Paper 193-30

Introducing SAS® Forecast Studio

Brenda Wolfe, SAS Institute Inc., Cary, NC
Michael Leonard, SAS Institute Inc., Cary, NC
Paddy Fahey, SAS Institute Inc., Cary, NC

ABSTRACT

Organizations depend on statistical forecasting to provide a solid foundation for many important planning and decision-making processes. SAS® Forecast Studio facilitates and speeds the statistical forecasting process by providing a convenient, user-friendly interface to the large-scale automatic forecasting, model building, and time series exploration capabilities available in the SAS System. This new user interface addresses the needs of novice forecasters by being largely automated, yet still meets the needs of more experienced analysts by providing layers of sophistication that can be accessed as needed. SAS Forecast Studio enables users to set up forecasting projects, perform large-scale automatic forecasting, identify exceptions, override forecasts, and construct their own models if desired. Given the scale of many forecasting problems, manually customizing many statistical models may not be feasible. SAS Forecast Studio provides an automated system that selects appropriate models and intelligently chooses influential variables that improve the model fit. The system supports hierarchical forecasting processes by providing top-down, middle-out, and bottom-up forecast reconciliation and generates SAS code so that projects can be run in a batch environment.

INTRODUCTION

Within an organization, generating forecasts is an important first step in many planning and decision-making processes. Forecasting is typically performed on a regularly repeating basis for a wide variety of planning purposes. For example, future demand for products and services may be forecast in order to support production planning, marketing activities, resource scheduling, and financial planning.

Forecasters often follow the same iterative process with each forecasting and planning cycle. The process usually involves generating updated forecasts based on the most recent data, reconciling forecasts in a hierarchical manner, identifying and remedying problematic forecasts, adding judgmental overrides to the forecasts based on business knowledge, and publishing the forecasts to other systems or as reports.

Upon completion of the forecasting process, the planning process determines what actions the organization will take in light of the forecasts. Planning processes not only take into account the forecasts, but also the constraints upon the business and overall corporate goals. Much of an organization's time and effort is spent in the planning process. Improving the reliability of statistical forecasts that feed these processes can result in huge rewards. Improved forecasting often leads to greater operational efficiency, reduced expenses, and increased profits.

Although forecasting is an important function, many organizations rely on a relatively small number of forecasters to generate large numbers of statistical forecasts. Given the relatively small number of forecasters in most organizations, a large degree of automation is often required in order to complete the forecasting process in the time available each planning period. SAS Forecast Studio provides large-scale automatic forecasting coupled with analytical sophistication, making the forecasting process more manageable. This paper shows how SAS Forecast Studio facilitates forecasting by walking you through the steps in a typical forecasting process, illustrating key features along the way.

TARGET AUDIENCE

SAS Forecast Studio has been designed to enable the novice forecaster to quickly move through the statistical forecasting process in a largely automated fashion. The forecaster does not need to know any SAS programming or how to build time series models; the system can be entirely automated. Of course, more benefit may be gained from the system as the user's knowledge of forecasting methods grows.

Though the user interface is streamlined to enable less-experienced forecasters to move quickly through the forecasting process, the application has all of the depth that more advanced analysts expect. For those who wish to spend more time on analysis and model building, SAS Forecast Studio provides an array of easy-to-use diagnostic and model-building tools, enabling more experienced analysts to benefit from the productivity of the user interface.

GETTING STARTED

All work done in SAS Forecast Studio is stored in projects. A SAS Forecast Studio project contains information about the source data used for forecasting, specifics about the forecasting process, user-specified preferences, any notes entered by users, and the resulting SAS Forecast Studio project code and output data sets. It should be noted that the source data is not stored within the project itself and does *not* need to be moved or copied to where projects are stored. The information in project folders consists mostly of SAS programs, SAS data sets, and XML files.

SETTING UP A PROJECT

To help users get started, SAS Forecast Studio includes a wizard to guide users through the steps required in setting up a project. The steps include:

- Selection of the input data set to be forecast
- Assignment of variable roles
- Set-up of a forecasting hierarchy (optional)
- Specification of project-wide forecasting options
- Specification of a location where the project is to be saved
- Creation of event variables for use in the forecasting process (optional)
- Creation of filters to be used to identify exception forecasts (optional)

SELECTING A DATA SET

Users can select a data set by navigating through SAS libraries or by browsing folders. There is no limit on the size of the data set or the number of series to be forecast. There is also no limit on the length of history for each series. In fact, for forecasting, more history is often preferable.

