IBM SPSS Forecasting 22

Product Information

This edition applies to version 22, release 0, modification 0 of IBM SPSS Statistics and to all subsequent releases and modifications until otherwise indicated in new editions.

Contents

Chapter 1. Introduction to Time Series

A **time series** is a set of observations obtained by measuring a single variable regularly over a period of time. In a series of inventory data, for example, the observations might represent daily inventory levels for several months. A series showing the market share of a product might consist of weekly market share taken over a few years. A series of total sales figures might consist of one observation per month for many years. What each of these examples has in common is that some variable was observed at regular, known intervals over a certain length of time. Thus, the form of the data for a typical time series is a single sequence or list of observations representing measurements taken at regular intervals.

Table 1. Daily inventory time series

Time	Week	Day	Inventory level
t_1	1	Monday	160
t_2	1	Tuesday	135
t_3	1	Wednesday	129
t_4	1	Thursday	122
t_5	1	Friday	108
t_6	2	Monday	150
		...	
t_{60}	12	Friday	120

One of the most important reasons for doing time series analysis is to try to forecast future values of the series. A model of the series that explained the past values may also predict whether and how much the next few values will increase or decrease. The ability to make such predictions successfully is obviously important to any business or scientific field.

Time Series Data

When you define time series data for use with the Forecasting add-on module, each series corresponds to a separate variable. For example, to define a time series in the Data Editor, click the **Variable View** tab and enter a variable name in any blank row. Each observation in a time series corresponds to a case (a row in the Data Editor).

If you open a spreadsheet containing time series data, each series should be arranged in a column in the spreadsheet. If you already have a spreadsheet with time series arranged in rows, you can open it anyway and use Transpose on the Data menu to flip the rows into columns.

Data Transformations

A number of data transformation procedures provided in the Core system are useful in time series analysis.

- The Define Dates procedure (on the Data menu) generates date variables used to establish periodicity and to distinguish between historical, validation, and forecasting periods. Forecasting is designed to work with the variables created by the Define Dates procedure.
- The Create Time Series procedure (on the Transform menu) creates new time series variables as functions of existing time series variables. It includes functions that use neighboring observations for smoothing, averaging, and differencing.

- The Replace Missing Values procedure (on the Transform menu) replaces system- and user-missing values with estimates based on one of several methods. Missing data at the beginning or end of a series pose no particular problem; they simply shorten the useful length of the series. Gaps in the middle of a series (*embedded* missing data) can be a much more serious problem.

See the *Core System User's Guide* for detailed information concerning data transformations for time series.

Estimation and Validation Periods

It is often useful to divide your time series into an *estimation*, or *historical*, period and a *validation* period. You develop a model on the basis of the observations in the estimation (historical) period and then test it to see how well it works in the validation period. By forcing the model to make predictions for points you already know (the points in the validation period), you get an idea of how well the model does at forecasting.

The cases in the validation period are typically referred to as holdout cases because they are held-back from the model-building process. The estimation period consists of the currently selected cases in the active dataset. Any remaining cases following the last selected case can be used as holdouts. Once you're satisfied that the model does an adequate job of forecasting, you can redefine the estimation period to include the holdout cases, and then build your final model.

Building Models and Producing Forecasts

The Forecasting add-on module provides two procedures for accomplishing the tasks of creating models and producing forecasts.
- The Chapter 2, "Time Series Modeler," on page 3 procedure creates models for time series, and produces forecasts. It includes an Expert Modeler that automatically determines the best model for each of your time series. For experienced analysts who want a greater degree of control, it also provides tools for custom model building.
- The Chapter 3, "Apply Time Series Models," on page 15 procedure applies existing time series models--created by the Time Series Modeler--to the active dataset. This allows you to obtain forecasts for series for which new or revised data are available, without rebuilding your models. If there's reason to think that a model has changed, it can be rebuilt using the Time Series Modeler.

Chapter 2. Time Series Modeler

The Time Series Modeler procedure estimates exponential smoothing, univariate Autoregressive Integrated Moving Average (ARIMA), and multivariate ARIMA (or transfer function models) models for time series, and produces forecasts. The procedure includes an Expert Modeler that attempts to automatically identify and estimate the best-fitting ARIMA or exponential smoothing model for one or more dependent variable series, thus eliminating the need to identify an appropriate model through trial and error. Alternatively, you can specify a custom ARIMA or exponential smoothing model.

Example. You are a product manager responsible for forecasting next month's unit sales and revenue for each of 100 separate products, and have little or no experience in modeling time series. Your historical unit sales data for all 100 products is stored in a single Excel spreadsheet. After opening your spreadsheet in IBM® SPSS® Statistics, you use the Expert Modeler and request forecasts one month into the future. The Expert Modeler finds the best model of unit sales for each of your products, and uses those models to produce the forecasts. Since the Expert Modeler can handle multiple input series, you only have to run the procedure once to obtain forecasts for all of your products. Choosing to save the forecasts to the active dataset, you can easily export the results back to Excel.

Statistics. Goodness-of-fit measures: stationary R-square, R-square (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), maximum absolute error (MaxAE), maximum absolute percentage error (MaxAPE), normalized Bayesian information criterion (BIC). Residuals: autocorrelation function, partial autocorrelation function, Ljung-Box Q. For ARIMA models: ARIMA orders for dependent variables, transfer function orders for independent variables, and outlier estimates. Also, smoothing parameter estimates for exponential smoothing models.

Plots. Summary plots across all models: histograms of stationary R-square, R-square (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), maximum absolute error (MaxAE), maximum absolute percentage error (MaxAPE), normalized Bayesian information criterion (BIC); box plots of residual autocorrelations and partial autocorrelations. Results for individual models: forecast values, fit values, observed values, upper and lower confidence limits, residual autocorrelations and partial autocorrelations.

Time Series Modeler Data Considerations

Data. The dependent variable and any independent variables should be numeric.

Assumptions. The dependent variable and any independent variables are treated as time series, meaning that each case represents a time point, with successive cases separated by a constant time interval.

- **Stationarity.** For custom ARIMA models, the time series to be modeled should be stationary. The most effective way to transform a nonstationary series into a stationary one is through a difference transformation--available from the Create Time Series dialog box .
- **Forecasts.** For producing forecasts using models with independent (predictor) variables, the active dataset should contain values of these variables for all cases in the forecast period. Additionally, independent variables should not contain any missing values in the estimation period.

Defining Dates

Although not required, it's recommended to use the Define Dates dialog box to specify the date associated with the first case and the time interval between successive cases. This is done prior to using the Time Series Modeler and results in a set of variables that label the date associated with each case. It also sets an assumed periodicity of the data--for example, a periodicity of 12 if the time interval between successive cases is one month. This periodicity is required if you're interested in creating seasonal models.

If you're not interested in seasonal models and don't require date labels on your output, you can skip the Define Dates dialog box. The label associated with each case is then simply the case number.

To Use the Time Series Modeler

1. From the menus choose:

   Analyze > **Forecasting** > **Create Models...**
2. On the Variables tab, select one or more dependent variables to be modeled.
3. From the Method drop-down box, select a modeling method. For automatic modeling, leave the default method of **Expert Modeler**. This will invoke the Expert Modeler to determine the best-fitting model for each of the dependent variables.

   To produce forecasts:
4. Click the **Options** tab.
5. Specify the forecast period. This will produce a chart that includes forecasts and observed values.

Optionally, you can:

- Select one or more independent variables. Independent variables are treated much like predictor variables in regression analysis but are optional. They can be included in ARIMA models but not exponential smoothing models. If you specify **Expert Modeler** as the modeling method and include independent variables, only ARIMA models will be considered.
- Click **Criteria** to specify modeling details.
- Save predictions, confidence intervals, and noise residuals.
- Save the estimated models in XML format. Saved models can be applied to new or revised data to obtain updated forecasts without rebuilding models.
- Obtain summary statistics across all estimated models.
- Specify transfer functions for independent variables in custom ARIMA models.
- Enable automatic detection of outliers.
- Model specific time points as outliers for custom ARIMA models.

Modeling Methods

The available modeling methods are:

Expert Modeler. The Expert Modeler automatically finds the best-fitting model for each dependent series. If independent (predictor) variables are specified, the Expert Modeler selects, for inclusion in ARIMA models, those that have a statistically significant relationship with the dependent series. Model variables are transformed where appropriate using differencing and/or a square root or natural log transformation. By default, the Expert Modeler considers both exponential smoothing and ARIMA models. You can, however, limit the Expert Modeler to only search for ARIMA models or to only search for exponential smoothing models. You can also specify automatic detection of outliers.

Exponential Smoothing. Use this option to specify a custom exponential smoothing model. You can choose from a variety of exponential smoothing models that differ in their treatment of trend and seasonality.

ARIMA. Use this option to specify a custom ARIMA model. This involves explicitly specifying autoregressive and moving average orders, as well as the degree of differencing. You can include independent (predictor) variables and define transfer functions for any or all of them. You can also specify automatic detection of outliers or specify an explicit set of outliers.

