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:

estimated regression coefficient ± (critical t-value)(coefficient standard error)

The value of dependent variable Y is predicted as:

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

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:

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

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: $B_1 \neq 0$ 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:

	R^2	*F*-statistic
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.

Variable	*Coefficient*	*t-Statistic*
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 + (8.00×1) + (4.00×0.00) + (12.35×0.14) – (0.50×1.5) = 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

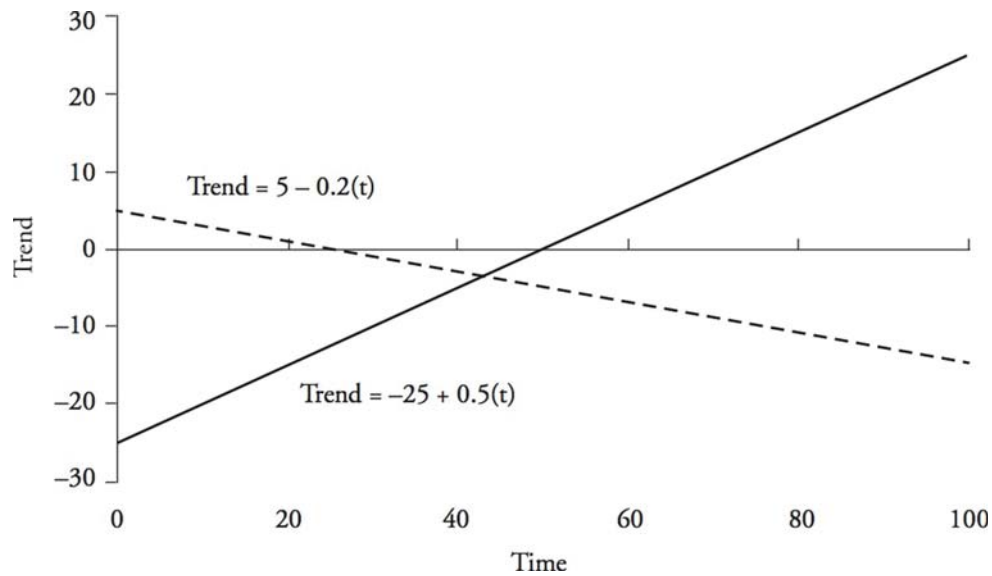
A *linear trend* is a time series pattern that can be graphed using a straight line. The simplest form of a linear trend is represented by the following model:

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

where:
y_t = the value of the time series (the dependent variable at time t)
β_0 = regression intercept at the vertical axis
β_1 = regression slope coefficient (or trend coefficient)
t = time trend or time dummy (the independent variable): t = 1, 2, 3, … ,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 β_1 in absolute value terms (e.g., 0.5) indicates a steeper slope, while a lower β_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