ASSIGNING VARIABLE ROLES

The role to which a variable in the input data set is assigned determines its usage throughout the project. Figure 1 illustrates how users assign variables to the different roles in a project. A date, time, or datetime variable is required for any forecasting project. If one is not available in the input data set, users are provided with multiple options for creating one. Classification, or BY variables, can be used to obtain separate analyses for groups of observations defined by the BY variables or they can be used to set up a hierarchical forecasting process. Numeric variables can serve as either dependent or independent variables. A dependent variable is required. In the case of a hierarchical forecasting project, only one dependent variable can be specified; otherwise multiple dependent variables can be specified. Independent variables are optional.

By clicking on the More Roles button, additional variable roles are available to the user. Users can specify which variables, if any, are to be used for reporting purposes. These variables will not be used in the analysis, but may be useful to the forecaster when generating reports. Additionally, users can specify variables to be used to adjust the dependent variable. An adjustment variable is used to transform the dependent variable before forecasting takes place and then transform the dependent variable again after forecasting has occurred. (For example, the user may specify a variable to be used to deflate the dependent variable or to convert the dependent variable to a different currency.)

Often forecasts are available from other sources. Users can assign these forecasts to the role of externally generated forecasts and SAS Forecast Studio will evaluate the performance of external forecasts in the same manner as forecasts generated by SAS Forecast Studio itself. In this way, users can compare the performance of forecasts from other third-party systems or judgmental forecasts within the organization to statistical forecasts generated by SAS Forecast Studio.

SETTING UP A FORECASTING HIERARCHY (OPTIONAL)

BY variables have a special usage in the case of a hierarchical forecasting project. Once BY variables are assigned in the project set-up wizard, they are used to construct the levels of a forecasting hierarchy. The ordering of the BY variables determines the structure of the hierarchy. Based on the ordering of BY variables, SAS Forecast Studio aggregates data in the data set to create the levels of the hierarchy. As shown in Figure 2, users have the option of previewing the layout of the hierarchy before the actual hierarchy is generated using the entire data set.

