

IBM SPSS Forecasting 32



Note

Before using this information and the product it supports, read the information in [“Notices” on page 53.](#)

Product Information

This edition applies to version 32, release 0, modification 0 of IBM® SPSS® Statistics and to all subsequent releases and modifications until otherwise indicated in new editions.

© **Copyright International Business Machines Corporation .**

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Chapter 1. Forecasting.....	1
Introduction to Time Series.....	1
Time Series Data.....	1
Data Transformations.....	2
Estimation and Validation Periods.....	2
Building Models and Producing Forecasts.....	2
Time Series Modeler.....	3
Specifying Options for the Expert Modeler.....	5
Custom Exponential Smoothing Models.....	6
Custom ARIMA Models.....	7
Output.....	9
Saving Model Predictions and Model Specifications.....	11
Options.....	12
TSMODEL Command Additional Features.....	13
Apply Time Series Models.....	13
Output.....	15
Saving Model Predictions and Model Specifications.....	17
Options.....	17
TSAPPLY Command Additional Features.....	18
Seasonal Decomposition.....	18
Seasonal Decomposition Save.....	19
SEASON Command Additional Features.....	19
Spectral Plots.....	20
SPECTRA Command Additional Features.....	21
Temporal Causal Models.....	22
To Obtain a Temporal Causal Model.....	22
Time Series to Model	23
Observations	24
Time Interval for Analysis	25
Aggregation and Distribution	25
Missing Values	26
General Data Options	27
General Build Options	27
Series to Display	27
Output Options	28
Estimation Period	30
Forecast	30
Save	31
Interactive Output	31
Applying Temporal Causal Models.....	32
Applying Temporal Causal Models.....	32
Temporal Causal Model Forecasting.....	32
Temporal Causal Model Scenarios.....	37
Vector Autoregressive Models.....	43
Variables.....	44
Statistics.....	45
Plots.....	45
Save.....	46
Options.....	46
Goodness-of-Fit Measures.....	47
Outlier Types.....	48

Guide to ACF/PACF Plots.....	48
Notices.....	53
Trademarks.....	54
Index.....	55

Chapter 1. Forecasting

The following forecasting features are included in SPSS Statistics Professional Edition or the Forecasting option.

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

Column-based data

Each time series field contains the data for a single time series. This structure is the traditional structure of time series data, as used by the Time Series Modeler procedure, the Seasonal Decomposition procedure, and the Spectral Plots procedure. 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 that contains 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.

Multidimensional data

For multidimensional data, each time series field contains the data for multiple time series. Separate time series, within a particular field, are then identified by a set of values of categorical fields referred to as *dimension* fields.

For example, sales data for different regions and brands might be stored in a single *sales* field, so that the dimensions in this case are *region* and *brand*. Each combination of *region* and *brand* identifies a

particular time series for *sales*. For example, in the following table, the records that have 'north' for *region* and 'brandX' for *brand* define a single time series.

<i>Table 2. Multidimensional data</i>			
date	region	brand	sales
01/01/2014	north	brandX	82350
01/01/2014	north	brandY	86380
01/01/2014	south	brandX	91375
01/01/2014	south	brandY	70320
01/02/2014	north	brandX	83275
01/02/2014	north	brandY	85260
01/02/2014	south	brandX	94760
01/02/2014	south	brandY	69870

Note: Data that is imported from OLAP cubes, such as from IBM Cognos[®] TM1[®], is represented as multidimensional data.

Data Transformations

A number of data transformation procedures that are provided in the Core system are useful in time series analysis. These transformations apply only to column-based data, where each time series field contains the data for a single time series.

- The Define Dates procedure (on the Data menu) generates date variables that are 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 about 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. 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 the following procedures for accomplishing the tasks of creating models and producing forecasts:

- The “[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 “[Apply Time Series Models](#)” on page 13 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.
- The “[Temporal Causal Models](#)” on page 22 procedure builds autoregressive time series models for each target and automatically determines the best inputs that have a causal relationship with the target. The procedure produces interactive output that you can use to explore the causal relationships. The procedure can also generate forecasts, detect outliers, and determine the series that most likely causes an outlier.
- The “[Temporal Causal Model Forecasting](#)” on page 32 procedure applies a temporal causal model to the active dataset. You can use this procedure to obtain forecasts for series for which more current data is available, without rebuilding your models. You can also use it to determine series that most likely cause outliers that were detected by the Temporal Causal Models procedure.

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 Traditional Models...

Note: The fields highlighted in red are required. The **Paste** and **OK** buttons are enabled after you enter valid values in all required fields.

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

Custom Exponential Smoothing Models

Model Type. Exponential smoothing models¹ 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.

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

- **None.** No transformation is performed.
- **Square root.** Square root transformation.
- **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-Jenkins² 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.

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.

- **Square root.** Square root transformation.
- **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 outliers³: 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

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
- Mean absolute percentage error
- Mean absolute error
- Maximum absolute percentage error
- Maximum absolute 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.

- Normalized BIC

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)

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.

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 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 Traditional 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

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)

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.

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.

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...

Note: The fields highlighted in red are required. The **Paste** and **OK** buttons are enabled after you enter valid values in all required fields.

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 $k = 0, \dots, p$, 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 $k = 0, \dots, p$, 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 $k = 0, \dots, p$, 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 $k = 0, \dots, p$, 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.
- *Power spectral density.* Power spectral density (or power density or power density spectrum or power spectrum) of the signal describes the power present in the signal as a function of frequency, per unit frequency.
- *Coherence.* Coherence is a statistic that can be used to examine the relation between two signals or data sets in signal processing, . It is commonly used to estimate the power transfer between the input and output of a linear system.
- *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.

Temporal Causal Models

Temporal causal modeling attempts to discover key causal relationships in time series data. In temporal causal modeling, you specify a set of target series and a set of candidate inputs to those targets. The procedure then builds an autoregressive time series model for each target and includes only those inputs that have a causal relationship with the target. This approach differs from traditional time series modeling where you must explicitly specify the predictors for a target series. Since temporal causal modeling typically involves building models for multiple related time series, the result is referred to as a *model system*.

In the context of temporal causal modeling, the term *causal* refers to Granger causality. A time series X is said to "Granger cause" another time series Y if regressing for Y in terms of past values of both X and Y results in a better model for Y than regressing only on past values of Y.

Examples

Business decision makers can use temporal causal modeling to uncover causal relationships within a large set of time-based metrics that describe the business. The analysis might reveal a few controllable inputs, which have the largest impact on key performance indicators.

Managers of large IT systems can use temporal causal modeling to detect anomalies in a large set of interrelated operational metrics. The causal model then allows going beyond anomaly detection and discovering the most likely root causes of the anomalies.

Field requirements

There must be at least one target. By default, fields with a predefined role of None are not used.

Data structure

Temporal causal modeling supports two types of data structures.

Column-based data

For column-based data, each time series field contains the data for a single time series. This structure is the traditional structure of time series data, as used by the Time Series Modeler.

Multidimensional data

For multidimensional data, each time series field contains the data for multiple time series. Separate time series, within a particular field, are then identified by a set of values of categorical fields referred to as *dimension* fields. For example, sales data for two different sales channels (retail and web) might be stored in a single *sales* field. A dimension field that is named *channel*, with values 'retail' and 'web', identifies the records that are associated with each of the two sales channels.

To Obtain a Temporal Causal Model

This feature requires the Statistics Forecasting option.

From the menus, choose:

Analyze > Forecasting > Create Temporal Causal Models...

1. If the observations are defined by a date/time field, then specify the field.
2. If the data are multidimensional, then specify the dimension fields that identify the time series.
 - The specified order of the dimension fields defines the order in which they appear in all subsequent dialogs and output. Use the up and down arrow buttons to reorder the dimension fields.
 - You can specify subsets of dimension values that limit the analysis to particular values of the dimension fields. For example, if you have dimensions for region and brand, you can limit the analysis to a specific region. Dimension subsets apply to all metric fields used in the analysis.
 - You can also customize the analysis by specifying dimension values by role at the metric field level. For example, if you have a dimension for sales channel (with values 'retail' and 'web') and metrics

for sales and advertising in those channels, then you can specify web advertising as an input for both retail and web sales. By default, this type of customization is enabled and limited to selecting from up to a specified number of distinct values (250 by default) of each dimension field.

3. Click **Continue**.

Note: Steps 1, 2 and 3 do not apply if the active dataset has a date specification. Date specifications are created from the Define Dates dialog or the DATE command.

4. Click **Fields** to specify the time series to include in the model and to specify how the observations are defined. At least one field must be specified as either a target or as both an input and a target.
5. Click **Data Specifications** to specify optional settings that include the time interval for analysis, aggregation and distribution settings, and handling of missing values.
6. Click **Build Options** to define the estimation period, specify the content of the output, and specify build settings such as the maximum number of inputs per target.
7. Click **Model Options** to request forecasts, save predictions, and export the model system to an external file.
8. Click **Run** to run the procedure.

Time Series to Model

On the Fields tab, use the **Time Series** settings to specify the series to include in the model system.

For column-based data, the term *series* has the same meaning as the term *field*. For multidimensional data, fields that contain time series are referred to as *metric* fields. A time series, for multidimensional data, is defined by a metric field and a value for each of the dimension fields. The following considerations apply to both column-based and multidimensional data.

- Series that are specified as candidate inputs or as both target and input are considered for inclusion in the model for each target. The model for each target always includes lagged values of the target itself.
- Series that are specified as forced inputs are always included in the model for each target.
- At least one series must be specified as either a target or as both target and input.
- When **Use predefined roles** is selected, fields that have a role of Input are set as candidate inputs. No predefined role maps to a forced input.

Multidimensional data

For multidimensional data, you specify metric fields and associated roles in a grid, where each row in the grid specifies a single metric and role. By default, the model system includes series for all combinations of the dimension fields for each row in the grid. For example, if there are dimensions for *region* and *brand* then, by default, specifying the metric *sales* as a target means that there is a separate sales target series for each combination of *region* and *brand*.

For each row in the grid, you can customize the set of values for any of the dimension fields by clicking the ellipsis button for a dimension. This action opens the Select Dimension Values subdialog. You can also add, delete, or copy grid rows.

The **Series Count** column displays the number of sets of dimension values that are currently specified for the associated metric. The displayed value can be larger than the actual number of series (one series per set). This condition occurs when some of the specified combinations of dimension values do not correspond to series contained by the associated metric.