Estimation and Forecast Periods

Estimation Period. The estimation period defines the set of cases used to determine the model. By default, the estimation period includes all cases in the active dataset. To set the estimation period, select **Based on time or case range** in the Select Cases dialog box. Depending on available data, the estimation period used by the procedure may vary by dependent variable and thus differ from the displayed value. For a given dependent variable, the true estimation period is the period left after eliminating any contiguous missing values of the variable occurring at the beginning or end of the specified estimation period.

Forecast Period. The forecast period begins at the first case after the estimation period, and by default goes through to the last case in the active dataset. You can set the end of the forecast period from the Options tab.

Specifying Options for the Expert Modeler

The Expert Modeler provides options for constraining the set of candidate models, specifying the handling of outliers, and including event variables.

Model Selection and Event Specification

The Model tab allows you to specify the types of models considered by the Expert Modeler and to specify event variables.

Model Type. The following options are available:
- **All models.** The Expert Modeler considers both ARIMA and exponential smoothing models.
- **Exponential smoothing models only.** The Expert Modeler only considers exponential smoothing models.
- **ARIMA models only.** The Expert Modeler only considers ARIMA models.

Expert Modeler considers seasonal models. This option is only enabled if a periodicity has been defined for the active dataset. When this option is selected (checked), the Expert Modeler considers both seasonal and nonseasonal models. If this option is not selected, the Expert Modeler only considers nonseasonal models.

Current Periodicity. Indicates the periodicity (if any) currently defined for the active dataset. The current periodicity is given as an integer--for example, 12 for annual periodicity, with each case representing a month. The value *None* is displayed if no periodicity has been set. Seasonal models require a periodicity. You can set the periodicity from the Define Dates dialog box.

Events. Select any independent variables that are to be treated as event variables. For event variables, cases with a value of 1 indicate times at which the dependent series are expected to be affected by the event. Values other than 1 indicate no effect.

Handling Outliers with the Expert Modeler

The Outliers tab allows you to choose automatic detection of outliers as well as the type of outliers to detect.

Detect outliers automatically. By default, automatic detection of outliers is not performed. Select (check) this option to perform automatic detection of outliers, then select one or more of the following outlier types:
- Additive
- Level shift
- Innovational
- Transient
- Seasonal additive

- Local trend
- Additive patch

See the topic Chapter 7, "Outlier Types," on page 27 for more information.

Custom Exponential Smoothing Models

Model Type. Exponential smoothing models [1] are classified as either seasonal or nonseasonal. Seasonal models are only available if a periodicity has been defined for the active dataset (see "Current Periodicity" below).

- *Simple*. This model is appropriate for series in which there is no trend or seasonality. Its only smoothing parameter is level. Simple exponential smoothing is most similar to an ARIMA model with zero orders of autoregression, one order of differencing, one order of moving average, and no constant.
- *Holt's linear trend*. This model is appropriate for series in which there is a linear trend and no seasonality. Its smoothing parameters are level and trend, which are not constrained by each other's values. Holt's model is more general than Brown's model but may take longer to compute for large series. Holt's exponential smoothing is most similar to an ARIMA model with zero orders of autoregression, two orders of differencing, and two orders of moving average.
- *Brown's linear trend*. This model is appropriate for series in which there is a linear trend and no seasonality. Its smoothing parameters are level and trend, which are assumed to be equal. Brown's model is therefore a special case of Holt's model. Brown's exponential smoothing is most similar to an ARIMA model with zero orders of autoregression, two orders of differencing, and two orders of moving average, with the coefficient for the second order of moving average equal to the square of one-half of the coefficient for the first order.
- *Damped trend*. This model is appropriate for series with a linear trend that is dying out and with no seasonality. Its smoothing parameters are level, trend, and damping trend. Damped exponential smoothing is most similar to an ARIMA model with 1 order of autoregression, 1 order of differencing, and 2 orders of moving average.
- *Simple seasonal*. This model is appropriate for series with no trend and a seasonal effect that is constant over time. Its smoothing parameters are level and season. Simple seasonal exponential smoothing is most similar to an ARIMA model with zero orders of autoregression, one order of differencing, one order of seasonal differencing, and orders 1, p, and p + 1 of moving average, where p is the number of periods in a seasonal interval (for monthly data, p = 12).
- *Winters' additive*. This model is appropriate for series with a linear trend and a seasonal effect that does not depend on the level of the series. Its smoothing parameters are level, trend, and season. Winters' additive exponential smoothing is most similar to an ARIMA model with zero orders of autoregression, one order of differencing, one order of seasonal differencing, and p + 1 orders of moving average, where p is the number of periods in a seasonal interval (for monthly data, p = 12).
- *Winters' multiplicative*. This model is appropriate for series with a linear trend and a seasonal effect that depends on the level of the series. Its smoothing parameters are level, trend, and season. Winters' multiplicative exponential smoothing is not similar to any ARIMA model.

Current Periodicity. Indicates the periodicity (if any) currently defined for the active dataset. The current periodicity is given as an integer--for example, 12 for annual periodicity, with each case representing a month. The value *None* is displayed if no periodicity has been set. Seasonal models require a periodicity. You can set the periodicity from the Define Dates dialog box.

Dependent Variable Transformation. You can specify a transformation performed on each dependent variable before it is modeled.

- **None.** No transformation is performed.
- **Square root.** Square root transformation.

---

1. Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting,* 4, 1-28.

- **Natural log.** Natural log transformation.

Custom ARIMA Models

The Time Series Modeler allows you to build custom nonseasonal or seasonal ARIMA (Autoregressive Integrated Moving Average) models--also known as Box-Jenkins [2] models--with or without a fixed set of predictor variables. You can define transfer functions for any or all of the predictor variables, and specify automatic detection of outliers, or specify an explicit set of outliers.

- All independent (predictor) variables specified on the Variables tab are explicitly included in the model. This is in contrast to using the Expert Modeler where independent variables are only included if they have a statistically significant relationship with the dependent variable.

Model Specification for Custom ARIMA Models

The Model tab allows you to specify the structure of a custom ARIMA model.

ARIMA Orders. Enter values for the various ARIMA components of your model into the corresponding cells of the Structure grid. All values must be non-negative integers. For autoregressive and moving average components, the value represents the maximum order. All positive lower orders will be included in the model. For example, if you specify 2, the model includes orders 2 and 1. Cells in the Seasonal column are only enabled if a periodicity has been defined for the active dataset (see "Current Periodicity" below).

- **Autoregressive (p).** The number of autoregressive orders in the model. Autoregressive orders specify which previous values from the series are used to predict current values. For example, an autoregressive order of 2 specifies that the value of the series two time periods in the past be used to predict the current value.
- **Difference (d).** Specifies the order of differencing applied to the series before estimating models. Differencing is necessary when trends are present (series with trends are typically nonstationary and ARIMA modeling assumes stationarity) and is used to remove their effect. The order of differencing corresponds to the degree of series trend--first-order differencing accounts for linear trends, second-order differencing accounts for quadratic trends, and so on.
- **Moving Average (q).** The number of moving average orders in the model. Moving average orders specify how deviations from the series mean for previous values are used to predict current values. For example, moving-average orders of 1 and 2 specify that deviations from the mean value of the series from each of the last two time periods be considered when predicting current values of the series.

Seasonal Orders. Seasonal autoregressive, moving average, and differencing components play the same roles as their nonseasonal counterparts. For seasonal orders, however, current series values are affected by previous series values separated by one or more seasonal periods. For example, for monthly data (seasonal period of 12), a seasonal order of 1 means that the current series value is affected by the series value 12 periods prior to the current one. A seasonal order of 1, for monthly data, is then the same as specifying a nonseasonal order of 12.

Current Periodicity. Indicates the periodicity (if any) currently defined for the active dataset. The current periodicity is given as an integer--for example, 12 for annual periodicity, with each case representing a month. The value *None* is displayed if no periodicity has been set. Seasonal models require a periodicity. You can set the periodicity from the Define Dates dialog box.

Dependent Variable Transformation. You can specify a transformation performed on each dependent variable before it is modeled.

- **None.** No transformation is performed.
- **Square root.** Square root transformation.

---

2. Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.

- **Natural log.** Natural log transformation.

Include constant in model. Inclusion of a constant is standard unless you are sure that the overall mean series value is 0. Excluding the constant is recommended when differencing is applied.

Transfer Functions in Custom ARIMA Models

The Transfer Function tab (only present if independent variables are specified) allows you to define transfer functions for any or all of the independent variables specified on the Variables tab. Transfer functions allow you to specify the manner in which past values of independent (predictor) variables are used to forecast future values of the dependent series.

Transfer Function Orders. Enter values for the various components of the transfer function into the corresponding cells of the Structure grid. All values must be non-negative integers. For numerator and denominator components, the value represents the maximum order. All positive lower orders will be included in the model. In addition, order 0 is always included for numerator components. For example, if you specify 2 for numerator, the model includes orders 2, 1, and 0. If you specify 3 for denominator, the model includes orders 3, 2, and 1. Cells in the Seasonal column are only enabled if a periodicity has been defined for the active dataset (see "Current Periodicity" below).