When aggregating data up the hierarchy, users can choose which statistic to apply to numeric variables in the project. In most cases it will be natural to sum the time series, but in a few cases it may be more appropriate to aggregate the series by averaging them. For example, if one were to use a variable such as price as an independent variable, it might make more sense to aggregate the average price up the hierarchy.
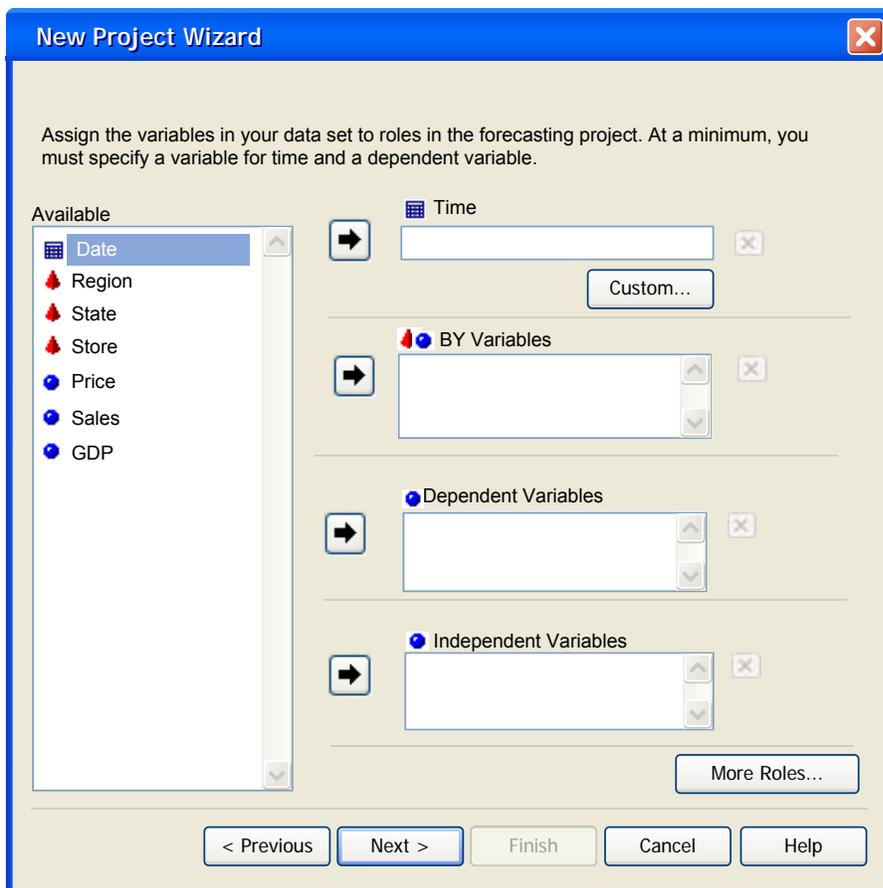
Figure 1: Assigning Variable Roles in the Set-up Wizard
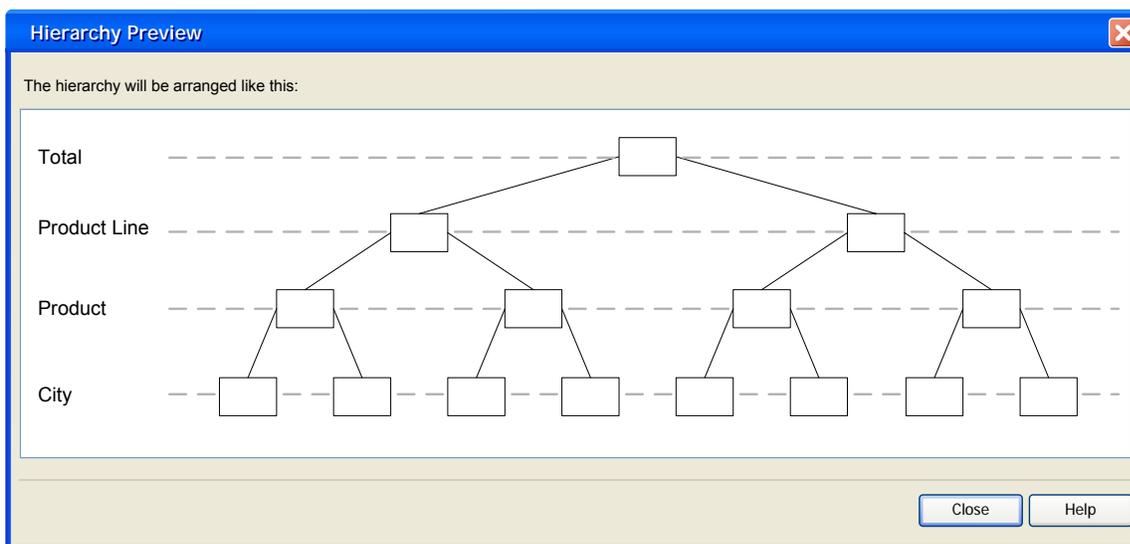
Figure 2: Hierarchy Preview Window

By default, SAS Forecast Studio produces statistical forecasts at each level of the hierarchy using the aggregated data. Due to the natural variance of most data, it is highly unlikely that forecasts generated using aggregated data at a higher level of the hierarchy will equal the sum of forecasts generated using individual series at a lower level of the hierarchy. However, when the forecasts are to be used for reporting or planning purposes, it is desirable for the forecasts to reconcile up and down the hierarchy.

SAS Forecast Studio performs hierarchical forecast reconciliation using either top-down, bottom-up, or middle-out techniques. If the forecaster specifies the bottom-up reconciliation method, forecasts generated at the lowest level of the hierarchy are aggregated to produce forecasts at higher levels of the hierarchy. If top-down reconciliation is specified, forecasts are apportioned down the hierarchy starting from the highest level. For middle-out reconciliation, the user selects the level of the hierarchy from which a combination of both top-down and bottom-up reconciliation will be performed.

In the case of top-down and middle-out reconciliation, multiple options are available for disaggregating forecasts from higher levels of the hierarchy down to lower levels. As shown in Figure 3, users can choose to use either the historical or forecast values of the time series at the lower level to generate time-varying percentages by which forecasts at the next higher level are divided. Users can also use simple static percentages to disaggregate forecasts if they desire. SAS Forecast Studio calculates the disaggregation percentages automatically.