Select Dimension Values

For multidimensional data, you can customize the analysis by specifying which dimension values apply to a particular metric field with a particular role. For example, if *sales* is a metric field and *channel* is a dimension with values 'retail' and 'web', you can specify that 'web' sales is an input and 'retail' sales is a target. You can also specify dimension subsets that apply to all metric fields used in the analysis. For

example, if *region* is a dimension field that indicates geographical region, then you can limit the analysis to particular regions.

All values

Specifies that all values of the current dimension field are included. This option is the default.

Select values to include or exclude

Use this option to specify the set of values for the current dimension field. When **Include** is selected for the **Mode**, only values that are specified in the **Selected values** list are included. When **Exclude** is selected for the **Mode**, all values other than the values that are specified in the **Selected values** list are included.

You can filter the set of values from which to choose. Values that meet the filter condition appear in the **Matched** tab and values that do not meet the filter condition appear in the **Unmatched** tab of the **Unselected values** list. The **All** tab lists all unselected values, regardless of any filter condition.

- You can use asterisks (*) to indicate wildcard characters when you specify a filter.
- To clear the current filter, specify an empty value for the search term on the Filter Displayed Values dialog.

Observations

On the Fields tab, use the **Observations** settings to specify the fields that define the observations.

Note: If the active dataset has a date specification, then the observations are defined by the date specification and cannot be modified in the temporal causal modeling procedure. Date specifications are created from the Define Dates dialog or the DATE command.

Observations that are defined by date/times

You can specify that the observations are defined by a field with a date, time, or datetime format, or by a string field that represents a date/time. String fields can represent a date in YYYY-MM-DD format, a time in HH:MM:SS format, or a datetime in YYYY-MM-DD HH:MM:SS format. Leading zeros can be omitted, in the string representation. For example, the string 2014-9-01 is equivalent to 2014-09-01.

In addition to the field that defines the observations, select the appropriate time interval that describes the observations. Depending on the specified time interval, you can also specify other settings, such as the interval between observations (increment) or the number of days per week. The following considerations apply to the time interval:

- Use the value **Irregular** when the observations are irregularly spaced in time, such as the time at which a sales order is processed. When **Irregular** is selected, you must specify the time interval that is used for the analysis, from the **Time Interval** settings on the Data Specifications tab.
- When the observations represent a date and time and the time interval is hours, minutes, or seconds, then use **Hours per day**, **Minutes per day**, or **Seconds per day**. When the observations represent a time (duration) without reference to a date and the time interval is hours, minutes, or seconds, then use **Hours (non-periodic)**, **Minutes (non-periodic)**, or **Seconds (non-periodic)**.
- Based on the selected time interval, the procedure can detect missing observations. Detecting missing observations is necessary since the procedure assumes that all observations are equally spaced in time and that there are no missing observations. For example, if the time interval is Days and the date 2014-10-27 is followed by 2014-10-29, then there is a missing observation for 2014-10-28. Values are imputed for any missing observations. Settings for handling missing values can be specified from the Data Specifications tab.
- The specified time interval allows the procedure to detect multiple observations in the same time interval that need to be aggregated together and to align observations on an interval boundary, such as the first of the month, to ensure that the observations are equally spaced. For example, if the time interval is Months, then multiple dates in the same month are aggregated together. This type of aggregation is referred to as *grouping*. By default, observations are summed when grouped. You can specify a different method for grouping, such as the mean of the observations, from the **Aggregation and Distribution** settings on the Data Specifications tab.

- For some time intervals, the additional settings can define breaks in the normal equally spaced intervals. For example, if the time interval is Days, but only weekdays are valid, you can specify that there are five days in a week, and the week begins on Monday.

Observations that are defined by periods or cyclic periods

Observations can be defined by one or more integer fields that represent periods or repeating cycles of periods, up to an arbitrary number of cycle levels. With this structure, you can describe series of observations that don't fit one of the standard time intervals. For example, a fiscal year with only 10 months can be described with a cycle field that represents years and a period field that represents months, where the length of one cycle is 10.

Fields that specify cyclic periods define a hierarchy of periodic levels, where the lowest level is defined by the **Period** field. The next highest level is specified by a cycle field whose level is 1, followed by a cycle field whose level is 2, and so on. Field values for each level, except the highest, must be periodic with respect to the next highest level. Values for the highest level cannot be periodic. For example, in the case of the 10-month fiscal year, months are periodic within years and years are not periodic.

- The length of a cycle at a particular level is the periodicity of the next lowest level. For the fiscal year example, there is only one cycle level and the cycle length is 10 since the next lowest level represents months and there are 10 months in the specified fiscal year.
- Specify the starting value for any periodic field that does not start from 1. This setting is necessary for detecting missing values. For example, if a periodic field starts from 2 but the starting value is specified as 1, then the procedure assumes that there is a missing value for the first period in each cycle of that field.

Observations that are defined by record order

For column-based data, you can specify that the observations are defined by record order so that the first record represents the first observation, the second record represents the second observation, and so on. It is then assumed that the records represent observations that are equally spaced in time.

Time Interval for Analysis

The time interval that is used for the analysis can differ from the time interval of the observations. For example, if the time interval of the observations is Days, you might choose Months for the time interval for analysis. The data are then aggregated from daily to monthly data before the model is built. You can also choose to distribute the data from a longer to a shorter time interval. For example, if the observations are quarterly then you can distribute the data from quarterly to monthly data.

The available choices for the time interval at which the analysis is done depend on how the observations are defined and the time interval of those observations. In particular, when the observations are defined by cyclic periods or when a date specification is defined for the active dataset, then only aggregation is supported. In that case, the time interval of the analysis must be greater than or equal to the time interval of the observations.

The time interval for the analysis is specified from the **Time Interval** settings on the Data Specifications tab. The method by which the data are aggregated or distributed is specified from the **Aggregation and Distribution** settings on the Data Specifications tab.

Aggregation and Distribution

Aggregation functions

When the time interval that is used for the analysis is longer than the time interval of the observations, the input data are aggregated. For example, aggregation is done when the time interval of the observations is Days and the time interval for analysis is Months. The following aggregation functions are available: mean, sum, mode, min, or max.

Distribution functions

When the time interval that is used for the analysis is shorter than the time interval of the observations, the input data are distributed. For example, distribution is done when the time interval

of the observations is Quarters and the time interval for analysis is Months. The following distribution functions are available: mean or sum.

Grouping functions

Grouping is applied when observations are defined by date/times and multiple observations occur in the same time interval. For example, if the time interval of the observations is Months, then multiple dates in the same month are grouped and associated with the month in which they occur. The following grouping functions are available: mean, sum, mode, min, or max. Grouping is always done when the observations are defined by date/times and the time interval of the observations is specified as Irregular.

Note: Although grouping is a form of aggregation, it is done before any handling of missing values whereas formal aggregation is done after any missing values are handled. When the time interval of the observations is specified as Irregular, aggregation is done only with the grouping function.

Aggregate cross-day observations to previous day

Specifies whether observations with times that cross a day boundary are aggregated to the values for the previous day. For example, for hourly observations with an eight-hour day that starts at 20:00, this setting specifies whether observations between 00:00 and 04:00 are included in the aggregated results for the previous day. This setting applies only if the time interval of the observations is Hours per day, Minutes per day or Seconds per day and the time interval for analysis is Days.

Custom settings for specified fields

You can specify aggregation, distribution, and grouping functions on a field by field basis. These settings override the default settings for the aggregation, distribution, and grouping functions.

Missing Values

Missing values in the input data are replaced with an imputed value. The following replacement methods are available:

Linear interpolation

Replaces missing values by using a linear interpolation. The last valid value before the missing value and the first valid value after the missing value are used for the interpolation. If the first or last observation in the series has a missing value, then the two nearest non-missing values at the beginning or end of the series are used.

Series mean

Replaces missing values with the mean for the entire series.

Mean of nearby points

Replaces missing values with the mean of valid surrounding values. The span of nearby points is the number of valid values before and after the missing value that are used to compute the mean.

Median of nearby points

Replaces missing values with the median of valid surrounding values. The span of nearby points is the number of valid values before and after the missing value that are used to compute the median.

Linear trend

This option uses all non-missing observations in the series to fit a simple linear regression model, which is then used to impute the missing values.

Other settings:

Maximum percentage of missing values (%)

Specifies the maximum percentage of missing values that are allowed for any series. Series with more missing values than the specified maximum are excluded from the analysis.

User-missing values

This option specifies whether user-missing values are treated as valid data and therefore included in the series. By default, user-missing values are excluded and are treated like system-missing values, which are then imputed.

General Data Options

Maximum number of distinct values per dimension field

This setting applies to multidimensional data and specifies the maximum number of distinct values that are allowed for any one dimension field. By default, this limit is set to 10000 but it can be increased to an arbitrarily large number.

General Build Options

Confidence interval width (%)

This setting controls the confidence intervals for both forecasts and model parameters. You can specify any positive value less than 100. By default, a 95% confidence interval is used.

Maximum number of inputs for each target

This setting specifies the maximum number of inputs that are allowed in the model for each target. You can specify an integer in the range 1 - 20. The model for each target always includes lagged values of itself, so setting this value to 1 specifies that the only input is the target itself.

Model tolerance

This setting controls the iterative process that is used for determining the best set of inputs for each target. You can specify any value that is greater than zero. The default is 0.001.

Outlier threshold (%)

An observation is flagged as an outlier if the probability, as calculated from the model, that it is an outlier exceeds this threshold. You can specify a value in the range 50 - 100.

Number of Lags for Each Input

This setting specifies the number of lag terms for each input in the model for each target. By default, the number of lag terms is automatically determined from the time interval that is used for the analysis. For example, if the time interval is months (with an increment of one month) then the number of lags is 12. Optionally, you can explicitly specify the number of lags. The specified value must be an integer in the range 1 - 20.

Series to Display

These options specify the series (targets or inputs) for which output is displayed. The content of the output for the specified series is determined by the **Output Options** settings.

Display targets associated with best-fitting models

By default, output is displayed for the targets that are associated with the 10 best-fitting models, as determined by the R square value. You can specify a different fixed number of best-fitting models or you can specify a percentage of best-fitting models. You can also choose from the following goodness of fit measures:

R square

Goodness-of-fit measure of a linear model, sometimes called the coefficient of determination. It is the proportion of variation in the target variable explained by the model. It ranges in value from 0 to 1. Small values indicate that the model does not fit the data well.

Root mean square percentage error

A measure of how much the model-predicted values differ from the observed values of the series. It is independent of the units that are used and can therefore be used to compare series with different units.

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.

BIC

Bayesian Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The BIC also "penalizes" overparameterized models (complex models with a large number of inputs, for example), but more strictly than the AIC.

AIC

Akaike Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The AIC "penalizes" overparameterized models (complex models with a large number of inputs, for example).