- **Numerator.** The numerator order of the transfer function. Specifies which previous values from the selected independent (predictor) series are used to predict current values of the dependent series. For example, a numerator order of 1 specifies that the value of an independent series one time period in the past--as well as the current value of the independent series--is used to predict the current value of each dependent series.
- **Denominator.** The denominator order of the transfer function. Specifies how deviations from the series mean, for previous values of the selected independent (predictor) series, are used to predict current values of the dependent series. For example, a denominator order of 1 specifies that deviations from the mean value of an independent series one time period in the past be considered when predicting the current value of each dependent series.
- **Difference.** Specifies the order of differencing applied to the selected independent (predictor) series before estimating models. Differencing is necessary when trends are present and is used to remove their effect.

Seasonal Orders. Seasonal numerator, denominator, and differencing components play the same roles as their nonseasonal counterparts. For seasonal orders, however, current series values are affected by previous series values separated by one or more seasonal periods. For example, for monthly data (seasonal period of 12), a seasonal order of 1 means that the current series value is affected by the series value 12 periods prior to the current one. A seasonal order of 1, for monthly data, is then the same as specifying a nonseasonal order of 12.

Current Periodicity. Indicates the periodicity (if any) currently defined for the active dataset. The current periodicity is given as an integer--for example, 12 for annual periodicity, with each case representing a month. The value *None* is displayed if no periodicity has been set. Seasonal models require a periodicity. You can set the periodicity from the Define Dates dialog box.

Delay. Setting a delay causes the independent variable's influence to be delayed by the number of intervals specified. For example, if the delay is set to 5, the value of the independent variable at time t doesn't affect forecasts until five periods have elapsed ($t + 5$).

Transformation. Specification of a transfer function, for a set of independent variables, also includes an optional transformation to be performed on those variables.
- **None.** No transformation is performed.
- **Square root.** Square root transformation.
- **Natural log.** Natural log transformation.

Outliers in Custom ARIMA Models

The Outliers tab provides the following choices for the handling of outliers [3]: detect them automatically, specify particular points as outliers, or do not detect or model them.

Do not detect outliers or model them. By default, outliers are neither detected nor modeled. Select this option to disable any detection or modeling of outliers.

Detect outliers automatically. Select this option to perform automatic detection of outliers, and select one or more of the following outlier types:

- Additive
- Level shift
- Innovational
- Transient
- Seasonal additive
- Local trend
- Additive patch

See the topic Chapter 7, "Outlier Types," on page 27 for more information.

Model specific time points as outliers. Select this option to specify particular time points as outliers. Use a separate row of the Outlier Definition grid for each outlier. Enter values for all of the cells in a given row.

- **Type.** The outlier type. The supported types are: additive (default), level shift, innovational, transient, seasonal additive, and local trend.

Note 1: If no date specification has been defined for the active dataset, the Outlier Definition grid shows the single column *Observation*. To specify an outlier, enter the row number (as displayed in the Data Editor) of the relevant case.

Note 2: The *Cycle* column (if present) in the Outlier Definition grid refers to the value of the *CYCLE_* variable in the active dataset.

Output

Available output includes results for individual models as well as results calculated across all models. Results for individual models can be limited to a set of best- or poorest-fitting models based on user-specified criteria.

Statistics and Forecast Tables

The Statistics tab provides options for displaying tables of the modeling results.

Display fit measures, Ljung-Box statistic, and number of outliers by model. Select (check) this option to display a table containing selected fit measures, Ljung-Box value, and the number of outliers for each estimated model.

Fit Measures. You can select one or more of the following for inclusion in the table containing fit measures for each estimated model:

- Stationary *R*-square
- *R*-square
- Root mean square error

---

3. Pena, D., G. C. Tiao, and R. S. Tsay, eds. 2001. *A course in time series analysis*. New York: John Wiley and Sons.

- Mean absolute percentage error
- Mean absolute error
- Maximum absolute percentage error
- Maximum absolute error
- Normalized BIC

See the topic Chapter 6, "Goodness-of-Fit Measures," on page 25 for more information.

Statistics for Comparing Models. This group of options controls display of tables containing statistics calculated across all estimated models. Each option generates a separate table. You can select one or more of the following options:

- **Goodness of fit.** Table of summary statistics and percentiles for stationary *R*-square, *R*-square, root mean square error, mean absolute percentage error, mean absolute error, maximum absolute percentage error, maximum absolute error, and normalized Bayesian Information Criterion.
- **Residual autocorrelation function (ACF).** Table of summary statistics and percentiles for autocorrelations of the residuals across all estimated models.
- **Residual partial autocorrelation function (PACF).** Table of summary statistics and percentiles for partial autocorrelations of the residuals across all estimated models.

Statistics for Individual Models. This group of options controls display of tables containing detailed information for each estimated model. Each option generates a separate table. You can select one or more of the following options:

- **Parameter estimates.** Displays a table of parameter estimates for each estimated model. Separate tables are displayed for exponential smoothing and ARIMA models. If outliers exist, parameter estimates for them are also displayed in a separate table.
- **Residual autocorrelation function (ACF).** Displays a table of residual autocorrelations by lag for each estimated model. The table includes the confidence intervals for the autocorrelations.
- **Residual partial autocorrelation function (PACF).** Displays a table of residual partial autocorrelations by lag for each estimated model. The table includes the confidence intervals for the partial autocorrelations.

Display forecasts. Displays a table of model forecasts and confidence intervals for each estimated model. The forecast period is set from the Options tab.

Plots

The Plots tab provides options for displaying plots of the modeling results.

Plots for Comparing Models

This group of options controls display of plots containing statistics calculated across all estimated models. Each option generates a separate plot. You can select one or more of the following options:

- Stationary *R*-square
- *R*-square
- Root mean square error
- Mean absolute percentage error
- Mean absolute error
- Maximum absolute percentage error
- Maximum absolute error
- Normalized BIC
- Residual autocorrelation function (ACF)

- Residual partial autocorrelation function (PACF)

See the topic Chapter 6, "Goodness-of-Fit Measures," on page 25 for more information.

Plots for Individual Models

Series. Select (check) this option to obtain plots of the predicted values for each estimated model. You can select one or more of the following for inclusion in the plot:
- **Observed values.** The observed values of the dependent series.
- **Forecasts.** The model predicted values for the forecast period.
- **Fit values.** The model predicted values for the estimation period.
- **Confidence intervals for forecasts.** The confidence intervals for the forecast period.
- **Confidence intervals for fit values.** The confidence intervals for the estimation period.

Residual autocorrelation function (ACF). Displays a plot of residual autocorrelations for each estimated model.

Residual partial autocorrelation function (PACF). Displays a plot of residual partial autocorrelations for each estimated model.

Limiting Output to the Best- or Poorest-Fitting Models

The Output Filter tab provides options for restricting both tabular and chart output to a subset of the estimated models. You can choose to limit output to the best-fitting and/or the poorest-fitting models according to fit criteria you provide. By default, all estimated models are included in the output.

Best-fitting models. Select (check) this option to include the best-fitting models in the output. Select a goodness-of-fit measure and specify the number of models to include. Selecting this option does not preclude also selecting the poorest-fitting models. In that case, the output will consist of the poorest-fitting models as well as the best-fitting ones.
- **Fixed number of models.** Specifies that results are displayed for the n best-fitting models. If the number exceeds the number of estimated models, all models are displayed.
- **Percentage of total number of models.** Specifies that results are displayed for models with goodness-of-fit values in the top n percent across all estimated models.

Poorest-fitting models. Select (check) this option to include the poorest-fitting models in the output. Select a goodness-of-fit measure and specify the number of models to include. Selecting this option does not preclude also selecting the best-fitting models. In that case, the output will consist of the best-fitting models as well as the poorest-fitting ones.
- **Fixed number of models.** Specifies that results are displayed for the n poorest-fitting models. If the number exceeds the number of estimated models, all models are displayed.
- **Percentage of total number of models.** Specifies that results are displayed for models with goodness-of-fit values in the bottom n percent across all estimated models.

Goodness of Fit Measure. Select the goodness-of-fit measure to use for filtering models. The default is stationary R square.

Saving Model Predictions and Model Specifications

The Save tab allows you to save model predictions as new variables in the active dataset and save model specifications to an external file in XML format.

Save Variables. You can save model predictions, confidence intervals, and residuals as new variables in the active dataset. Each dependent series gives rise to its own set of new variables, and each new variable

contains values for both the estimation and forecast periods. New cases are added if the forecast period extends beyond the length of the dependent variable series. Choose to save new variables by selecting the associated Save check box for each. By default, no new variables are saved.

- **Predicted Values.** The model predicted values.
- **Lower Confidence Limits.** Lower confidence limits for the predicted values.
- **Upper Confidence Limits.** Upper confidence limits for the predicted values.
- **Noise Residuals.** The model residuals. When transformations of the dependent variable are performed (for example, natural log), these are the residuals for the transformed series.
- **Variable Name Prefix.** Specify prefixes to be used for new variable names, or leave the default prefixes. Variable names consist of the prefix, the name of the associated dependent variable, and a model identifier. The variable name is extended if necessary to avoid variable naming conflicts. The prefix must conform to the rules for valid variable names.

