denominator along side row

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>12</b>	<b>15</b>	<b>20</b>	<b>24</b>	<b>30</b>	<b>40</b>
<b>1</b>	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
<b>2</b>	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
<b>3</b>	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
<b>4</b>	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
<b>5</b>	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
<b>6</b>	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
<b>7</b>	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
<b>8</b>	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
<b>9</b>	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	6.01	2.94	2.90	2.86	2.83
<b>10</b>	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
<b>11</b>	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
<b>12</b>	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
<b>13</b>	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
<b>14</b>	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
<b>15</b>	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
<b>16</b>	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
<b>17</b>	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
<b>18</b>	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
<b>19</b>	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
<b>20</b>	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99
<b>21</b>	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96
<b>22</b>	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94
<b>23</b>	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91
<b>24</b>	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89
<b>25</b>	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87
<b>30</b>	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79
<b>40</b>	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69
<b>60</b>	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59
<b>120</b>	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50
<b><math>\infty</math></b>	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39

# F-TABLE AT 2.5%

Critical values of the *F*-distribution at a 2.5% level of significance

Degrees of freedom for the numerator along top row

Degrees of freedom for the denominator along side row

	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	648	799	864	900	922	937	948	957	963	969	977	985	993	997	1001	1006
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.41	4.36	4.31
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61
$\infty$	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48

# CHI-SQUARED TABLE

Values of  $\chi^2$  (Degrees of Freedom, Level of Significance)  
Probability in Right Tail

Degrees of Freedom	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000157	0.000982	0.003932	0.0158	2.706	3.841	5.024	6.635	7.879
2	0.020100	0.050636	0.102586	0.2107	4.605	5.991	7.378	9.210	10.597
3	0.1148	0.2158	0.3518	0.5844	6.251	7.815	9.348	11.345	12.838
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
100	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.170

---

# INDEX

---

## A

addition rule 20  
adjusted  $R^2$  161, 181  
Akaike information criterion 199  
alternative hypothesis 101  
annuity 5  
ANOVA table 178  
antithetic variate technique 271  
arithmetic means 30  
asymptotic efficiency 200  
autocorrelation function 215  
autocovariance 215  
autocovariance function 215  
autoregression 215  
autoregressive conditional heteroskedasticity model 242  
autoregressive moving average process 233  
autoregressive representation 230

## B

backtesting 121  
Bayes' theorem 75  
Bernoulli distribution 55  
best linear unbiased estimator (BLUE) 48, 149  
binomial distribution 55  
binomial random variable 55  
bootstrapping method 274  
Box-Pierce Q-statistic 223

## C

central limit theorem 66, 134  
central moments 43  
central tendency 29  
Chebyshev's inequality 120  
chi-squared distribution 68  
chi-squared test 114  
coefficient of determination 135, 160  
coefficient of multiple correlation 179  
cokurtosis 46  
compounding 2  
compound interest 1  
conditional heteroskedasticity 147  
conditional probability 18, 75  
confidence interval 60, 96, 145  
    regression coefficient 142, 174  
consistency 200  
consistent estimator 48, 134

continuous random variable 14  
continuous uniform distribution 53  
control variate technique 272  
copula 258  
copula correlation 259  
correlation 40, 137, 251  
correlogram 222  
coskewness 46  
cost of capital 4  
covariance 38, 251  
covariance stationary 214  
critical value 101  
cumulative distribution function 15

## D

data generating process 200, 269  
data mining 198  
decision rule 103  
default risk 3  
degrees of freedom 170, 174  
dependent events 19  
dependent variable 128  
    predicting 175  
descriptive statistics 29  
deterministic trends 189  
discount factor 4  
discounting 2  
discount rates 3  
discrete probability function 16  
discrete random variable 13  
discrete uniform random variable 16  
displacement 214  
distributed lag 220  
dummy variables 147, 208  
dummy variable trap 163