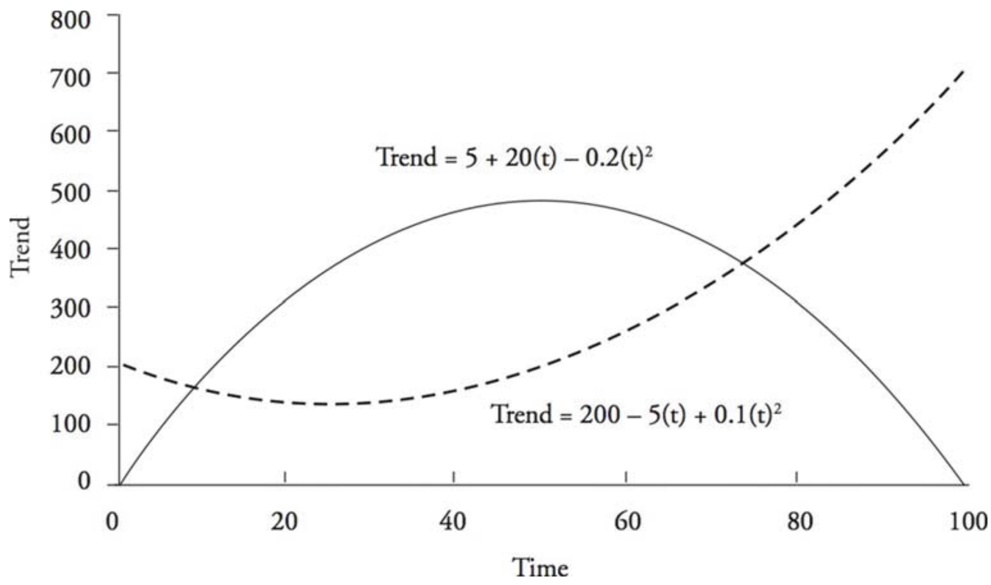
A nonlinear trend is a time series pattern that can be graphed with a curve. For example, a nonlinear trend would result if a variable increases at an increasing rate. When estimating and forecasting trends, a trend is not required to be linear; however, it should exhibit a smooth pattern. Nonlinear 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 β_1 and β_2 are positive, the trend increases at an increasing rate over time. Conversely, when both β_1 and β_2 are negative, the trend decreases at an increasing rate over time. When β_1 is negative and β_2 is positive, the trend will resemble a "U" shape. Finally, when β_1 is positive and β_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.

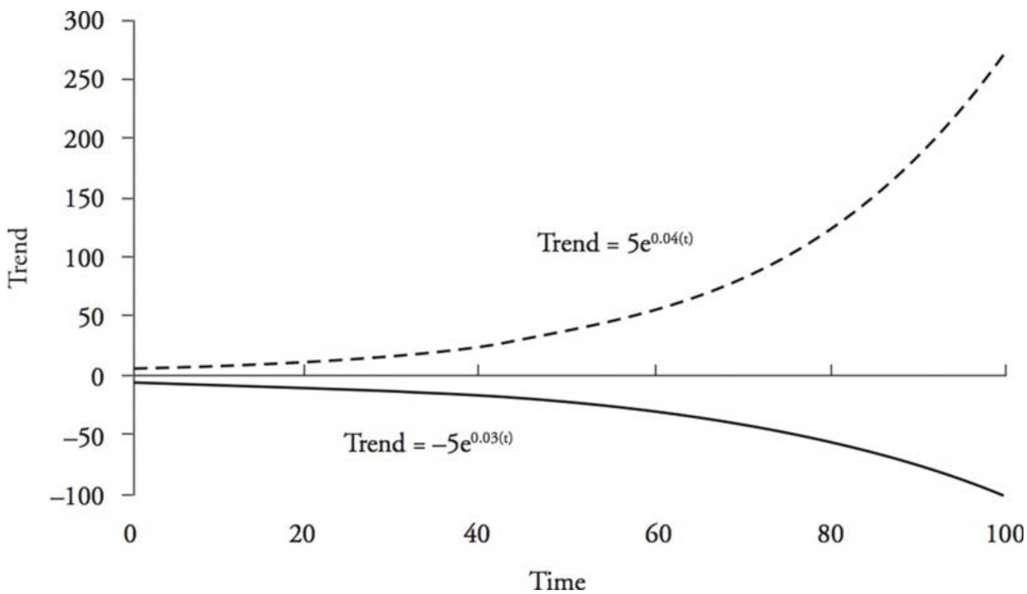
When a series exhibits exponential growth, it can be modeled using an exponential trend as follows:

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

where:
y_t = the value of the dependent variable at time t
β_0 = regression intercept term
β_1 = the constant rate of growth
t = time: t = 1, 2, 3, … ,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 results shown below.

Time Series Regression Results for Manufacturing Capacity Utilization

Regression model: $y_t = \beta_0 + \beta_1 t + \varepsilon_t$

R square	0.346		
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.