Specify Individual Series

You can specify individual series for which you want output.

- For column-based data, you specify the fields that contain the series that you want. The order of the specified fields defines the order in which they appear in the output.
- For multidimensional data, you specify a particular series by adding an entry to the grid for the metric field that contains the series. You then specify the values of the dimension fields that define the series.
 - You can enter the value for each dimension field directly into the grid or you can select from the list of available dimension values. To select from the list of available dimension values, click the ellipsis button in the cell for the dimension that you want. This action opens the Select Dimension Value subdialog.
 - You can search the list of dimension values, on the Select Dimension Value subdialog, by clicking the binoculars icon and specifying a search term. Spaces are treated as part of the search term. Asterisks (*) in the search term do not indicate wildcard characters.
 - The order of the series in the grid defines the order in which they appear in the output.

For both column-based data and multidimensional data, output is limited to 30 series. This limit includes individual series (inputs or targets) that you specify and targets that are associated with best-fitting models. Individually specified series take precedence over targets that are associated with best-fitting models.

Output Options

These options specify the content of the output. Options in the **Output for targets** group generate output for the targets that are associated with the best-fitting models on the **Series to Display** settings. Options in the **Output for series** group generate output for the individual series that are specified on the **Series to Display** settings.

Overall model system

Displays a graphical representation of the causal relations between series in the model system. Tables of both model fit statistics and outliers for the displayed targets are included as part of the output item. When this option is selected in the **Output for series** group, a separate output item is created for each individual series that is specified on the **Series to Display** settings.

Causal relations between series have an associated significance level, where a smaller significance level indicates a more significant connection. You can choose to hide relations with a significance level that is greater than a specified value.

Model fit statistics and outliers

Tables of model fit statistics and outliers for the target series that are selected for display. These tables contain the same information as the tables in the Overall Model System visualization. These tables support all standard features for pivoting and editing tables.

Model effects and model parameters

Tables of model effects tests and model parameters for the target series that are selected for display. Model effects tests include the F statistic and associated significance value for each input included in the model.

Impact diagram

Displays a graphical representation of the causal relations between a series of interest and other series that it affects or that affect it. Series that affect the series of interest are referred to as *causes*. Selecting **Effects** generates an impact diagram that is initialized to display effects. Selecting **Causes** generates an impact diagram that is initialized to display causes. Selecting **Both causes and effects**

generates two separate impact diagrams, one that is initialized to causes and one that is initialized to effects. You can interactively toggle between causes and effects in the output item that displays the impact diagram.

You can specify the number of levels of causes or effects to display, where the first level is just the series of interest. Each additional level shows more indirect causes or effects of the series of interest. For example, the third level in the display of effects consists of the series that contain series in the second level as a direct input. Series in the third level are then indirectly affected by the series of interest since the series of interest is a direct input to the series in the second level.

Series plot

Plots of observed and predicted values for the target series that are selected for display. When forecasts are requested, the plot also shows the forecasted values and the confidence intervals for the forecasts.

Residuals plot

Plots of the model residuals for the target series that are selected for display.

Top inputs

Plots of each displayed target, over time, along with the top 3 inputs for the target. The top inputs are the inputs with the lowest significance value. To accommodate different scales for the inputs and target, the y axis represents the z score for each series.

Forecast table

Tables of forecasted values and confidence intervals of those forecasts for the target series that are selected for display.

Outlier root cause analysis

Determines which series are most likely to be the cause of each outlier in a series of interest. Outlier root cause analysis is done for each target series that is included in the list of individual series on the **Series to Display** settings.

Output

Interactive outliers table and chart

Table and chart of outliers and root causes of those outliers for each series of interest. The table contains a single row for each outlier. The chart is an impact diagram. Selecting a row in the table highlights the path, in the impact diagram, from the series of interest to the series that most likely causes the associated outlier.

Pivot table of outliers

Table of outliers and root causes of those outliers for each series of interest. This table contains the same information as the table in the interactive display. This table supports all standard features for pivoting and editing tables.

Causal levels

You can specify the number of levels to include in the search for the root causes. The concept of levels that is used here is the same as described for impact diagrams.

Model fit across all models

Histogram of model fit for all models and for selected fit statistics. The following fit statistics are available:

R square

Goodness-of-fit measure of a linear model, sometimes called the coefficient of determination. It is the proportion of variation in the target variable explained by the model. It ranges in value from 0 to 1. Small values indicate that the model does not fit the data well.

Root mean square percentage error

A measure of how much the model-predicted values differ from the observed values of the series. It is independent of the units that are used and can therefore be used to compare series with different units.

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.

BIC

Bayesian Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The BIC also "penalizes" overparameterized models (complex models with a large number of inputs, for example), but more strictly than the AIC.

AIC

Akaike Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The AIC "penalizes" overparameterized models (complex models with a large number of inputs, for example).

Outliers over time

Bar chart of the number of outliers, across all targets, for each time interval in the estimation period.

Series transformations

Table of any transformations that were applied to the series in the model system. The possible transformations are missing value imputation, aggregation, and distribution.

Estimation Period

By default, the estimation period starts at the time of the earliest observation and ends at the time of the latest observation across all series.

By start and end times

You can specify both the start and end of the estimation period or you can specify just the start or just the end. If you omit the start or the end of the estimation period, the default value is used.

- If the observations are defined by a date/time field, then enter values for start and end in the same format that is used for the date/time field.
- For observations that are defined by cyclic periods, specify a value for each of the cyclic periods fields. Each field is displayed in a separate column.
- If there is a date specification in effect for the active dataset, then you must specify a value for each component (such as Month) of the date specification. Each component is displayed in a separate column.
- When the observations are defined by record order, the start and end of the estimation period are defined by the row number (as displayed in the Data Editor) of the relevant case.

By latest or earliest time intervals

Defines the estimation period as a specified number of time intervals that start at the earliest time interval or end at the latest time interval in the data, with an optional offset. In this context, the time interval refers to the time interval of the analysis. For example, assume that the observations are monthly but the time interval of the analysis is quarters. Specifying **Latest** and a value of 24 for the **Number of time intervals** means the latest 24 quarters.

Optionally, you can exclude a specified number of time intervals. For example, specifying the latest 24 time intervals and 1 for the number to exclude means that the estimation period consists of the 24 intervals that precede the last one.

Forecast

The option to **Extend records into the future** sets the number of time intervals to forecast beyond the end of the estimation period. The time interval in this case is the time interval of the analysis, which is specified on the Data Specifications tab. When forecasts are requested, autoregressive models are automatically built for any input series that are not also targets. These models are then used to generate values for those input series in the forecast period.

Save

Destination Options

You can save both transformations of the data (such as aggregation or imputation of missing values) and new variables (specified on the **Save Targets** settings) to an IBM SPSS Statistics data file or a new dataset in the current session. Date/time values in the saved data are aligned to the beginning of each time interval, such as the first of the month, and represent the time interval of the analysis for the model system. You can save any new variables to the active dataset only if the observations are defined by a date specification, or record order, and the data are not aggregated.

Save Targets

You can save model predictions, confidence intervals, and residuals as new variables. Each specified target to save generates its own set of new variables, and each new variable contains values for both the estimation and forecast periods. For predicted values, confidence intervals, and noise residuals you can specify the root name to be used as the prefix for the new variables. The full variable name is the concatenation of the root name and the name of the field that contains the target series. The root name must conform to the rules for valid variable names. The variable name is extended if necessary to avoid variable naming conflicts.

Indicate cases containing forecasts

Creates a variable that indicates whether a record contains forecast data. You can specify the variable name. The default is *ForecastIndicator*.

Targets to save

Specifies whether new variables are created for all target series in the model system or just the target series that are specified on the **Series to Display** settings.

Export model system

Saves the model system to a compressed file archive (.zip file). The model system file can be used by the Temporal Causal Model Forecasting procedure to obtain updated forecasts or to generate any of the available output. It can also be used by the Temporal Causal Model Scenarios procedure to run scenario analysis.

Interactive Output

The output from temporal causal modeling includes a number of interactive output objects. Interactive features are available by activating (double-clicking) the object in the Output Viewer.

Overall model system

Displays the causal relations between series in the model system. All lines that connect a particular target to its inputs have the same color. The thickness of the line indicates the significance of the causal connection, where thicker lines represent a more significant connection. Inputs that are not also targets are indicated with a black square.

- You can display relations for top models, a specified series, all series, or models with no inputs. The top models are the models that meet the criteria that were specified for best-fitting models on the **Series to Display** settings.
- You can generate impact diagrams for one or more series by selecting the series names in the chart, right-clicking, and then choosing **Create Impact Diagram** from the context menu.
- You can choose to hide causal relations that have a significance level that is greater than a specified value. Smaller significance levels indicate a more significant causal relation.
- You can display relations for a particular series by selecting the series name in the chart, right-clicking, and then choosing **Highlight relations for series** from the context menu.

Impact diagram

Displays a graphical representation of the causal relations between a series of interest and other series that it affects or that affect it. Series that affect the series of interest are referred to as *causes*.

- You can change the series of interest by specifying the name of the series that you want. Double-clicking any node in the impact diagram changes the series of interest to the series associated with that node.

- You can toggle the display between causes and effects and you can change the number of levels of causes or effects to display.
- Single-clicking any node opens a detailed sequence diagram for the series that is associated with the node.

Outlier root cause analysis

Determines which series are most likely to be the cause of each outlier in a series of interest.

- You can display the root cause for any outlier by selecting the row for the outlier in the Outliers table. You can also display the root cause by clicking the icon for the outlier in the sequence chart.
- Single-clicking any node opens a detailed sequence diagram for the series that is associated with the node.

Overall model quality

Histogram of model fit for all models, for a particular fit statistic. Clicking a bar in the bar chart filters the dot plot so that it displays only the models that are associated with the selected bar. You can find the model for a particular target series in the dot plot by specifying the series name.

Outlier distribution

Bar chart of the number of outliers, across all targets, for each time interval in the estimation period. Clicking a bar in the bar chart filters the dot plot so that it displays only the outliers that are associated with the selected bar.

Note:

- If you save an output document that contains interactive output from temporal causal modeling and you want to retain the interactive features, then make sure that **Store required model information with the output document** is selected on the Save Output As dialog.
- Some interactive features require that the active dataset is the data that was used to build the temporal causal model system.

Applying Temporal Causal Models

Applying Temporal Causal Models

Two procedures are available for applying models that were created with the Temporal Causal Modeling procedure. Both procedures require the model system file, which can be saved as part of the Temporal Causal Modeling procedure.