Export Model File. Model specifications for all estimated models are exported to the specified file in XML format. Saved models can be used to obtain updated forecasts.

- **XML File.** Model specifications are saved in an XML file that can be used with IBM SPSS applications.
- **PMML File.** Model specifications are saved in a PMML-compliant XML file that can be used with PMML-compliant applications, including IBM SPSS applications.

Options

The Options tab allows you to set the forecast period, specify the handling of missing values, set the confidence interval width, specify a custom prefix for model identifiers, and set the number of lags shown for autocorrelations.

Forecast Period. The forecast period always begins with the first case after the end of the estimation period (the set of cases used to determine the model) and goes through either the last case in the active dataset or a user-specified date. By default, the end of the estimation period is the last case in the active dataset, but it can be changed from the Select Cases dialog box by selecting **Based on time or case range**.

- **First case after end of estimation period through last case in active dataset.** Select this option when the end of the estimation period is prior to the last case in the active dataset, and you want forecasts through the last case. This option is typically used to produce forecasts for a holdout period, allowing comparison of the model predictions with a subset of the actual values.
- **First case after end of estimation period through a specified date.** Select this option to explicitly specify the end of the forecast period. This option is typically used to produce forecasts beyond the end of the actual series. Enter values for all of the cells in the Date grid.

  If no date specification has been defined for the active dataset, the Date grid shows the single column *Observation*. To specify the end of the forecast period, enter the row number (as displayed in the Data Editor) of the relevant case.

  The *Cycle* column (if present) in the Date grid refers to the value of the *CYCLE_* variable in the active dataset.

User-Missing Values. These options control the handling of user-missing values.

- **Treat as invalid.** User-missing values are treated like system-missing values.
- **Treat as valid.** User-missing values are treated as valid data.

Missing Value Policy. The following rules apply to the treatment of missing values (includes system-missing values and user-missing values treated as invalid) during the modeling procedure:

- Cases with missing values of a dependent variable that occur within the estimation period are included in the model. The specific handling of the missing value depends on the estimation method.

- A warning is issued if an independent variable has missing values within the estimation period. For the Expert Modeler, models involving the independent variable are estimated without the variable. For custom ARIMA, models involving the independent variable are not estimated.
- If any independent variable has missing values within the forecast period, the procedure issues a warning and forecasts as far as it can.

Confidence Interval Width (%). Confidence intervals are computed for the model predictions and residual autocorrelations. You can specify any positive value less than 100. By default, a 95% confidence interval is used.

Prefix for Model Identifiers in Output. Each dependent variable specified on the Variables tab gives rise to a separate estimated model. Models are distinguished with unique names consisting of a customizable prefix along with an integer suffix. You can enter a prefix or leave the default of *Model*.

Maximum Number of Lags Shown in ACF and PACF Output. You can set the maximum number of lags shown in tables and plots of autocorrelations and partial autocorrelations.

TSMODEL Command Additional Features

You can customize your time series modeling if you paste your selections into a syntax window and edit the resulting TSMODEL command syntax. The command syntax language allows you to:
- Specify the seasonal period of the data (with the SEASONLENGTH keyword on the AUXILIARY subcommand). This overrides the current periodicity (if any) for the active dataset.
- Specify nonconsecutive lags for custom ARIMA and transfer function components (with the ARIMA and TRANSFERFUNCTION subcommands). For example, you can specify a custom ARIMA model with autoregressive lags of orders 1, 3, and 6; or a transfer function with numerator lags of orders 2, 5, and 8.
- Provide more than one set of modeling specifications (for example, modeling method, ARIMA orders, independent variables, and so on) for a single run of the Time Series Modeler procedure (with the MODEL subcommand).

See the *Command Syntax Reference* for complete syntax information.

Chapter 3. Apply Time Series Models

The Apply Time Series Models procedure loads existing time series models from an external file and applies them to the active dataset. You can use this procedure to obtain forecasts for series for which new or revised data are available, without rebuilding your models. Models are generated using the Time Series Modeler procedure.

Example. You are an inventory manager with a major retailer, and responsible for each of 5,000 products. You've used the Expert Modeler to create models that forecast sales for each product three months into the future. Your data warehouse is refreshed each month with actual sales data which you'd like to use to produce monthly updated forecasts. The Apply Time Series Models procedure allows you to accomplish this using the original models, and simply reestimating model parameters to account for the new data.

Statistics. Goodness-of-fit measures: stationary R-square, R-square (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), maximum absolute error (MaxAE), maximum absolute percentage error (MaxAPE), normalized Bayesian information criterion (BIC). Residuals: autocorrelation function, partial autocorrelation function, Ljung-Box Q.

Plots. Summary plots across all models: histograms of stationary R-square, R-square (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), maximum absolute error (MaxAE), maximum absolute percentage error (MaxAPE), normalized Bayesian information criterion (BIC); box plots of residual autocorrelations and partial autocorrelations. Results for individual models: forecast values, fit values, observed values, upper and lower confidence limits, residual autocorrelations and partial autocorrelations.

Apply Time Series Models Data Considerations

Data. Variables (dependent and independent) to which models will be applied should be numeric.

Assumptions. Models are applied to variables in the active dataset with the same names as the variables specified in the model. All such variables are treated as time series, meaning that each case represents a time point, with successive cases separated by a constant time interval.

- **Forecasts.** For producing forecasts using models with independent (predictor) variables, the active dataset should contain values of these variables for all cases in the forecast period. If model parameters are reestimated, then independent variables should not contain any missing values in the estimation period.

Defining Dates

The Apply Time Series Models procedure requires that the periodicity, if any, of the active dataset matches the periodicity of the models to be applied. If you're simply forecasting using the same dataset (perhaps with new or revised data) as that used to the build the model, then this condition will be satisfied. If no periodicity exists for the active dataset, you will be given the opportunity to navigate to the Define Dates dialog box to create one. If, however, the models were created without specifying a periodicity, then the active dataset should also be without one.

To Apply Models

1. From the menus choose:

   Analyze > **Forecasting** > **Apply Models...**

2. Enter the file specification for a model file or click **Browse** and select a model file (model files are created with the Time Series Modeler procedure).

Optionally, you can:

- Reestimate model parameters using the data in the active dataset. Forecasts are created using the reestimated parameters.
- Save predictions, confidence intervals, and noise residuals.
- Save reestimated models in XML format.

Model Parameters and Goodness of Fit Measures

Load from model file. Forecasts are produced using the model parameters from the model file without reestimating those parameters. Goodness of fit measures displayed in output and used to filter models (best- or worst-fitting) are taken from the model file and reflect the data used when each model was developed (or last updated). With this option, forecasts do not take into account historical data--for either dependent or independent variables--in the active dataset. You must choose **Reestimate from data** if you want historical data to impact the forecasts. In addition, forecasts do not take into account values of the dependent series in the forecast period--but they do take into account values of independent variables in the forecast period. If you have more current values of the dependent series and want them to be included in the forecasts, you need to reestimate, adjusting the estimation period to include these values.

Reestimate from data. Model parameters are reestimated using the data in the active dataset. Reestimation of model parameters has no effect on model structure. For example, an ARIMA(1,0,1) model will remain so, but the autoregressive and moving-average parameters will be reestimated. Reestimation does not result in the detection of new outliers. Outliers, if any, are always taken from the model file.

- **Estimation Period.** The estimation period defines the set of cases used to reestimate the model parameters. By default, the estimation period includes all cases in the active dataset. To set the estimation period, select **Based on time or case range** in the Select Cases dialog box. Depending on available data, the estimation period used by the procedure may vary by model and thus differ from the displayed value. For a given model, the true estimation period is the period left after eliminating any contiguous missing values, from the model's dependent variable, occurring at the beginning or end of the specified estimation period.

Forecast Period

The forecast period for each model always begins with the first case after the end of the estimation period and goes through either the last case in the active dataset or a user-specified date. If parameters are not reestimated (this is the default), then the estimation period for each model is the set of cases used when the model was developed (or last updated).

- **First case after end of estimation period through last case in active dataset.** Select this option when the end of the estimation period is prior to the last case in the active dataset, and you want forecasts through the last case.
- **First case after end of estimation period through a specified date.** Select this option to explicitly specify the end of the forecast period. Enter values for all of the cells in the Date grid.

  If no date specification has been defined for the active dataset, the Date grid shows the single column *Observation*. To specify the end of the forecast period, enter the row number (as displayed in the Data Editor) of the relevant case.

  The *Cycle* column (if present) in the Date grid refers to the value of the *CYCLE_* variable in the active dataset.

Output

Available output includes results for individual models as well as results across all models. Results for individual models can be limited to a set of best- or poorest-fitting models based on user-specified criteria.