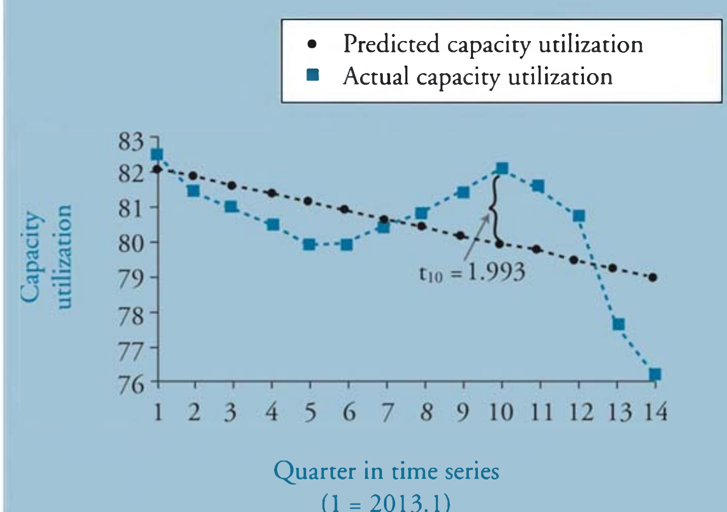
The projections (i.e., values generated by the model) for all quarters are compared to the actual values below.

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.

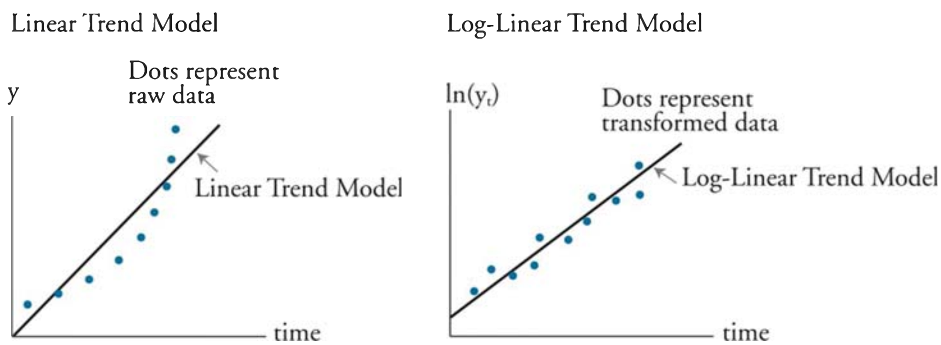
Selecting the Correct Trend Model

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.

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:

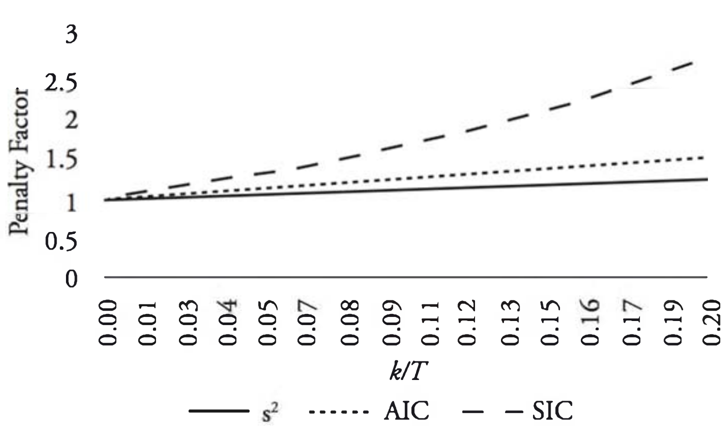
$$AIC = e^{\left(\frac{2k}{T}\right)}\frac{\sum_{t=1}^{T} e_t^2}{T}$$

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

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. $lnSUVsales_t = ln(\beta_0) + \beta_1(t)$.
C. $lnSUVsales_t = \beta_0 + \beta_1(SUVsales_{t-1})$.
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.