Temporal Causal Model Forecasting

You can use this procedure to obtain forecasts for series for which more current data is available, without rebuilding your models. You can also generate any of the output that is available with the Temporal Causal Modeling procedure.

Temporal Causal Model Scenarios

Use this procedure to investigate how specific values of a particular time series, in a model system, affect the predicted values of the time series that are causally related to it.

Temporal Causal Model Forecasting

The Temporal Causal Model Forecasting procedure loads a model system file that was created by the Temporal Causal Modeling procedure and applies the models to the active dataset. You can use this procedure to obtain forecasts for series for which more current data is available, without rebuilding your models. You can also generate any of the output that is available with the Temporal Causal Modeling procedure.

Assumptions

- The structure of the data in the active dataset, either column-based or multidimensional, must be the same structure that was used when the model system was built. For multidimensional data, the

dimension fields must be the same as used to build the model system. Also, the dimension values that were used to build the model system must exist in the active dataset.

- Models are applied to fields in the active dataset with the same names as the fields specified in the model system.
- The field, or fields, that defined the observations when the model system was built must exist in the active dataset. The time interval between observations is assumed to be the same as when the models were built. If the observations were defined by a date specification, then the same date specification must exist in the active dataset. Date specifications are created from the Define Dates dialog or the DATE command.
- The time interval of the analysis and any settings for aggregation, distribution, and missing values are the same as when the models were built.

To Use Temporal Causal Model Forecasting

This feature requires the Statistics Forecasting option.

From the menus, choose:

Analyze > Forecasting > Apply Temporal Causal Models...

1. Enter the file specification for a model system file or click **Browse** and select a model system file. Model system files are created with the Temporal Causal Modeling procedure.
2. Click the option to reestimate models, create forecasts, and generate output.
3. Click **Continue**.
4. Specify whether you want to use existing model parameters or reestimate model parameters from the data in the active dataset.
5. Specify how far into the future to forecast, or specify not to forecast.
6. Click **Options** to specify the content of the output.
7. Click **Save** to save predictions, and export the updated model system to an external file when model parameters are reestimated.
8. Click **Run** to run the procedure.

Model Parameters and Forecasts

Load from model file

Forecasts are produced by using the model parameters from the model system file, and the data from the active dataset, without reestimating the model parameters. Goodness of fit measures that are displayed in output and used to select best-fitting models are taken from the model system file. The fit measures then reflect the data that was used when each model was developed (or last updated). This option is appropriate for generating forecasts and output from the data that was used to build the model system.

Reestimate from data

Model parameters are reestimated by using the data in the active dataset. Reestimation of model parameters does not affect which inputs are included in the model for each target. This option is appropriate when you have new data beyond the original estimation period and you want to generate forecasts or other output with the updated data.

All observations

Specifies that the estimation period starts at the time of the earliest observation and ends at the time of the latest observation across all series.

By start and end times

You can specify both the start and end of the estimation period or you can specify just the start or just the end. If you omit the start or the end of the estimation period, the default value is used.

- If the observations are defined by a date/time field, then enter values for start and end in the same format that is used for the date/time field.

- For observations that are defined by cyclic periods, specify a value for each of the cyclic periods fields. Each field is displayed in a separate column.
- If there is a date specification in effect for the active dataset, then you must specify a value for each component (such as Month) of the date specification. Each component is displayed in a separate column.
- When the observations are defined by record order, the start and end of the estimation period are defined by the row number (as displayed in the Data Editor) of the relevant case.

By latest or earliest time intervals

Defines the estimation period as a specified number of time intervals that start at the earliest time interval or end at the latest time interval in the data, with an optional offset. In this context, the time interval refers to the time interval of the analysis. For example, assume that the observations are monthly but the time interval of the analysis is quarters. Specifying **Latest** and a value of 24 for the **Number of time intervals** means the latest 24 quarters.

Optionally, you can exclude a specified number of time intervals. For example, specifying the latest 24 time intervals and 1 for the number to exclude means that the estimation period consists of the 24 intervals that precede the last one.

Extend records into the future

Sets the number of time intervals to forecast beyond the end of the estimation period. The time interval in this case is the time interval of the analysis. When forecasts are requested, autoregressive models are automatically built for any input series that are not also targets. These models are then used to generate values for those input series in the forecast period to obtain the forecasts for the targets of those inputs.

General Options

Confidence interval width (%)

This setting controls the confidence intervals for both forecasts and model parameters. You can specify any positive value less than 100. By default, a 95% confidence interval is used.

Outlier threshold (%)

An observation is flagged as an outlier if the probability, as calculated from the model, that it is an outlier exceeds this threshold. You can specify a value in the range 50 - 100.

Series to Display

These options specify the series (targets or inputs) for which output is displayed. The content of the output for the specified series is determined by the **Output Options** settings.

Display targets associated with best-fitting models

By default, output is displayed for the targets that are associated with the 10 best-fitting models, as determined by the R square value. You can specify a different fixed number of best-fitting models or you can specify a percentage of best-fitting models. You can also choose from the following goodness of fit measures:

R square

Goodness-of-fit measure of a linear model, sometimes called the coefficient of determination. It is the proportion of variation in the target variable explained by the model. It ranges in value from 0 to 1. Small values indicate that the model does not fit the data well.

Root mean square percentage error

A measure of how much the model-predicted values differ from the observed values of the series. It is independent of the units that are used and can therefore be used to compare series with different units.

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.

BIC

Bayesian Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The BIC also "penalizes" overparameterized models (complex models with a large number of inputs, for example), but more strictly than the AIC.

AIC

Akaike Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The AIC "penalizes" overparameterized models (complex models with a large number of inputs, for example).

Specify Individual Series

You can specify individual series for which you want output.

- For column-based data, you specify the fields that contain the series that you want. The order of the specified fields defines the order in which they appear in the output.
- For multidimensional data, you specify a particular series by adding an entry to the grid for the metric field that contains the series. You then specify the values of the dimension fields that define the series.
 - You can enter the value for each dimension field directly into the grid or you can select from the list of available dimension values. To select from the list of available dimension values, click the ellipsis button in the cell for the dimension that you want. This action opens the Select Dimension Value subdialog.
 - You can search the list of dimension values, on the Select Dimension Value subdialog, by clicking the binoculars icon and specifying a search term. Spaces are treated as part of the search term. Asterisks (*) in the search term do not indicate wildcard characters.
 - The order of the series in the grid defines the order in which they appear in the output.

For both column-based data and multidimensional data, output is limited to 30 series. This limit includes individual series (inputs or targets) that you specify and targets that are associated with best-fitting models. Individually specified series take precedence over targets that are associated with best-fitting models.

Output Options

These options specify the content of the output. Options in the **Output for targets** group generate output for the targets that are associated with the best-fitting models on the **Series to Display** settings. Options in the **Output for series** group generate output for the individual series that are specified on the **Series to Display** settings.

Overall model system

Displays a graphical representation of the causal relations between series in the model system. Tables of both model fit statistics and outliers for the displayed targets are included as part of the output item. When this option is selected in the **Output for series** group, a separate output item is created for each individual series that is specified on the **Series to Display** settings.

Causal relations between series have an associated significance level, where a smaller significance level indicates a more significant connection. You can choose to hide relations with a significance level that is greater than a specified value.

Model fit statistics and outliers

Tables of model fit statistics and outliers for the target series that are selected for display. These tables contain the same information as the tables in the Overall Model System visualization. These tables support all standard features for pivoting and editing tables.

Model effects and model parameters

Tables of model effects tests and model parameters for the target series that are selected for display. Model effects tests include the F statistic and associated significance value for each input included in the model.

Impact diagram

Displays a graphical representation of the causal relations between a series of interest and other series that it affects or that affect it. Series that affect the series of interest are referred to as *causes*. Selecting **Effects** generates an impact diagram that is initialized to display effects. Selecting **Causes** generates an impact diagram that is initialized to display causes. Selecting **Both causes and effects** generates two separate impact diagrams, one that is initialized to causes and one that is initialized to effects. You can interactively toggle between causes and effects in the output item that displays the impact diagram.

You can specify the number of levels of causes or effects to display, where the first level is just the series of interest. Each additional level shows more indirect causes or effects of the series of interest. For example, the third level in the display of effects consists of the series that contain series in the second level as a direct input. Series in the third level are then indirectly affected by the series of interest since the series of interest is a direct input to the series in the second level.

Series plot

Plots of observed and predicted values for the target series that are selected for display. When forecasts are requested, the plot also shows the forecasted values and the confidence intervals for the forecasts.

Residuals plot

Plots of the model residuals for the target series that are selected for display.

Top inputs

Plots of each displayed target, over time, along with the top 3 inputs for the target. The top inputs are the inputs with the lowest significance value. To accommodate different scales for the inputs and target, the y axis represents the z score for each series.

Forecast table

Tables of forecasted values and confidence intervals of those forecasts for the target series that are selected for display.

Outlier root cause analysis

Determines which series are most likely to be the cause of each outlier in a series of interest. Outlier root cause analysis is done for each target series that is included in the list of individual series on the **Series to Display** settings.

Output

Interactive outliers table and chart

Table and chart of outliers and root causes of those outliers for each series of interest. The table contains a single row for each outlier. The chart is an impact diagram. Selecting a row in the table highlights the path, in the impact diagram, from the series of interest to the series that most likely causes the associated outlier.

Pivot table of outliers

Table of outliers and root causes of those outliers for each series of interest. This table contains the same information as the table in the interactive display. This table supports all standard features for pivoting and editing tables.

Causal levels

You can specify the number of levels to include in the search for the root causes. The concept of levels that is used here is the same as described for impact diagrams.

Model fit across all models

Histogram of model fit for all models and for selected fit statistics. The following fit statistics are available:

R square

Goodness-of-fit measure of a linear model, sometimes called the coefficient of determination. It is the proportion of variation in the target variable explained by the model. It ranges in value from 0 to 1. Small values indicate that the model does not fit the data well.

Root mean square percentage error

A measure of how much the model-predicted values differ from the observed values of the series. It is independent of the units that are used and can therefore be used to compare series with different units.

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.

BIC

Bayesian Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The BIC also "penalizes" overparameterized models (complex models with a large number of inputs, for example), but more strictly than the AIC.

AIC

Akaike Information Criterion. A measure for selecting and comparing models based on the -2 reduced log likelihood. Smaller values indicate better models. The AIC "penalizes" overparameterized models (complex models with a large number of inputs, for example).

Outliers over time

Bar chart of the number of outliers, across all targets, for each time interval in the estimation period.

Series transformations

Table of any transformations that were applied to the series in the model system. The possible transformations are missing value imputation, aggregation, and distribution.