Statistics and Forecast Tables

The Statistics tab provides options for displaying tables of model fit statistics, model parameters, autocorrelation functions, and forecasts. Unless model parameters are reestimated (**Reestimate from data** on the Models tab), displayed values of fit measures, Ljung-Box values, and model parameters are those from the model file and reflect the data used when each model was developed (or last updated). Outlier information is always taken from the model file.

Display fit measures, Ljung-Box statistic, and number of outliers by model. Select (check) this option to display a table containing selected fit measures, Ljung-Box value, and the number of outliers for each model.

Fit Measures. You can select one or more of the following for inclusion in the table containing fit measures for each model:
- Stationary *R*-square
- *R*-square
- Root mean square error
- Mean absolute percentage error
- Mean absolute error
- Maximum absolute percentage error
- Maximum absolute error
- Normalized BIC

See the topic Chapter 6, "Goodness-of-Fit Measures," on page 25 for more information.

Statistics for Comparing Models. This group of options controls the display of tables containing statistics across all models. Each option generates a separate table. You can select one or more of the following options:
- **Goodness of fit.** Table of summary statistics and percentiles for stationary *R*-square, *R*-square, root mean square error, mean absolute percentage error, mean absolute error, maximum absolute percentage error, maximum absolute error, and normalized Bayesian Information Criterion.
- **Residual autocorrelation function (ACF).** Table of summary statistics and percentiles for autocorrelations of the residuals across all estimated models. This table is only available if model parameters are reestimated (**Reestimate from data** on the Models tab).
- **Residual partial autocorrelation function (PACF).** Table of summary statistics and percentiles for partial autocorrelations of the residuals across all estimated models. This table is only available if model parameters are reestimated (**Reestimate from data** on the Models tab).

Statistics for Individual Models. This group of options controls display of tables containing detailed information for each model. Each option generates a separate table. You can select one or more of the following options:
- **Parameter estimates.** Displays a table of parameter estimates for each model. Separate tables are displayed for exponential smoothing and ARIMA models. If outliers exist, parameter estimates for them are also displayed in a separate table.
- **Residual autocorrelation function (ACF).** Displays a table of residual autocorrelations by lag for each estimated model. The table includes the confidence intervals for the autocorrelations. This table is only available if model parameters are reestimated (**Reestimate from data** on the Models tab).
- **Residual partial autocorrelation function (PACF).** Displays a table of residual partial autocorrelations by lag for each estimated model. The table includes the confidence intervals for the partial autocorrelations. This table is only available if model parameters are reestimated (**Reestimate from data** on the Models tab).

Display forecasts. Displays a table of model forecasts and confidence intervals for each model.

Plots

The Plots tab provides options for displaying plots of model fit statistics, autocorrelation functions, and series values (including forecasts).

Plots for Comparing Models

This group of options controls the display of plots containing statistics across all models. Unless model parameters are reestimated (**Reestimate from data** on the Models tab), displayed values are those from the model file and reflect the data used when each model was developed (or last updated). In addition, autocorrelation plots are only available if model parameters are reestimated. Each option generates a separate plot. You can select one or more of the following options:

- Stationary *R*-square
- *R*-square
- Root mean square error
- Mean absolute percentage error
- Mean absolute error
- Maximum absolute percentage error
- Maximum absolute error
- Normalized BIC
- Residual autocorrelation function (ACF)
- Residual partial autocorrelation function (PACF)

See the topic Chapter 6, "Goodness-of-Fit Measures," on page 25 for more information.

Plots for Individual Models

Series. Select (check) this option to obtain plots of the predicted values for each model. Observed values, fit values, confidence intervals for fit values, and autocorrelations are only available if model parameters are reestimated (**Reestimate from data** on the Models tab). You can select one or more of the following for inclusion in the plot:

- **Observed values.** The observed values of the dependent series.
- **Forecasts.** The model predicted values for the forecast period.
- **Fit values.** The model predicted values for the estimation period.
- **Confidence intervals for forecasts.** The confidence intervals for the forecast period.
- **Confidence intervals for fit values.** The confidence intervals for the estimation period.

Residual autocorrelation function (ACF). Displays a plot of residual autocorrelations for each estimated model.

Residual partial autocorrelation function (PACF). Displays a plot of residual partial autocorrelations for each estimated model.

Limiting Output to the Best- or Poorest-Fitting Models

The Output Filter tab provides options for restricting both tabular and chart output to a subset of models. You can choose to limit output to the best-fitting and/or the poorest-fitting models according to fit criteria you provide. By default, all models are included in the output. Unless model parameters are reestimated (**Reestimate from data** on the Models tab), values of fit measures used for filtering models are those from the model file and reflect the data used when each model was developed (or last updated).

Best-fitting models. Select (check) this option to include the best-fitting models in the output. Select a goodness-of-fit measure and specify the number of models to include. Selecting this option does not preclude also selecting the poorest-fitting models. In that case, the output will consist of the poorest-fitting models as well as the best-fitting ones.

- **Fixed number of models.** Specifies that results are displayed for the *n* best-fitting models. If the number exceeds the total number of models, all models are displayed.
- **Percentage of total number of models.** Specifies that results are displayed for models with goodness-of-fit values in the top *n* percent across all models.

Poorest-fitting models. Select (check) this option to include the poorest-fitting models in the output. Select a goodness-of-fit measure and specify the number of models to include. Selecting this option does not preclude also selecting the best-fitting models. In that case, the output will consist of the best-fitting models as well as the poorest-fitting ones.

- **Fixed number of models.** Specifies that results are displayed for the *n* poorest-fitting models. If the number exceeds the total number of models, all models are displayed.
- **Percentage of total number of models.** Specifies that results are displayed for models with goodness-of-fit values in the bottom *n* percent across all models.

Goodness of Fit Measure. Select the goodness-of-fit measure to use for filtering models. The default is stationary *R*-square.

Saving Model Predictions and Model Specifications

The Save tab allows you to save model predictions as new variables in the active dataset and save model specifications to an external file in XML format.

Save Variables. You can save model predictions, confidence intervals, and residuals as new variables in the active dataset. Each model gives rise to its own set of new variables. New cases are added if the forecast period extends beyond the length of the dependent variable series associated with the model. Unless model parameters are reestimated (**Reestimate from data** on the Models tab), predicted values and confidence limits are only created for the forecast period. Choose to save new variables by selecting the associated Save check box for each. By default, no new variables are saved.

- **Predicted Values.** The model predicted values.
- **Lower Confidence Limits.** Lower confidence limits for the predicted values.
- **Upper Confidence Limits.** Upper confidence limits for the predicted values.
- **Noise Residuals.** The model residuals. When transformations of the dependent variable are performed (for example, natural log), these are the residuals for the transformed series. This choice is only available if model parameters are reestimated (**Reestimate from data** on the Models tab).
- **Variable Name Prefix.** Specify prefixes to be used for new variable names or leave the default prefixes. Variable names consist of the prefix, the name of the associated dependent variable, and a model identifier. The variable name is extended if necessary to avoid variable naming conflicts. The prefix must conform to the rules for valid variable names.

Export Model File Model specifications, containing reestimated parameters and fit statistics, are exported to the specified file in XML format. This option is only available if model parameters are reestimated (**Reestimate from data** on the Models tab).

- **XML File.** Model specifications are saved in an XML file that can be used with IBM SPSS applications.
- **PMML File.** Model specifications are saved in a PMML-compliant XML file that can be used with PMML-compliant applications, including IBM SPSS applications.

Options

The Options tab allows you to specify the handling of missing values, set the confidence interval width, and set the number of lags shown for autocorrelations.

User-Missing Values. These options control the handling of user-missing values.

- **Treat as invalid.** User-missing values are treated like system-missing values.
- **Treat as valid.** User-missing values are treated as valid data.

Missing Value Policy. The following rules apply to the treatment of missing values (includes system-missing values and user-missing values treated as invalid):

- Cases with missing values of a dependent variable that occur within the estimation period are included in the model. The specific handling of the missing value depends on the estimation method.
- For ARIMA models, a warning is issued if a predictor has any missing values within the estimation period. Any models involving the predictor are not reestimated.
- If any independent variable has missing values within the forecast period, the procedure issues a warning and forecasts as far as it can.

Confidence Interval Width (%). Confidence intervals are computed for the model predictions and residual autocorrelations. You can specify any positive value less than 100. By default, a 95% confidence interval is used.

Maximum Number of Lags Shown in ACF and PACF Output. You can set the maximum number of lags shown in tables and plots of autocorrelations and partial autocorrelations. This option is only available if model parameters are reestimated (**Reestimate from data** on the Models tab).

TSAPPLY Command Additional Features

Additional features are available if you paste your selections into a syntax window and edit the resulting TSAPPLY command syntax. The command syntax language allows you to:

- Specify that only a subset of the models in a model file are to be applied to the active dataset (with the DROP and KEEP keywords on the MODEL subcommand).
- Apply models from two or more model files to your data (with the MODEL subcommand). For example, one model file might contain models for series that represent unit sales, and another might contain models for series that represent revenue.

See the *Command Syntax Reference* for complete syntax information.