## E

economic significance 109  
efficient estimator 48  
error term 130  
estimator 48  
event 13  
exceedance 121  
exception 121  
excess kurtosis 44, 46  
exhaustive events 13  
expectations 34  
expected value 34

explained sum of squares 135, 178  
explained variable 128  
explanatory variable 128  
exponentially weighted moving average 243, 253  
exponential trend 191

## F

factor model 258  
*F*-distribution 69  
first-order autoregressive process 231  
frequentist approach 80  
*F*-statistic 176  
*F*-test 117  
future value 1

## G

GARCH model 244, 255  
Gaussian copula 260  
Gaussian white noise 218  
Gauss-Markov theorem 149  
general linear process 221  
geometric mean 33

## H

heteroskedasticity 147, 159  
holiday variations 209  
homoskedasticity 147, 159  
homoskedasticity-only *F*-statistic 182  
h-step-ahead point forecast 210  
hypothesis 100  
hypothesis testing 143

## I

implied volatility 240  
independent and identically distributed (i.i.d.)  
    random variables 66  
independent events 19  
independent variable 128  
independent white noise 218  
inferential statistics 29  
infinite distributed lag 220  
innovations 221  
intercept 130, 158  
inverse cumulative distribution function 16

## J

joint probability 18, 21

## K

kurtosis 43, 45

## L

lag 214  
    distributed 220  
    infinite distributed 220  
    rational distributed 221  
lag operator 220  
lag operator polynomial 220  
leptokurtic distributions 45  
level of significance 96  
linear trend models 189  
liquidity risk 3  
Ljung-Box *Q*-statistic 223  
log-linear trend 192  
lognormal distribution 65

## M

marginal distributions 258  
maturity risk 3  
maximum likelihood estimator 245  
mean 42  
mean reversion 245  
mean squared error 197  
measures of fit 159  
median 31  
mesokurtic distributions 45  
mixture distributions 70  
mode 32  
Monte Carlo simulation 269  
moving average process 229  
moving average representation 230  
moving block bootstrap 275  
multicollinearity 162  
multiple regression 157  
    analysis 157  
    formulation 158  
    interpretation 158  
multiplication rule 18  
multivariate copula 261  
mutually exclusive events 13, 20

## N

negative skew 44  
noise component 130  
nominal risk-free rate 3  
non-independent data 275  
nonparametric distributions 53  
normal distribution 59  
normal white noise 218  
null hypothesis 101

## O

OLS estimators 157  
omitted variable bias 156

- one-factor copula 262
- one-tailed test 102, 174
- opportunity cost 3, 4
- ordinary least squares 132, 192
- outcome 13
- outliers 44, 275
- P**
  - parameter 128
  - parametric distributions 53
  - partial autocorrelation function 215
  - partial slope coefficients 158
  - penalty factor 199
  - perpetuity 6
  - persistence 245
  - platykurtic distributions 45
  - point estimate 48, 96
  - Poisson distribution 58
  - population 29
  - population mean 30, 90
  - population variance 91
  - positive-semidefinite 256
  - positive skew 44
  - posterior probabilities 82
  - power of a test 107
  - predicted values 145
  - present value 1
  - present value factor 4
  - price relatives 66
  - probability density function 15
  - probability distribution 13
  - probability function 13
  - probability matrix 23
  - pseudo-random number generators 275
  - p*-value 110, 144, 172, 179
- Q**
  - quadratic trend 190
- R**
  - $R^2$ , adjusted 161, 181
  - $R^2$ , coefficient of determination 135, 160
  - random number generation 275
  - random variable 13
  - rational distributed lag 221
  - real risk-free rate 3
  - regression analysis 128
    - heteroskedasticity 148
    - multicollinearity 163
  - regression coefficient 129
    - confidence interval 142, 174
    - hypothesis testing 170
  - t*-test 143
- required rate of return 3, 4
- residual 131
- residual plot 148
- restricted  $R^2$  182
- robust standard errors 149
- S**
  - $s^2$  measure 198
  - sample 29
  - sample autocorrelation 222
  - sample autocorrelation function 222
  - sample covariance 95
  - sample mean 30, 90
  - sample partial autocorrelation 222
  - sample regression function 130
  - sample standard deviation 93
  - sample variance 92
  - sampling distribution 89
  - scatter plot 41, 128
  - Schwarz information criterion 199
  - seasonal dummy variables 208
  - seasonality 206
  - seasonally adjusted time series 207
  - seed 276
  - serially uncorrelated 218
  - skewness 43, 44
  - slope coefficient 130
  - standard deviation 37
  - standard error 90, 93
  - standard error of the forecast 146
  - standard error of the regression 137, 159
  - statistical significance 109, 171
  - stochastic seasonality 206
  - Student's *t*-copula 261
  - Student's *t*-distribution 66
  - sum of squared residuals 134, 178
  - symmetrical distributions 44
- T**
  - tail dependence 262
  - t*-distribution 66
  - test statistic 101
  - the power law 240
  - time value of money 1
  - total sum of squares 135, 178
  - trading-day variations 209
  - trend 189
  - t*-test 110, 143
  - two-tailed test 102, 173
  - Type I error 107
  - Type II error 107, 163