Save**Save Targets**

You can save model predictions, confidence intervals, and residuals as new variables. Each specified target to save generates its own set of new variables, and each new variable contains values for both the estimation and forecast periods. For predicted values, confidence intervals, and noise residuals you can specify the root name to be used as the prefix for the new variables. The full variable name is the concatenation of the root name and the name of the field that contains the target series. The root name must conform to the rules for valid variable names. The variable name is extended if necessary to avoid variable naming conflicts.

Indicate cases containing forecasts

Creates a variable that indicates whether a record contains forecast data. You can specify the variable name. The default is *ForecastIndicator*.

Targets to save

Specifies whether new variables are created for all target series in the model system or just the target series that are specified on the **Series to Display** settings.

Destination Options

You can save both transformations of the data (such as aggregation or imputation of missing values) and new variables (specified on the **Save Targets** settings) to an IBM SPSS Statistics data file or a new dataset in the current session. Date/time values in the saved data are aligned to the beginning of each time interval, such as the first of the month, and represent the time interval of the analysis for the model system. You can save any new variables to the active dataset only if the observations are defined by a date specification, or record order, and the data are not aggregated.

Export model system

Saves the model system to a compressed file archive (.zip file). The model system file can be reused by this procedure. It can also be used by the Temporal Causal Model Scenarios procedure to run scenario analysis. This option is only available when model parameters are reestimated.

Temporal Causal Model Scenarios

The Temporal Causal Model Scenarios procedure runs user-defined scenarios for a temporal causal model system, with the data from the active dataset. A *scenario* is defined by a time series, that is referred to as the *root series*, and a set of user-defined values for that series over a specified time range.

The specified values are then used to generate predictions for the time series that are affected by the root series. The procedure requires a model system file that was created by the Temporal Causal Modeling procedure. It is assumed that the active dataset is the same data as was used to create the model system file.

Example

Using the Temporal Causal Modeling procedure, a business decision maker discovered a key metric that affects a number of important performance indicators. The metric is controllable, so the decision maker wants to investigate the effect of various sets of values for the metric over the next quarter. The investigation is easily done by loading the model system file into the Temporal Causal Model Scenarios procedure and specifying the sets of values for the key metric.

To Run Temporal Causal Model Scenarios

This feature requires the Statistics Forecasting option.

From the menus, choose:

Analyze > Forecasting > Apply Temporal Causal Models...

1. Enter the file specification for a model system file or click **Browse** and select a model system file. Model system files are created with the Temporal Causal Modeling procedure.
2. Click the option to run scenarios.
3. Click **Continue**.
4. On the Scenarios tab (in the Temporal Causal Model Scenarios dialog), click **Define Scenario Period** and specify the scenario period.
5. For column-based data, click **Add Scenario** to define each scenario. For multidimensional data, click **Add Scenario** to define each individual scenario, and click **Add Scenario Group** to define each scenario group.
6. Click **Options** to specify the content of the output and to specify the breadth of the series that will be affected by the scenario.
7. Click **Run** to run the procedure.

Defining the Scenario Period

The scenario period is the period over which you specify the values that are used to run your scenarios. It can start before or after the end of the estimation period. You can optionally specify to predict beyond the end of the scenario period. By default, predictions are generated through the end of the scenario period. All scenarios use the same scenario period and specifications for how far to predict.

Note: Predictions start at the first time period after the beginning of the scenario period. For example, if the scenario period starts on 2014-11-01 and the time interval is months, then the first prediction is for 2014-12-01.

Specify by start, end and predict through times

- If the observations are defined by a date/time field, then enter values for start, end, and predict through in the same format that is used for the date/time field. Values for date/time fields are aligned to the beginning of the associated time interval. For example, if the time interval of the analysis is months, then the value 10/10/2014 is adjusted to 10/01/2014, which is the beginning of the month.
- For observations that are defined by cyclic periods, specify a value for each of the cyclic periods fields. Each field is displayed in a separate column.
- If there is a date specification in effect for the active dataset, then you must specify a value for each component (such as Month) of the date specification. Each component is displayed in a separate column.

- When the observations are defined by record order, start, end, and predict through are defined by the row number (as displayed in the Data Editor) of the relevant case.

Specify by time intervals relative to end of estimation period

Defines start and end in terms of the number of time intervals relative to the end of the estimation period, where the time interval is the time interval of the analysis. The end of the estimation period is defined as time interval 0. Time intervals before the end of the estimation period have negative values and intervals after the end of the estimation period have positive values. You can also specify how many intervals to predict beyond the end of the scenario period. The default is 0.

For example, assume that the time interval of the analysis is months and that you specify 1 for starting interval, 3 for ending interval, and 1 for how far to predict beyond that. The scenario period is then the 3 months that follow the end of the estimation period. Predictions are generated for the second and third month of the scenario period and for 1 more month past the end of the scenario period.

Adding Scenarios and Scenario Groups

The Scenarios tab specifies the scenarios that are to be run. To define scenarios, you must first define the scenario period by clicking **Define Scenario Period**. Scenarios and scenario groups (applies only to multidimensional data) are created by clicking the associated **Add Scenario** or **Add Scenario Group** button. By selecting a particular scenario or scenario group in the associated grid, you can edit it, make a copy of it, or delete it.

Column-based data

The **Root field** column in the grid specifies the time series field whose values are replaced with the scenario values. The **Scenario values** column displays the specified scenario values in the order of earliest to latest. If the scenario values are defined by an expression, then the column displays the expression.

Multidimensional data

Individual Scenarios

Each row in the Individual Scenarios grid specifies a time series whose values are replaced by the specified scenario values. The series is defined by the combination of the field that is specified in the **Root Metric** column and the specified value for each of the dimension fields. The content of the **Scenario values** column is the same as for column-based data.

Scenario Groups

A *scenario group* defines a set of scenarios that are based on a single root metric field and multiple sets of dimension values. Each set of dimension values (one value per dimension field), for the specified metric field, defines a time series. An individual scenario is then generated for each such time series, whose values are then replaced by the scenario values. Scenario values for a scenario group are specified by an expression, which is then applied to each time series in the group.

The **Series Count** column displays the number of sets of dimension values that are associated with a scenario group. The displayed value can be larger than the actual number of time series that are associated with the scenario group (one series per set). This condition occurs when some of the specified combinations of dimension values do not correspond to series contained by the root metric for the group.

As an example of a scenario group, consider a metric field *advertising* and two dimension fields *region* and *brand*. You can define a scenario group that is based on *advertising* as the root metric and that includes all combinations of *region* and *brand*. You might then specify $advertising * 1.2$ as the expression to investigate the effect of increasing *advertising* by 20 percent for each of the time series that are associated with the *advertising* field. If there are 4 values of *region* and 2 values of *brand*, then there are 8 such time series and thus 8 scenarios defined by the group.

Scenario Definition

Settings for defining a scenario depend on whether your data are column-based or multidimensional.

Root series

Specifies the root series for the scenario. Each scenario is based on a single root series. For column-based data, you select the field that defines the root series. For multidimensional data, you specify the root series by adding an entry to the grid for the metric field that contains the series. You then specify the values of the dimension fields that define the root series. The following apply to specifying the dimension values:

- You can enter the value for each dimension field directly into the grid or you can select from the list of available dimension values. To select from the list of available dimension values, click the ellipsis button in the cell for the dimension that you want. This action opens the Select Dimension Value subdialog.
- You can search the list of dimension values, on the Select Dimension Value subdialog, by clicking the binoculars icon and specifying a search term. Spaces are treated as part of the search term. Asterisks (*) in the search term do not indicate wildcard characters.

Specify affected targets

Use this option when you know specific targets that are affected by the root series and you want to investigate the effects on those targets only. By default, targets that are affected by the root series are automatically determined. You can specify the breadth of the series that are affected by the scenario with settings on the Options tab.

For column-based data, select the targets that you want. For multidimensional data, you specify target series by adding an entry to the grid for the target metric field that contains the series. By default, all series that are contained in the specified metric field are included. You can customize the set of included series by customizing the included values for one or more of the dimension fields. To customize the dimension values that are included, click the ellipsis button for the dimension that you want. This action opens the Select Dimension Values dialog.

The **Series Count** column (for multidimensional data) displays the number of sets of dimension values that are currently specified for the associated target metric. The displayed value can be larger than the actual number of affected target series (one series per set). This condition occurs when some of the specified combinations of dimension values do not correspond to series contained by the associated target metric.

Scenario ID

Each scenario must have a unique identifier. The identifier is displayed in output that is associated with the scenario. There are no restrictions, other than uniqueness, on the value of the identifier.

Specify scenario values for root series

Use this option to specify explicit values for the root series in the scenario period. You must specify a numeric value for each time interval that is listed in the grid. You can obtain the values of the root series (actual or forecasted) for each interval in the scenario period by clicking **Read**, **Forecast**, or **Read\Forecast**.

Specify expression for scenario values for root series

You can define an expression for computing the values of the root series in the scenario period. You can enter the expression directly or click the calculator button and create the expression from the Scenario Values Expression Builder.

- The expression can contain any target or input in the model system.
- When the scenario period extends beyond the existing data, the expression is applied to forecasted values of the fields in the expression.
- For multidimensional data, each field in the expression specifies a time series that is defined by the field and the dimension values that were specified for the root metric. It is those time series that are used to evaluate the expression.

As an example, assume that the root field is *advertising* and the expression is $advertising \times 1.2$. The values for *advertising* that are used in the scenario represent a 20 percent increase over the existing values.

Note: Scenarios are created by clicking **Add Scenario** on the Scenarios tab.

Scenario Values Expression Builder

Use the Scenario Values Expression Builder to create an expression that is used to compute scenario values for a single scenario or a scenario group. To build an expression, either paste components into the **Expression** field or type directly in the **Expression** field.

- The expression can contain any target or input in the model system.
- You can paste functions by selecting a group from the Function group list and double-clicking the function in the Functions list (or select the function and click the arrow adjacent to the Function group list). Enter any parameters indicated by question marks. The function group labeled **All** provides a listing of all available functions. A brief description of the currently selected function is displayed in a reserved area in the dialog box.
- String constants must be enclosed in quotation marks.
- If values contain decimals, a period (.) must be used as the decimal indicator.

To access the Scenario Values Expression Builder, click the calculator button on the Scenario Definition or Scenario Group Definition dialog.

Select Dimension Values

For multidimensional data, you can customize the dimension values that define the targets that are affected by a scenario or scenario group. You can also customize the dimension values that define the set of root series for a scenario group.

All values

Specifies that all values of the current dimension field are included. This option is the default.