Chapter 4. Seasonal Decomposition

The Seasonal Decomposition procedure decomposes a series into a seasonal component, a combined trend and cycle component, and an "error" component. The procedure is an implementation of the Census Method I, otherwise known as the ratio-to-moving-average method.

Example. A scientist is interested in analyzing monthly measurements of the ozone level at a particular weather station. The goal is to determine if there is any trend in the data. In order to uncover any real trend, the scientist first needs to account for the variation in readings due to seasonal effects. The Seasonal Decomposition procedure can be used to remove any systematic seasonal variations. The trend analysis is then performed on a seasonally adjusted series.

Statistics. The set of seasonal factors.

Seasonal Decomposition Data Considerations

Data. The variables should be numeric.

Assumptions. The variables should not contain any embedded missing data. At least one periodic date component must be defined.

Estimating Seasonal Factors

1. From the menus choose:

   Analyze > **Forecasting** > **Seasonal Decomposition...**

2. Select one or more variables from the available list and move them into the Variable(s) list. Note that the list includes only numeric variables.

Model Type. The Seasonal Decomposition procedure offers two different approaches for modeling the seasonal factors: multiplicative or additive.

- *Multiplicative*. The seasonal component is a factor by which the seasonally adjusted series is multiplied to yield the original series. In effect, seasonal components that are proportional to the overall level of the series. Observations without seasonal variation have a seasonal component of 1.

- *Additive*. The seasonal adjustments are added to the seasonally adjusted series to obtain the observed values. This adjustment attempts to remove the seasonal effect from a series in order to look at other characteristics of interest that may be "masked" by the seasonal component. In effect, seasonal components that do not depend on the overall level of the series. Observations without seasonal variation have a seasonal component of 0.

Moving Average Weight. The Moving Average Weight options allow you to specify how to treat the series when computing moving averages. These options are available only if the periodicity of the series is even. If the periodicity is odd, all points are weighted equally.

- *All points equal*. Moving averages are calculated with a span equal to the periodicity and with all points weighted equally. This method is always used if the periodicity is odd.

- *Endpoints weighted by .5*. Moving averages for series with even periodicity are calculated with a span equal to the periodicity plus 1 and with the endpoints of the span weighted by 0.5.

Optionally, you can:

- Click **Save** to specify how new variables should be saved.

Seasonal Decomposition Save

Create Variables. Allows you to choose how to treat new variables.

- *Add to file*. The new series created by Seasonal Decomposition are saved as regular variables in your active dataset. Variable names are formed from a three-letter prefix, an underscore, and a number.
- *Replace existing*. The new series created by Seasonal Decomposition are saved as temporary variables in your active dataset. At the same time, any existing temporary variables created by the Forecasting procedures are dropped. Variable names are formed from a three-letter prefix, a pound sign (#), and a number.
- *Do not create*. The new series are not added to the active dataset.

New Variable Names

The Seasonal Decomposition procedure creates four new variables (series), with the following three-letter prefixes, for each series specified:

SAF. *Seasonal adjustment factors.* These values indicate the effect of each period on the level of the series.

SAS. *Seasonally adjusted series.* These are the values obtained after removing the seasonal variation of a series.

STC. *Smoothed trend-cycle components.* These values show the trend and cyclical behavior present in the series.

ERR. *Residual or "error" values.* The values that remain after the seasonal, trend, and cycle components have been removed from the series.

SEASON Command Additional Features

The command syntax language also allows you to:

- Specify any periodicity within the SEASON command rather than select one of the alternatives offered by the Define Dates procedure.

See the *Command Syntax Reference* for complete syntax information.

Chapter 5. Spectral Plots

The Spectral Plots procedure is used to identify periodic behavior in time series. Instead of analyzing the variation from one time point to the next, it analyzes the variation of the series as a whole into periodic components of different frequencies. Smooth series have stronger periodic components at low frequencies; random variation ("white noise") spreads the component strength over all frequencies.

Series that include missing data cannot be analyzed with this procedure.

Example. The rate at which new houses are constructed is an important barometer of the state of the economy. Data for housing starts typically exhibit a strong seasonal component. But are there longer cycles present in the data that analysts need to be aware of when evaluating current figures?

Statistics. Sine and cosine transforms, periodogram value, and spectral density estimate for each frequency or period component. When bivariate analysis is selected: real and imaginary parts of cross-periodogram, cospectral density, quadrature spectrum, gain, squared coherency, and phase spectrum for each frequency or period component.

Plots. For univariate and bivariate analyses: periodogram and spectral density. For bivariate analyses: squared coherency, quadrature spectrum, cross amplitude, cospectral density, phase spectrum, and gain.

Spectral Plots Data Considerations

Data. The variables should be numeric.

Assumptions. The variables should not contain any embedded missing data. The time series to be analyzed should be stationary and any non-zero mean should be subtracted out from the series.

- *Stationary*. A condition that must be met by the time series to which you fit an ARIMA model. Pure MA series will be stationary; however, AR and ARMA series might not be. A stationary series has a constant mean and a constant variance over time.

Obtaining a Spectral Analysis
1. From the menus choose:

   Analysis > **Time Series** > **Spectral Analysis...**
2. Select one or more variables from the available list and move them to the Variable(s) list. Note that the list includes only numeric variables.
3. Select one of the Spectral Window options to choose how to smooth the periodogram in order to obtain a spectral density estimate. Available smoothing options are Tukey-Hamming, Tukey, Parzen, Bartlett, Daniell (Unit), and None.

- *Tukey-Hamming*. The weights are $W_k = .54D_p(2 \pi f_k) + .23D_p (2 \pi f_k + \pi/p) + .23D_p (2 \pi f_k - \pi/p)$, for $k = 0, ..., p$, where p is the integer part of half the span and D_p is the Dirichlet kernel of order p.
- *Tukey*. The weights are $W_k = 0.5D_p(2 \pi f_k) + 0.25D_p (2 \pi f_k + \pi/p) + 0.25D_p(2 \pi f_k - \pi/p)$, for $k = 0, ..., p$, where p is the integer part of half the span and D_p is the Dirichlet kernel of order p.
- *Parzen*. The weights are $W_k = 1/p(2 + \cos(2 \pi f_k)) (F[p/2] (2 \pi f_k))^{**2}$, for $k = 0, ... p$, where p is the integer part of half the span and $F[p/2]$ is the Fejer kernel of order p/2.
- *Bartlett*. The shape of a spectral window for which the weights of the upper half of the window are computed as $W_k = F_p (2*\pi*f_k)$, for $k = 0, ... p$, where p is the integer part of half the span and F_p is the Fejer kernel of order p. The lower half is symmetric with the upper half.
- *Daniell (Unit)*. The shape of a spectral window for which the weights are all equal to 1.

- *None*. No smoothing. If this option is chosen, the spectral density estimate is the same as the periodogram.

Span. The range of consecutive values across which the smoothing is carried out. Generally, an odd integer is used. Larger spans smooth the spectral density plot more than smaller spans.

Center variables. Adjusts the series to have a mean of 0 before calculating the spectrum and to remove the large term that may be associated with the series mean.

Bivariate analysis—first variable with each. If you have selected two or more variables, you can select this option to request bivariate spectral analyses.
- The first variable in the Variable(s) list is treated as the independent variable, and all remaining variables are treated as dependent variables.
- Each series after the first is analyzed with the first series independently of other series named. Univariate analyses of each series are also performed.

Plot. Periodogram and spectral density are available for both univariate and bivariate analyses. All other choices are available only for bivariate analyses.
- *Periodogram*. Unsmoothed plot of spectral amplitude (plotted on a logarithmic scale) against either frequency or period. Low-frequency variation characterizes a smooth series. Variation spread evenly across all frequencies indicates "white noise."
- *Squared coherency*. The product of the gains of the two series.
- *Quadrature spectrum*. The imaginary part of the cross-periodogram, which is a measure of the correlation of the out-of-phase frequency components of two time series. The components are out of phase by pi/2 radians.
- *Cross amplitude*. The square root of the sum of the squared cospectral density and the squared quadrature spectrum.
- *Spectral density*. A periodogram that has been smoothed to remove irregular variation.
- *Cospectral density*. The real part of the cross-periodogram, which is a measure of the correlation of the in-phase frequency components of two time series.
- *Phase spectrum*. A measure of the extent to which each frequency component of one series leads or lags the other.
- *Gain*. The quotient of dividing the cross amplitude by the spectral density for one of the series. Each of the two series has its own gain value.

By frequency. All plots are produced by frequency, ranging from frequency 0 (the constant or mean term) to frequency 0.5 (the term for a cycle of two observations).

By period. All plots are produced by period, ranging from 2 (the term for a cycle of two observations) to a period equal to the number of observations (the constant or mean term). Period is displayed on a logarithmic scale.

SPECTRA Command Additional Features

The command syntax language also allows you to:
- Save computed spectral analysis variables to the active dataset for later use.
- Specify custom weights for the spectral window.
- Produce plots by both frequency and period.
- Print a complete listing of each value shown in the plot.

See the *Command Syntax Reference* for complete syntax information.

Chapter 6. Goodness-of-Fit Measures

This section provides definitions of the goodness-of-fit measures used in time series modeling.