**U**

unbiased estimator 48, 134  
unconditional heteroskedasticity 147  
unconditional probability 18, 75  
uniform distribution 53  
unrestricted R<sup>2</sup> 182

**V**

variance 37, 43  
variance-covariance matrix 256  
variance rate 240  
Venn diagram 20  
volatility 239  
volatility index 240

**W**

white noise 218  
    Gaussian 218  
    independent 218  
    normal 218  
Wold's theorem (Wold's representation) 221

**Y**

Yule-Walker equation 232

**Z**

z-distribution 61  
z-test 111  
z-value 61

**Required Disclaimers:**

CFA Institute does not endorse, promote, or warrant the accuracy or quality of the products or services offered by Kaplan. CFA Institute, CFA®, and Chartered Financial Analyst® are trademarks owned by CFA Institute.

Certified Financial Planner Board of Standards Inc. owns the certification marks CFP®, CERTIFIED FINANCIAL PLANNER™, and federally registered CFP (with flame design) in the U.S., which it awards to individuals who successfully complete initial and ongoing certification requirements. Kaplan does not certify individuals to use the CFP®, CERTIFIED FINANCIAL PLANNER™, and CFP (with flame design) certification marks. CFP® certification is granted only by Certified Financial Planner Board of Standards Inc. to those persons who, in addition to completing an educational requirement such as this CFP® Board-Registered Program, have met its ethics, experience, and examination requirements.

Kaplan is a review course provider for the CFP® Certification Examination administered by Certified Financial Planner Board of Standards Inc. CFP Board does not endorse any review course or receive financial remuneration from review course providers.

GARP® does not endorse, promote, review, or warrant the accuracy of the products or services offered by Kaplan or FRM® related information, nor does it endorse any pass rates claimed by the provider. Further, GARP® is not responsible for any fees or costs paid by the user to Kaplan, nor is GARP® responsible for any fees or costs of any person or entity providing any services to Kaplan. FRM®, GARP®, and Global Association of Risk Professionals™ are trademarks owned by the Global Association of Risk Professionals, Inc.

CAIAA does not endorse, promote, review or warrant the accuracy of the products or services offered by Kaplan, nor does it endorse any pass rates claimed by the provider. CAIAA is not responsible for any fees or costs paid by the user to Kaplan nor is CAIAA responsible for any fees or costs of any person or entity providing any services to Kaplan. CAIA®, CAIA Association®, Chartered Alternative Investment Analyst™, and Chartered Alternative Investment Analyst Association® are service marks and trademarks owned by CHARTERED ALTERNATIVE INVESTMENT ANALYST ASSOCIATION, INC., a Massachusetts non-profit corporation with its principal place of business at Amherst, Massachusetts, and are used by permission.