Select values

Use this option to specify the set of values for the current dimension field. You can filter the set of values from which to choose. Values that meet the filter condition appear in the **Matched** tab and values that do not meet the filter condition appear in the **Unmatched** tab of the **Unselected values** list. The **All** tab lists all unselected values, regardless of any filter condition.

- You can use asterisks (*) to indicate wildcard characters when you specify a filter.
- To clear the current filter, specify an empty value for the search term on the Filter Displayed Values dialog.

To customize dimension values for affected targets:

1. From the Scenario Definition or Scenario Group Definition dialog, select the target metric for which you want to customize dimension values.
2. Click the ellipsis button in the column for the dimension that you want to customize.

To customize dimension values for the root series of a scenario group:

1. From the Scenario Group Definition dialog, click the ellipsis button (in the root series grid) for the dimension that you want to customize.

Scenario Group Definition

Root Series

Specifies the set of root series for the scenario group. An individual scenario is generated for each time series in the set. You specify the root series by adding an entry to the grid for the metric field that contains the series that you want. You then specify the values of the dimension fields that define the set. By default, all series that are contained in the specified root metric field are included. You can customize the set of included series by customizing the included values for one or more of the dimension fields. To customize the dimension values that are included, click the ellipsis button for a dimension. This action opens the Select Dimension Values dialog.

The **Series Count** column displays the number of sets of dimension values that are currently included for the associated root metric. The displayed value can be larger than the actual number of root series for the scenario group (one series per set). This condition occurs when some of the specified combinations of dimension values do not correspond to series contained by the root metric.

Specify affected target series

Use this option when you know specific targets that are affected by the set of root series and you want to investigate the effects on those targets only. By default, targets that are affected by each root series are automatically determined. You can specify the breadth of the series that are affected by each individual scenario with settings on the Options tab.

You specify target series by adding an entry to the grid for the metric field that contains the series. By default, all series that are contained in the specified metric field are included. You can customize the set of included series by customizing the included values for one or more of the dimension fields. To customize the dimension values that are included, click the ellipsis button for the dimension that you want. This action opens the Select Dimension Values dialog.

The **Series Count** column displays the number of sets of dimension values that are currently specified for the associated target metric. The displayed value can be larger than the actual number of affected target series (one series per set). This condition occurs when some of the specified combinations of dimension values do not correspond to series contained by the associated target metric.

Scenario ID prefix

Each scenario group must have a unique prefix. The prefix is used to construct an identifier that is displayed in output that is associated with each individual scenario in the scenario group. The identifier for an individual scenario is the prefix, followed by an underscore, followed by the value of each dimension field that identifies the root series. The dimension values are separated by underscores. There are no restrictions, other than uniqueness, on the value of the prefix.

Expression for scenario values for root series

Scenario values for a scenario group are specified by an expression, which is then used to compute the values for each of the root series in the group. You can enter an expression directly or click the calculator button and create the expression from the Scenario Values Expression Builder.

- The expression can contain any target or input in the model system.
- When the scenario period extends beyond the existing data, the expression is applied to forecasted values of the fields in the expression.
- For each root series in the group, the fields in the expression specify time series that are defined by those fields and the dimension values that define the root series. It is those time series that are used to evaluate the expression. For example, if a root series is defined by `region= ' north '` and `brand= ' X '`, then the time series that are used in the expression are defined by those same dimension values.

As an example, assume that the root metric field is *advertising* and that there are two dimension fields *region* and *brand*. Also, assume that the scenario group includes all combinations of the dimension field values. You might then specify `advertising*1.2` as the expression to investigate the effect of increasing *advertising* by 20 percent for each of the time series that are associated with the *advertising* field.

Note: Scenario groups apply to multidimensional data only and are created by clicking **Add Scenario Group** on the Scenarios tab.

Options

Maximum level for affected targets

Specifies the maximum number of levels of affected targets. Each successive level, up to the maximum of 5, includes targets that are more indirectly affected by the root series. Specifically, the first level includes targets that have the root series as a direct input. Targets in the second level have targets in the first level as a direct input, and so on. Increasing the value of this setting increases the complexity of the computation and might affect performance.

Maximum auto-detected targets

Specifies the maximum number of affected targets that are automatically detected for each root series. Increasing the value of this setting increases the complexity of the computation and might affect performance.

Impact diagram

Displays a graphical representation of the causal relations between the root series for each scenario and the target series that it affects. Tables of both the scenario values and the predicted values for the affected targets are included as part of the output item. The graph includes plots of the predicted values of the affected targets. Single-clicking any node in the impact diagram opens a detailed sequence diagram for the series that is associated with the node. A separate impact diagram is generated for each scenario.

Series plots

Generates series plots of the predicted values for each of the affected targets in each scenario.

Forecast and scenario tables

Tables of predicted values and scenario values for each scenario. These tables contains the same information as the tables in the impact diagram. These tables support all standard features for pivoting and editing tables.

Include confidence intervals in plots and tables

Specifies whether confidence intervals for scenario predictions are included in both chart and table output.

Confidence interval width (%)

This setting controls the confidence intervals for the scenario predictions. You can specify any positive value less than 100. By default, a 95% confidence interval is used.

Vector Autoregressive Models

A Vector Autoregressive (VAR) model is a fundamental tool in multivariate time series analysis. The concept is to extend the idea of autoregressive models, which explain a variable in terms of its own past values, into the multivariate case, where multiple interrelated time series are modeled together. A VAR model captures the linear interdependencies among multiple time series variables. That is, a VAR(p) model expresses each variable as a linear function of its own past values and the past values of all other variables: Each variable in the system is explained by its own past values (auto regression) and the past values of all other variables in the system

The steps in VAR analysis include defining date and time, stationary check, model estimation, lag order selection, model diagnostics, and forecasting.

Statistics

- For each variable: Identifies the number of valid cases, measures the mean and standard deviation, and specifies the minimum and maximum values.
- For each model: Assesses the parameter estimates (coefficients), standard error of coefficients, t -statistic, p -value (Sig.), lower and upper confidence intervals, covariance matrix of residuals, determinant of residual covariance matrix, correlation matrix of residuals, eigen values of residual correlation matrix, residual sum of squares, R square, Adjusted R square, MSE, RMSE, MAE, MAPE, Ljung-Box test for autocorrelation of residuals, and Jarque-Bera test for normality of residuals.
- For goodness of fit: Calculates AIC, BIC, and HQIC for various lags.
- Plots: Time series, ACF, residual plot for each variable, combined residual plot, residual ACF plot for each variable, fit values, and forecast.

Data Considerations

- Endogenous variables must be scale. A minimum of 2 variables are required.
- Exogenous variables can be scale or categorical. This is optional

Obtaining Vector Autoregressive Models

1. VAR models are applicable only for time series data. Click **Define date and time** to generate date and time variables.
2. Stationarity is essential for VAR modeling. Non-stationary series often require differencing or cointegration analysis.

A stationary time series is one whose statistical properties such as mean, variance, and autocorrelation remain constant over time. Before you move to fit a VAR model, make sure that the time series is stationary.

For non-stationary series, take the first difference of actual data and check for Stationary. Take the difference again and continue the procedure until the data attains stationarity.

The Augmented Dickey-Fuller (ADF) test can be used for checking stationarity. ADF test is part of the extension procedure, **STATS > TSTESTS**. After you install **STATS > TSTESTS**, go to **Analyze > Forecasting > Stationarity and Cointegration Tests**. Select a variable and click **Tests**. Select **Compute test (one variable)** for Augmented Dickey-Fuller Test.. Run the test.

3. After the stationary test, click **Analyze > Forecasting > Mutivariate Time Series > Vector Autoregressive Models**.

Note: The fields highlighted in red are required. The **Paste** and **OK** buttons are enabled after you enter valid values in all required fields.

4. Select the Endogenous and Exogenous variables where, Exogenous variables are optional.
5. The optimal lag length must be chosen for better forecasting. This can be performed based on the selection by using the various goodness of fit criterion or information criterion: AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria, or HQIC (Hannan Quin Information Criteria).
6. Select the options from Statistics tab, Plots tab, Save tab, and Options tab.
7. Click **OK**.

Variables

Variables

Variables list act as a source from which you can select both Endogenous variables and Exogenous variables.

- Select two or more Endogenous variables. The Endogenous variables are the variables that are determined within the model or the system being studied. They are influenced by other variables in the model, including both exogenous and other endogenous variables.
- The **Exogenous variables** field is optional. The Exogenous variables are independent factors that affect the system from the outside. They are determined outside the model or the system being studied. These variables are not influenced by other variables within the model but can have an impact on the endogenous variables.

Lag Selection

The optimal lag length must be chosen for better forecasting. You can automatically select the best lag order by entering a value for maximum lag. By default, the minimum lag is set as 1. This is done by using any of the 3 goodness of fit criterion available in SPSS Statistics:

AIC (Akaike Information Criteria), **BIC** (Bayesian Information Criteria, or **HQIC** (Hannan Quin Information Criteria)

- **Maximum Lag** should be an integer. In IBM SPSS Statistics, the maximum lag allowed in 24. .

Method

The method refers to the estimation method in the VAR model. The Ordinary least square (OLS) method is the default method.

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 Endogenous variable and thus differ from the displayed value. For a given Endogenous 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.

Statistics

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

Fit Measures and Forecasts

Denotes the various fit measures for the developed VAR models. You can select one or more of the following options for inclusion in the table containing fit measures for estimated model:

- **Descriptive Statistics:** Determines the descriptive measures such as mean, standard deviation, and minimum and maximum of each variables selected under Endogenous variables.
- **Goodness of Fit:** Determines the AIC, BIC and HQIC for best lag order corresponding to the variables selected.
- **Parameter Estimates:** Determines the estimates (coefficients), standard error, t-statistic, p-value, lower and upper confidence intervals by OLS corresponding to each variables.

Note: Parameter estimates are required for VAR models. Therefore, the option **Parameter Estimates** is selected by default as built-in option and cannot be modified.

- **Residual Sum of Squares:** Determines the sum of squares of residuals (or errors), where, residuals are obtained by the difference between the actual and predicted (fitted) values.
- **Residual Autocorrelation Function (ACF):** Determines the autocorrelation function of residuals.
- **Mean Square Error:** Determines the mean square error, which is the average of square of residuals.
- **Root Mean Square Error:** Determines the root mean square error, which is the square root of mean square error.
- **Mean Absolute Error:** Determines the mean absolute error, which is the mean of absolute values of residuals.
- **Mean Absolute Percentage Error:** Determines the mean absolute percentage error.
- **R Square:** Determines the R Square value or Coefficient of determination.
- **Adjusted R Square:** Determines the Adjusted R Square value.
- **Correlation Matrix of Residuals:** Determines the correlation matrix of residuals for all variables. The covariance matrix of residuals and its determinant value are also included in the output.
- **Eigen Values of Residual Correlation Matrix:** Determines the eigen values of residual correlation matrix.
- **Test for Normality of Residuals:** Examines whether the residuals are normally distributed or not. The Jarque Bera test could be used for this.
- **Test for Autocorrelation:** Examines if a group of auto correlations for a series (or model residuals) are significantly different from zero or test whether the data is white noise (random) or has underlying patterns (serial correlation). The Ljung Box test could be used for this.
- **Forecasts:** Displays the predicted values of the variables.