- *Stationary R-squared*. A measure that compares the stationary part of the model to a simple mean model. This measure is preferable to ordinary R-squared when there is a trend or seasonal pattern. Stationary R-squared can be negative with a range of negative infinity to 1. Negative values mean that the model under consideration is worse than the baseline model. Positive values mean that the model under consideration is better than the baseline model.

- *R-squared*. An estimate of the proportion of the total variation in the series that is explained by the model. This measure is most useful when the series is stationary. R-squared can be negative with a range of negative infinity to 1. Negative values mean that the model under consideration is worse than the baseline model. Positive values mean that the model under consideration is better than the baseline model.

- *RMSE*. Root Mean Square Error. The square root of mean square error. A measure of how much a dependent series varies from its model-predicted level, expressed in the same units as the dependent series.

- *MAPE*. Mean Absolute Percentage Error. A measure of how much a dependent series varies from its model-predicted level. It is independent of the units used and can therefore be used to compare series with different units.

- *MAE*. Mean absolute error. Measures how much the series varies from its model-predicted level. MAE is reported in the original series units.

- *MaxAPE*. Maximum Absolute Percentage Error. The largest forecasted error, expressed as a percentage. This measure is useful for imagining a worst-case scenario for your forecasts.

- *MaxAE*. Maximum Absolute Error. The largest forecasted error, expressed in the same units as the dependent series. Like MaxAPE, it is useful for imagining the worst-case scenario for your forecasts. Maximum absolute error and maximum absolute percentage error may occur at different series points--for example, when the absolute error for a large series value is slightly larger than the absolute error for a small series value. In that case, the maximum absolute error will occur at the larger series value and the maximum absolute percentage error will occur at the smaller series value.

- *Normalized BIC*. Normalized Bayesian Information Criterion. A general measure of the overall fit of a model that attempts to account for model complexity. It is a score based upon the mean square error and includes a penalty for the number of parameters in the model and the length of the series. The penalty removes the advantage of models with more parameters, making the statistic easy to compare across different models for the same series.

Chapter 7. Outlier Types

This section provides definitions of the outlier types used in time series modeling.

- *Additive*. An outlier that affects a single observation. For example, a data coding error might be identified as an additive outlier.
- *Level shift*. An outlier that shifts all observations by a constant, starting at a particular series point. A level shift could result from a change in policy.
- *Innovational*. An outlier that acts as an addition to the noise term at a particular series point. For stationary series, an innovational outlier affects several observations. For nonstationary series, it may affect every observation starting at a particular series point.
- *Transient*. An outlier whose impact decays exponentially to 0.
- *Seasonal additive*. An outlier that affects a particular observation and all subsequent observations separated from it by one or more seasonal periods. All such observations are affected equally. A seasonal additive outlier might occur if, beginning in a certain year, sales are higher every January.
- *Local trend*. An outlier that starts a local trend at a particular series point.
- *Additive patch*. A group of two or more consecutive additive outliers. Selecting this outlier type results in the detection of individual additive outliers in addition to groups of them.

Chapter 8. Guide to ACF/PACF Plots

The plots shown here are those of pure or theoretical ARIMA processes. Here are some general guidelines for identifying the process:

- Nonstationary series have an ACF that remains significant for half a dozen or more lags, rather than quickly declining to 0. You must difference such a series until it is stationary before you can identify the process.
- Autoregressive processes have an exponentially declining ACF and spikes in the first one or more lags of the PACF. The number of spikes indicates the order of the autoregression.
- Moving average processes have spikes in the first one or more lags of the ACF and an exponentially declining PACF. The number of spikes indicates the order of the moving average.
- Mixed (ARMA) processes typically show exponential declines in both the ACF and the PACF.

At the identification stage, you do not need to worry about the sign of the ACF or PACF, or about the speed with which an exponentially declining ACF or PACF approaches 0. These depend upon the sign and actual value of the AR and MA coefficients. In some instances, an exponentially declining ACF alternates between positive and negative values.

ACF and PACF plots from real data are never as clean as the plots shown here. You must learn to pick out what is essential in any given plot. Always check the ACF and PACF of the residuals, in case your identification is wrong. Bear in mind that:

- Seasonal processes show these patterns at the seasonal lags (the multiples of the seasonal period).
- You are entitled to treat nonsignificant values as 0. That is, you can ignore values that lie within the confidence intervals on the plots. You do not have to ignore them, however, particularly if they continue the pattern of the statistically significant values.
- An occasional autocorrelation will be statistically significant by chance alone. You can ignore a statistically significant autocorrelation if it is isolated, preferably at a high lag, and if it does not occur at a seasonal lag.

Consult any text on ARIMA analysis for a more complete discussion of ACF and PACF plots.

Table 2. ARIMA(0,0,1), q>0
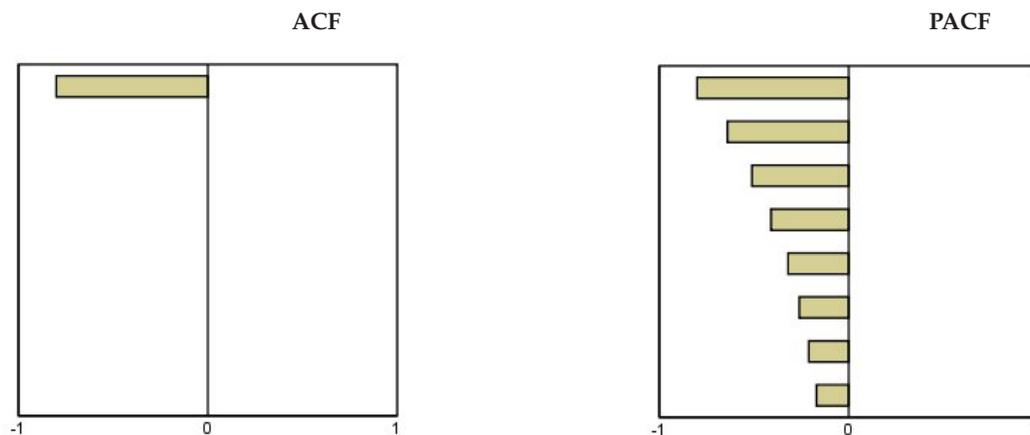
Table 3. ARIMA(0,0,1), q<0
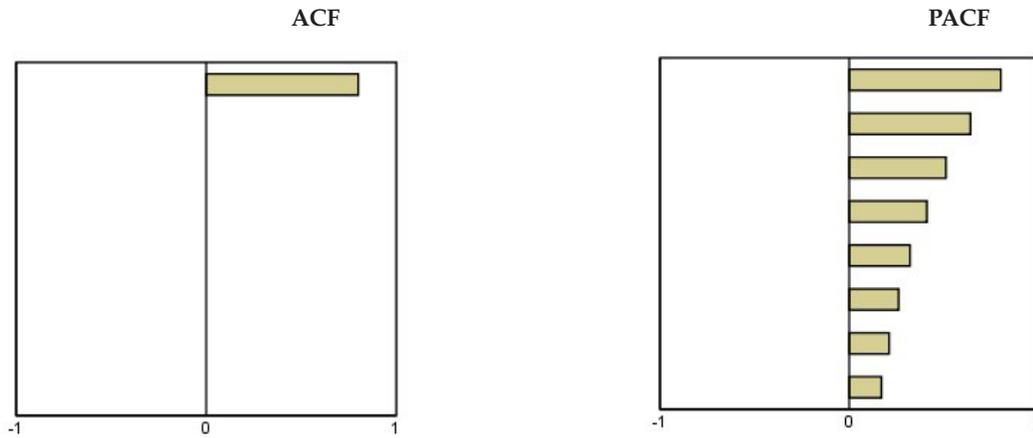
ACF	PACF

ARIMA(0,0,2), $\theta_1 \theta_2 > 0$

ACF	PACF

Table 4. ARIMA(1,0,0), f>0

ACF	PACF

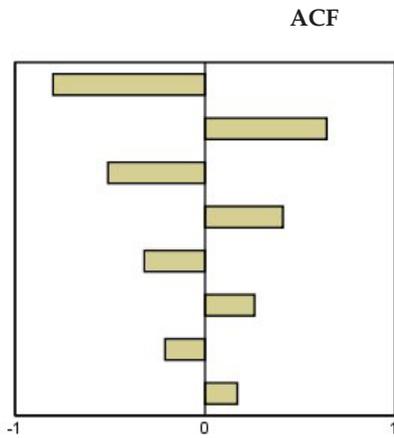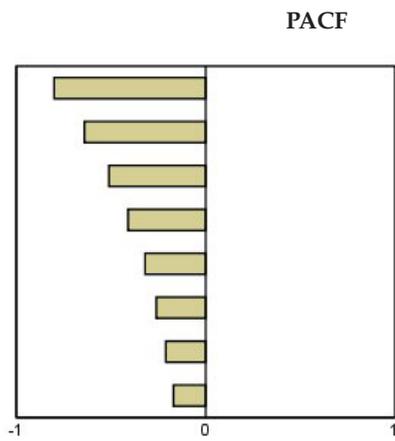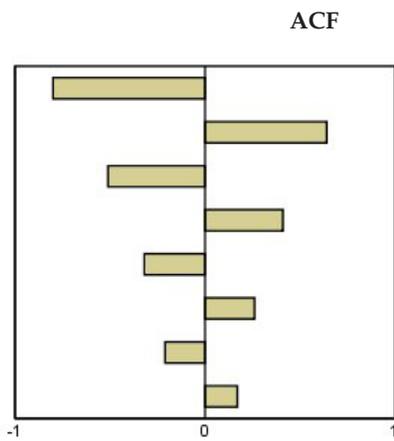*Table 5. ARIMA(1,0,0), f<0*