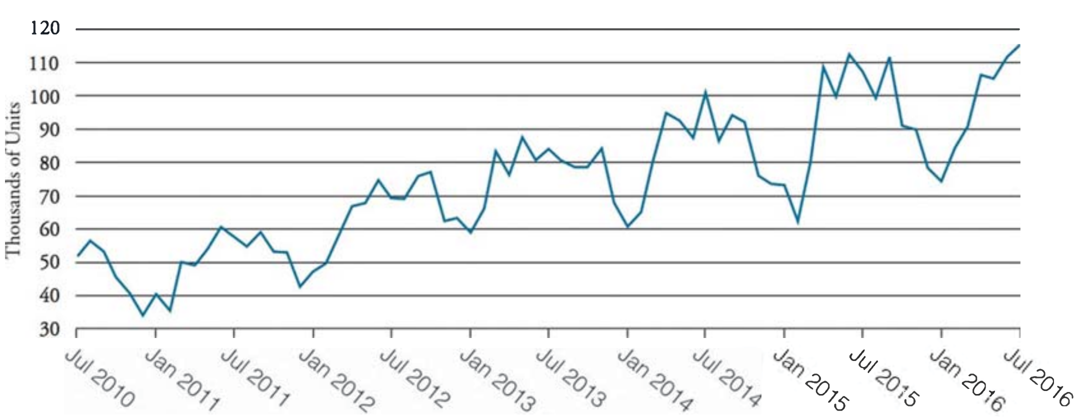
Annual changes can be approximate, as in the case of **stochastic seasonality**, or exact, as in the case of **deterministic seasonality**. Similar to the previous topic, where we focused on deterministic trends, the focus here will be exclusively on deterministic seasonality.

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.

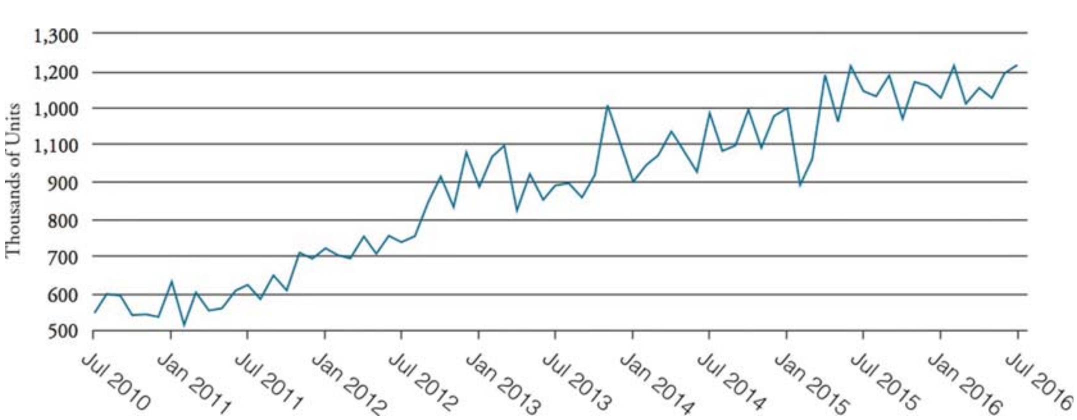
Figures 1 and 2 illustrate the difference between data that is not seasonally adjusted and data that is seasonally adjusted, using data for housing starts of privately owned homes between July 2010 and July 2016. As you can see from Figure 1, data that is not seasonally adjusted fluctuates greatly throughout 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 γ represent seasonal factors and the D represent the dummy variables. (If all of the γ_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:

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

where
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, β_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 β_0 = 1.25, β_1 = 0.75, β_2 = –0.20, and β_3 = 0.10:

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

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 $\widehat{EPS}_t = 1.25 + 0.75(1) - 0.20(0) + 0.10(0) = 2.00$. 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:

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

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

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

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

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

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:

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

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:

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

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:

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

   where:
   $SALES_t$ = 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 β_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 γ) 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 γ represent seasonal factors (i.e., the intercepts). If a time series does not exhibit seasonality, all γ_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.