Plots

In the **Plots** tab, specify the options for displaying plots of the modeling results. You can generate visualizations to better understand the relationships between the observed values, predicted values, residuals and other characteristics of variables. The tab includes various plots such as time series plots corresponding to each variable, ACF plot, residual plots for each variable, combined residual plot, residual ACF for each variable, fit values, and forecast.

Each visualization provides a unique perspective on the VAR model and is customizable based on user preferences.

Available plots

- **Time Series:** A time series plot is a graphical display of data points recorded in chronological order, where time is shown on the horizontal axis and the measured variable on the vertical axis. You can visualize the time series plot of each variable selected under Endogenous and Exogenous variables.
- **ACF:** An Autocorrelation Function (ACF) plot is a graphical tool that shows the correlation between a time series and its own lagged values across different time lags. The horizontal axis represents the lag number, and the vertical axis represents the autocorrelation coefficient at each lag.
- **Residual (each variable):** A residual plot is a diagnostic graph that displays the residuals (observed values minus fitted values or fitted values minus observed values) from a statistical (VAR) model, typically plotted against time. This plot is obtained by taking time along horizontal axis and the residuals along the vertical axis. The residual plot is applicable to all selected endogenous and exogenous variables.
- **Residual Combined:** This is a combined form of **Residual (each variable)**. Instead of displaying residual plots for each variables in separate graphs, the residuals of all variables is displayed in the same graph, with time marked along the x axis.
- **Residual ACF (each variable):** A residual ACF plot visualizes the correlation of a model's errors (or residuals) with their past values (lags) to check for remaining patterns, indicating if the model adequately captured the structure of the data. Ideally, significant spikes outside confidence bands suggest autocorrelation. This is plotted by taking the lag (time difference between observations) along the horizontal axis and the autocorrelation value along the vertical axis. The Confidence intervals (default 95%) are marked by blue dashed lines.
- **Fit Values:** This plot displays the plot of predicted values corresponding to each variable, at various time periods.
- **Forecast:** This plot displays the forecast at various time periods.

Time Series and **Fit Values** are selected by default.

Save

With the **Save** tab, you can choose options to save the predicted (fitted) values and residuals as additional columns in a new dataset.

Predicted Values

Save the Predicted values

Noise Residuals

Save the noise residuals.

Choose dataset

You can save the current dataset and along with predictive values and residuals as a new .sav file. In the **New dataset** field, click the ellipsis button ("..."). Select a file location, enter a file name, and click **Save**.

You can also save the new dataset to IBM SPSS Collaboration and Deployment Services Repository. See .

Options

The Options tab allows you to set the forecast period, specify how the missing values should be handled and set the confidence interval width.

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.

User-Missing Values

These options control the handling of user-missing values.

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

Confidence Interval Width (%)

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

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.

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.

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 3. $ARIMA(0,0,1), q>0$

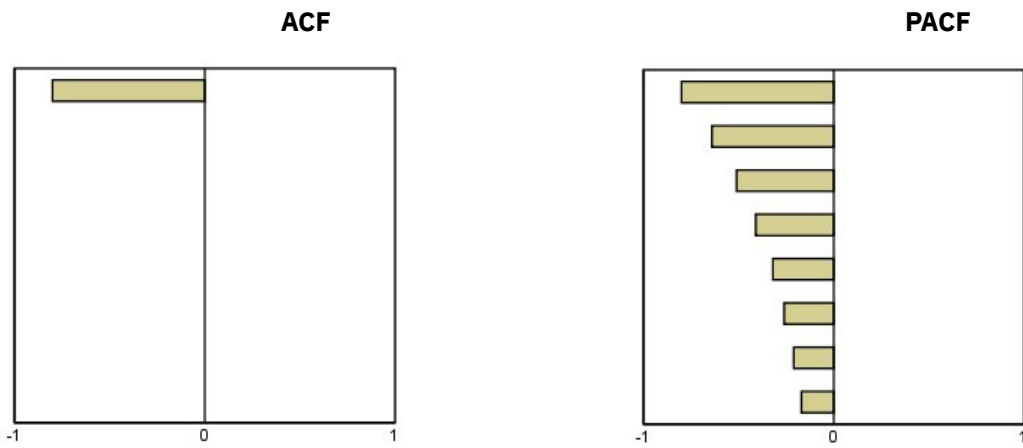
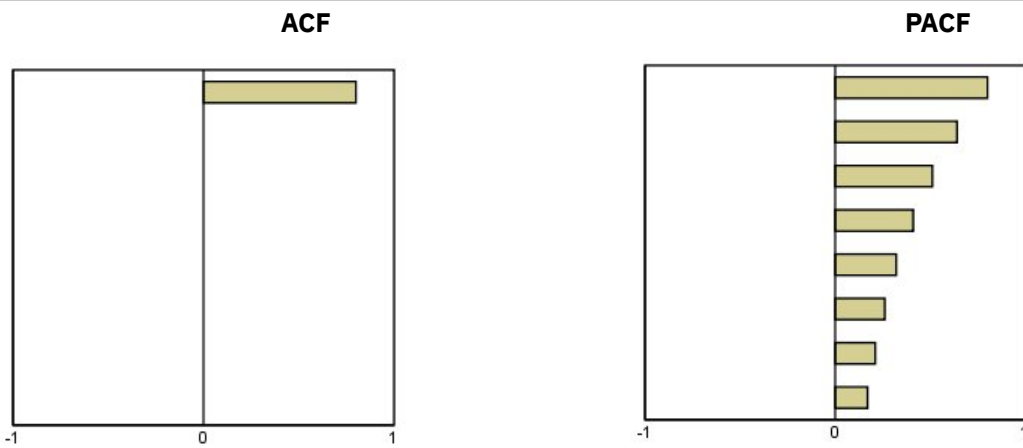


Table 4. $ARIMA(0,0,1), q<0$



$ARIMA(0,0,2), \rho_1 > 0$

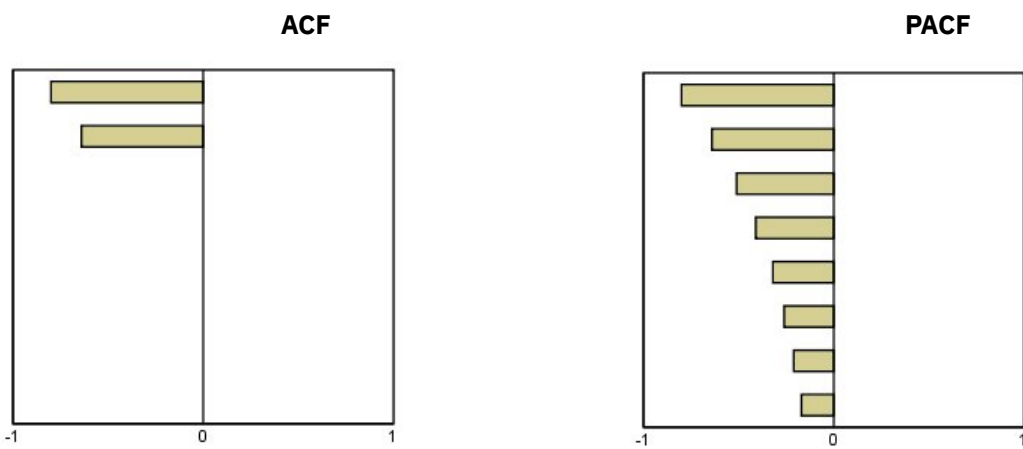


Table 5. ARIMA(1,0,0), $f > 0$

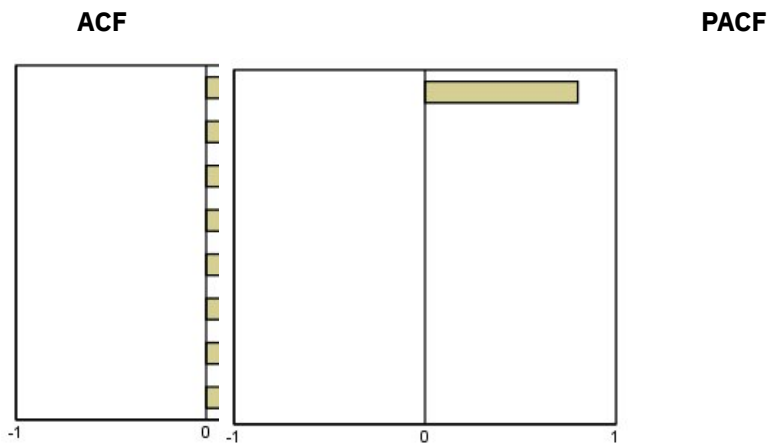
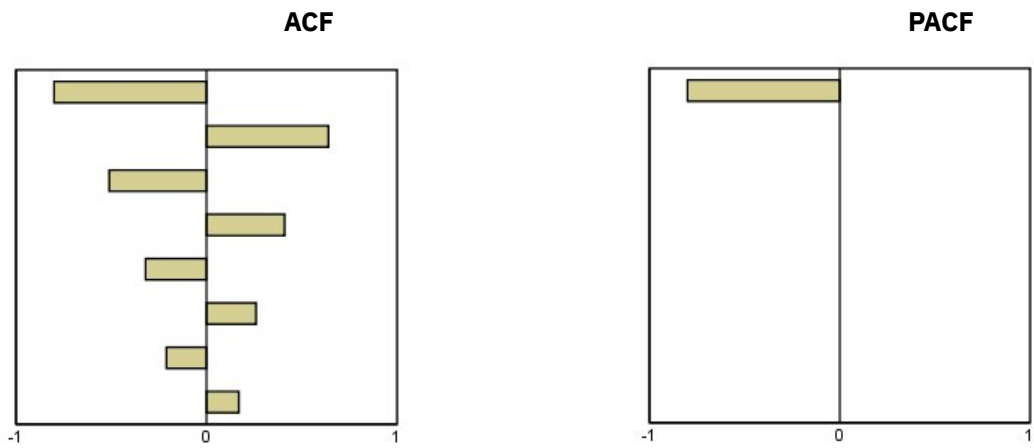
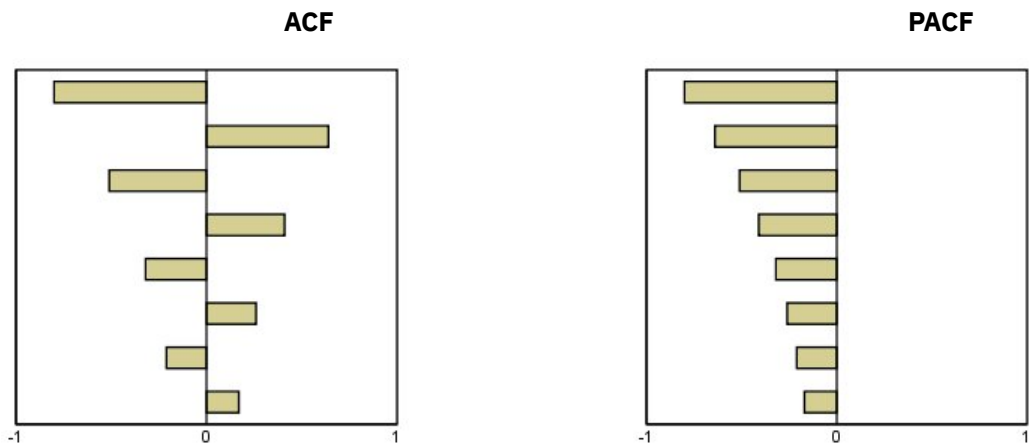