ACF   PACF

ARIMA(1,0,1), <0, >0

ACF   PACF

ARIMA(2,0,0), $_{1\ 2}$>0

ACF   PACF
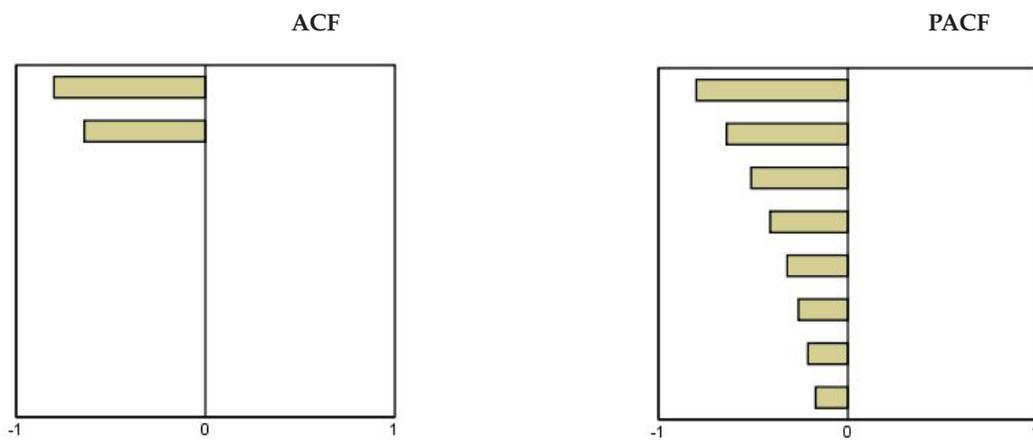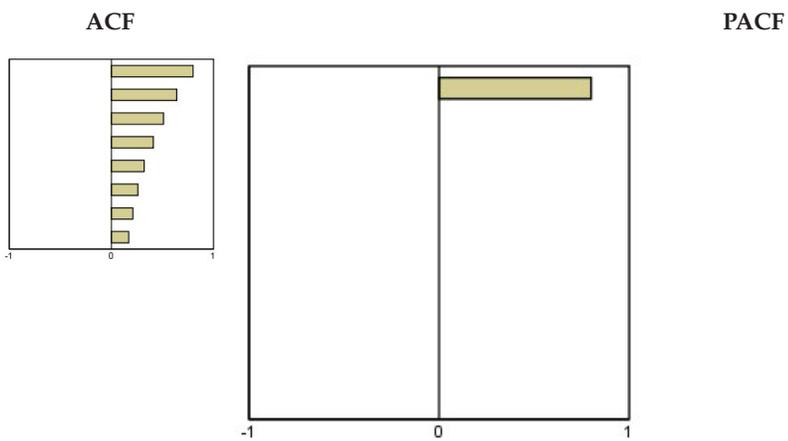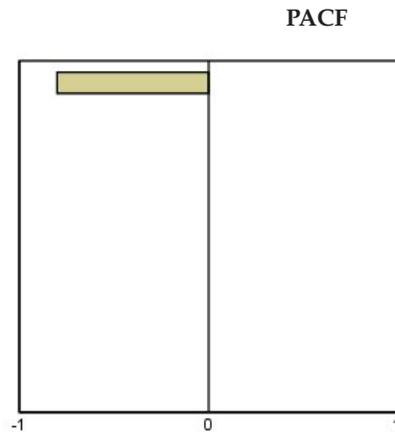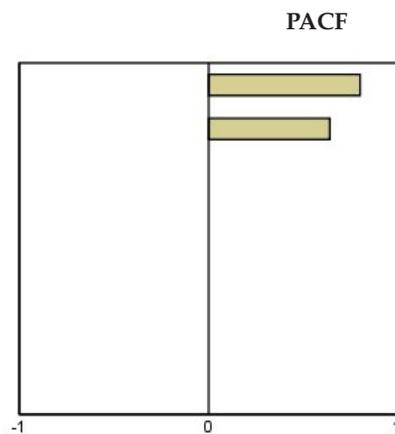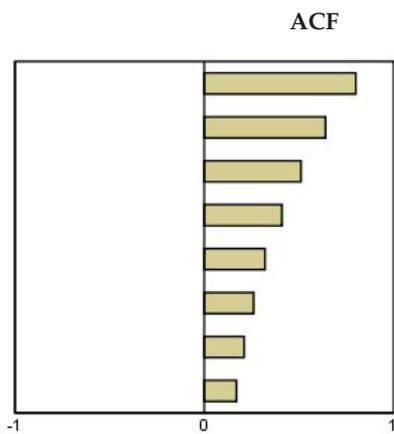
Table 6. ARIMA(0,1,0) (integrated series)

ACF
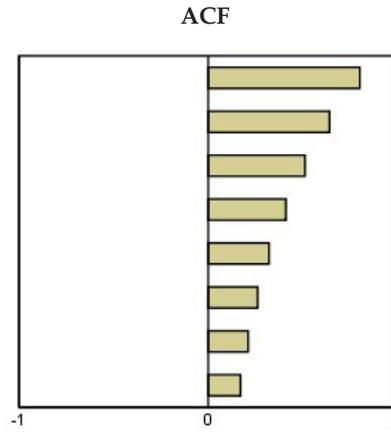
Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs.

© Copyright IBM Corp. _enter the year or years_. All rights reserved.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Index

A

ACF
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
    plots for pure ARIMA processes   29
additive outlier   27
    in Time Series Modeler   5, 9
additive patch outlier   27
    in Time Series Modeler   5, 9
Apply Time Series Models   15
    best- and poorest-fitting models   18
    Box-Ljung statistic   17
    confidence intervals   18, 20
    estimation period   15
    fit values   18
    forecast period   15
    forecasts   17, 18
    goodness-of-fit statistics   17, 18
    missing values   20
    model parameters   17
    new variable names   19
    reestimate model parameters   15
    residual autocorrelation function   17, 18
    residual partial autocorrelation function   17, 18
    saving predictions   19
    saving reestimated models in XML   19
    statistics across all models   17, 18
ARIMA models   3
    outliers   9
    transfer functions   8
autocorrelation function
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
    plots for pure ARIMA processes   29

B

Box-Ljung statistic
    in Apply Time Series Models   17
    in Time Series Modeler   9
Brown's exponential smoothing model   6

C

confidence intervals
    in Apply Time Series Models   18, 20
    in Time Series Modeler   10, 12

D

damped exponential smoothing model   6

E

estimation period   2
    in Apply Time Series Models   15

estimation period *(continued)*
    in Time Series Modeler   3
events   5
    in Time Series Modeler   5
Expert Modeler   3
    limiting the model space   5
    outliers   5
exponential smoothing models   3, 6

F

fit values
    in Apply Time Series Models   18
    in Time Series Modeler   10
forecast period
    in Apply Time Series Models   15
    in Time Series Modeler   3, 12
forecasts
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10

G

goodness of fit
    definitions   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10

H

harmonic analysis   23
historical data
    in Apply Time Series Models   18
    in Time Series Modeler   10
historical period   2
holdout cases   2
Holt's exponential smoothing model   6

I

innovational outlier   27
    in Time Series Modeler   5, 9

L

level shift outlier   27
    in Time Series Modeler   5, 9
local trend outlier   27
    in Time Series Modeler   5, 9
log transformation
    in Time Series Modeler   6, 7, 8

M

MAE   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
MAPE   25

MAPE *(continued)*
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
MaxAE   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
MaxAPE   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
maximum absolute error   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
maximum absolute percentage error   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
mean absolute error   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
mean absolute percentage error   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
missing values
    in Apply Time Series Models   20
    in Time Series Modeler   12
model names
    in Time Series Modeler   12
model parameters
    in Apply Time Series Models   17
    in Time Series Modeler   9
models
    ARIMA   3
    Expert Modeler   3
    exponential smoothing   3, 6

N

natural log transformation
    in Time Series Modeler   6, 7, 8
normalized BIC (Bayesian information criterion)   25
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10

O

outliers
    ARIMA models   9
    definitions   27
    Expert Modeler   5

P

PACF
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
    plots for pure ARIMA processes   29
partial autocorrelation function
    in Apply Time Series Models   17, 18
    in Time Series Modeler   9, 10
    plots for pure ARIMA processes   29

IBM ®

Printed in USA