Table 6. ARIMA(1,0,0), $f < 0$



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



ARIMA(2,0,0), $\rho_1 \rho_2 > 0$

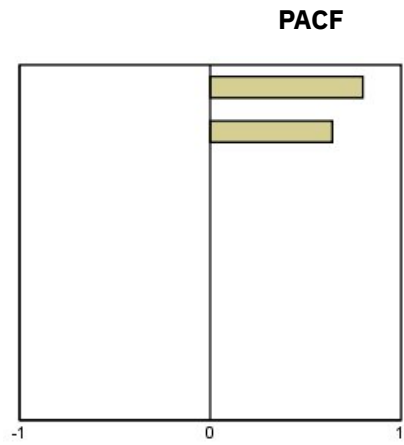
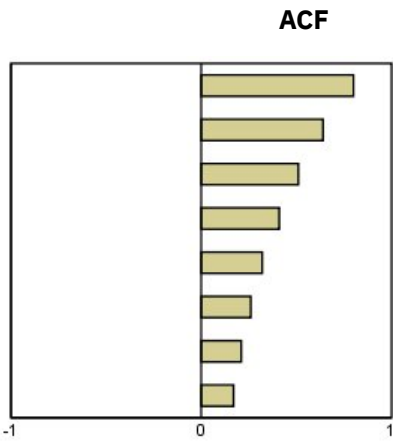
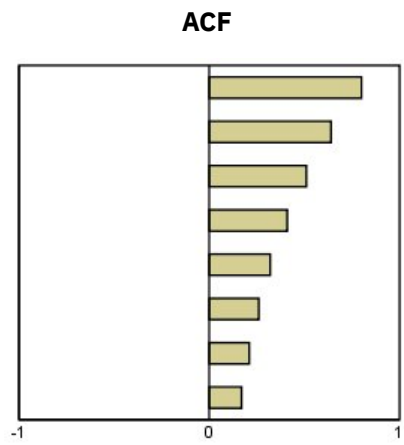


Table 7. $ARIMA(0,1,0)$ (integrated series)



Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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, MD-NC119
Armonk, NY 10504-1785
US*

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.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan*

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 jurisdictions 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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide 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 Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US*

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.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

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 actual people or business enterprises 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:

© Copyright IBM Corp. 2021. Portions of this code are derived from IBM Corp. Sample Programs.

© Copyright IBM Corp. 1989 - 2021. 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 [15](#), [16](#)
 - in Time Series Modeler [9](#), [10](#)
 - plots for pure ARIMA processes [48](#)
- additive outlier
 - in Time Series Modeler [5](#), [9](#)
- additive patch outlier
 - in Time Series Modeler [5](#), [9](#)
- Apply Time Series Models
 - best- and poorest-fitting models [16](#)
 - Box-Ljung statistic [15](#)
 - confidence intervals [16](#), [17](#)
 - estimation period [13](#)
 - fit values [16](#)
 - forecast period [13](#)
 - forecasts [15](#), [16](#)
 - goodness-of-fit statistics [15](#), [16](#)
 - missing values [17](#)
 - model parameters [15](#)
 - new variable names [17](#)
 - reestimate model parameters [13](#)
 - residual autocorrelation function [15](#), [16](#)
 - residual partial autocorrelation function [15](#), [16](#)
 - saving predictions [17](#)
 - saving reestimated models in XML [17](#)
 - statistics across all models [15](#), [16](#)
- ARIMA models
 - outliers [9](#)
 - transfer functions [8](#)
- autocorrelation function
 - in Apply Time Series Models [15](#), [16](#)
 - in Time Series Modeler [9](#), [10](#)
 - plots for pure ARIMA processes [48](#)

B

- Box-Ljung statistic
 - in Apply Time Series Models [15](#)
 - in Time Series Modeler [9](#)
- Brown's exponential smoothing model [6](#)

C

- confidence intervals
 - in Apply Time Series Models [16](#), [17](#)
 - in Time Series Modeler [10](#), [12](#)

D

- damped exponential smoothing model [6](#)

E

- estimation period

- estimation period (*continued*)
 - in Apply Time Series Models [13](#)
 - in Time Series Modeler [3](#)
- events
 - in Time Series Modeler [5](#)
- Expert Modeler
 - limiting the model space [5](#)
 - outliers [5](#)
- exponential smoothing models [3](#), [6](#)

F

- fit values
 - in Apply Time Series Models [16](#)
 - in Time Series Modeler [10](#)
- forecast period
 - in Apply Time Series Models [13](#)
 - in Time Series Modeler [3](#), [12](#)
- forecasts
 - in Apply Time Series Models [15](#), [16](#)
 - in Time Series Modeler [9](#), [10](#)

G

- goodness of fit
 - definitions [47](#)
 - in Apply Time Series Models [15](#), [16](#)
 - in Time Series Modeler [9](#), [10](#)

H

- harmonic analysis [20](#)
- historical data
 - in Apply Time Series Models [16](#)
 - in Time Series Modeler [10](#)
- historical period [2](#)
- holdout cases [2](#)
- Holt's exponential smoothing model [6](#)

I

- innovational outlier
 - in Time Series Modeler [5](#), [9](#)

L

- level shift outlier
 - in Time Series Modeler [5](#), [9](#)
- local trend outlier
 - in Time Series Modeler [5](#), [9](#)
- log transformation
 - in Time Series Modeler [6–8](#)

M

- MAE

MAE (*continued*)
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

MAPE
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

MaxAE
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

MaxAPE
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

maximum absolute error
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

maximum absolute percentage error
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

mean absolute error
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

mean absolute percentage error
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

missing values
 in Apply Time Series Models [17](#)
 in Time Series Modeler [12](#)

model names
 in Time Series Modeler [12](#)

model parameters
 in Apply Time Series Models [15](#)
 in Time Series Modeler [9](#)

models
 ARIMA [3](#)
 Expert Modeler [3](#)
 exponential smoothing [3](#), [6](#)

N

natural log transformation
 in Time Series Modeler [6–8](#)

normalized BIC (Bayesian information criterion)
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

O

outliers
 ARIMA models [9](#)
 definitions [48](#)
 Expert Modeler [5](#)

P

PACF
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)
 plots for pure ARIMA processes [48](#)

partial autocorrelation function
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)
 plots for pure ARIMA processes [48](#)

periodicity

periodicity (*continued*)
 in Time Series Modeler [5–8](#)

R

R2
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

reestimate model parameters
 in Apply Time Series Models [13](#)

residuals
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

RMSE
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

root mean square error
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

S

save
 model predictions [11](#), [17](#)
 model specifications in XML [11](#)
 new variable names [11](#), [17](#)
 reestimated models in XML [17](#)

seasonal additive outlier
 in Time Series Modeler [5](#), [9](#)

Seasonal Decomposition
 assumptions [18](#)
 computing moving averages [18](#)
 create variables [19](#)
 models [18](#)
 saving new variables [19](#)

simple exponential smoothing model [6](#)

simple seasonal exponential smoothing model [6](#)

Spectral Plots
 assumptions [20](#)
 bivariate spectral analysis [20](#)
 centering transformation [20](#)
 spectral windows [20](#)

square root transformation
 in Time Series Modeler [6–8](#)

stationary R2
 in Apply Time Series Models [15](#), [16](#)
 in Time Series Modeler [9](#), [10](#)

T

temporal causal model forecasting [32–35](#), [37](#)

temporal causal model scenarios [37–39](#), [41](#), [42](#)

temporal causal models [22–28](#), [30](#), [31](#)

time series analysis
 temporal causal models [22](#)

Time Series Modeler
 ARIMA [3](#), [7](#)
 best- and poorest-fitting models [11](#)
 Box-Ljung statistic [9](#)
 confidence intervals [10](#), [12](#)
 estimation period [3](#)
 events [5](#)
 Expert Modeler [3](#)

Time Series Modeler (*continued*)

- exponential smoothing [3](#), [6](#)
- fit values [10](#)
- forecast period [3](#), [12](#)
- forecasts [9](#), [10](#)
- goodness-of-fit statistics [9](#), [10](#)
- missing values [12](#)
- model names [12](#)
- model parameters [9](#)
- new variable names [11](#)
- outliers [5](#), [9](#)
- periodicity [5–8](#)
- residual autocorrelation function [9](#), [10](#)
- residual partial autocorrelation function [9](#), [10](#)
- saving model specifications in XML [11](#)
- saving predictions [11](#)
- series transformation [6–8](#)
- statistics across all models [9](#), [10](#)
- transfer functions [8](#)
- transfer functions
 - delay [8](#)
 - denominator orders [8](#)
 - difference orders [8](#)
 - numerator orders [8](#)
 - seasonal orders [8](#)
- transient outlier
 - in Time Series Modeler [5](#), [9](#)

V

- validation period [2](#)
- variable names
 - in Apply Time Series Models [17](#)
 - in Time Series Modeler [11](#)

W

- Winters' exponential smoothing model
 - additive [6](#)
 - multiplicative [6](#)

X

- XML
 - saving reestimated models in XML [17](#)
 - saving time series models in XML [11](#)