Figure 3: Hierarchy Set-up Options

CUSTOMIZING PROJECT SETTINGS (OPTIONAL)
The project set-up wizard allows users to change forecasting options for the entire project.
Among the options available, users can specify the following:

- How missing values in the data are to be interpreted
- The sensitivity of automatic outlier detection
- How leading and trailing zeros in the time series are to be treated
- Whether or not forecasts are allowed to go below zero
- The significance level of forecast confidence limits
- Holdout sample length

CREATING EVENT VARIABLES (OPTIONAL)
Users can define event, or intervention, variables to be added to the models during the automatic forecasting process. Events variables indicate when anything out of the ordinary occurred in the past or will occur again in the future. These variables enable the models to adjust appropriately to any past or future interruptions. Examples of events might be such things as advertisements, natural disasters, strikes, and so on. The event management dialogs are discussed later in the paper.

CREATING FILTERS (OPTIONAL)
The project set-up wizard makes it easy for users to set up rules for finding exception forecasts. As illustrated in Figure 4, among other things, users can set up rules for finding forecasts where the goodness-of-fit statistic exceeds a certain threshold value specified by the user. Once the series in the project has been forecast, users can use the filters to quickly identify exceptions and work on the models and forecasts that need the most attention.

Figure 4: Defining Exception Rules

AUTOMATIC FORECAST GENERATION
Upon closing the project wizard, SAS Forecast Studio automatically generates statistical forecasts for all series in the project. If a hierarchy has been specified, the system reconciles the statistical forecasts according to the options specified by the user in the project set-up wizard.

TIME SERIES DIAGNOSIS AND MODEL CONSTRUCTION
In order to automatically generate forecasts, SAS Forecast Studio must first determine an appropriate model for each time series being forecast. The system uses several statistical tests and heuristics to construct candidate models for each series. SAS Forecast Studio tests for intermittency of the dependent variable, unit and seasonal unit roots, functional transformation of the dependent variable, outliers, and structural shifts. If a series is determined to be intermittent, SAS Forecast Studio fits an intermittent demand model. Otherwise, the system constructs the appropriate exponential smoothing, ARIMAX, or unobserved components model. Users can control the sensitivity of the diagnostic tests.

If any independent or event variables were specified in the set-up wizard, SAS Forecast Studio automatically selects from among them those variables that improve the fit of the model. The system performs many of the same diagnostic tests on the independent variables as are performed on the dependent variable(s). An important feature of SAS Forecast Studio is its ability to determine whether any of the independent variables are *dynamically* related to the

dependent variable and to specify the model accordingly. In the language of time series analysis and forecasting, SAS Forecast Studio automatically identifies the transfer function specification for each independent variable in a given model. For example, if one were forecasting the demand for a product that has a multi-period response to a change in price, SAS Forecast Studio would be able to detect and specify this dynamic relationship in the forecasting model.

The final model is selected based on a goodness-of-fit statistic chosen by the user. If a holdout sample is specified, the model is selected based on the goodness-of-fit statistic in the holdout region (out-of-sample fit); otherwise in-sample fit is used to select the model. Model parameters are automatically optimized to provide the best fitting model. If a holdout sample was specified, the model parameters are estimated using the full range of data including the holdout sample. Users may request the system to keep the n best fitting models.

The following goodness-of-fit statistics are available in the system:

- Total Sum of Squares (Uncorrected)
- Total Sum of Squares (Corrected)
- Sum of Square Errors
- Mean Square Error
- Root Mean Square Error
- Mean Absolute Percent Error
- Mean Absolute Error
- R-Square
- Adjusted R-Square
- Amemiya's Adjusted R-Square
- Random Walk R-Square
- Akaike's Information Criterion
- Schwarz Bayesian Information Criterion
- Amemiya's Prediction Criterion
- Maximum Error
- Minimum Error
- Maximum Percent Error
- Minimum Percent Error
- Mean Error
- Mean Percent Error

VIEWING FORECASTING RESULTS

Results from the automatic forecasting process are displayed in the Forecast Summary window. The historical data, fitted model, and confidence limits are displayed graphically, as shown in Figure 5. More information about each forecast can be seen by expanding the data table below the forecast plot.

The data table contains the following:

- The historical series
- Any variables used as regressors in the model(s)
- The fitted model(s)
- The final selected model to use for forecasting
- The goodness-of-fit statistics for the model(s)
- The reconciled forecast in the case of a hierarchical project
- A data-entry line for manually adjusting the statistical forecast as necessary

Users can choose whether or not to display all models fitted to the data in the table or to only show the final model to be used.

To quickly navigate through the forecasts in a project, SAS Forecast Studio provides a Tree View of the project, as shown on the left side of Figure 5. If a project is hierarchical, the levels of the hierarchy can be collapsed or expanded in the Tree View.
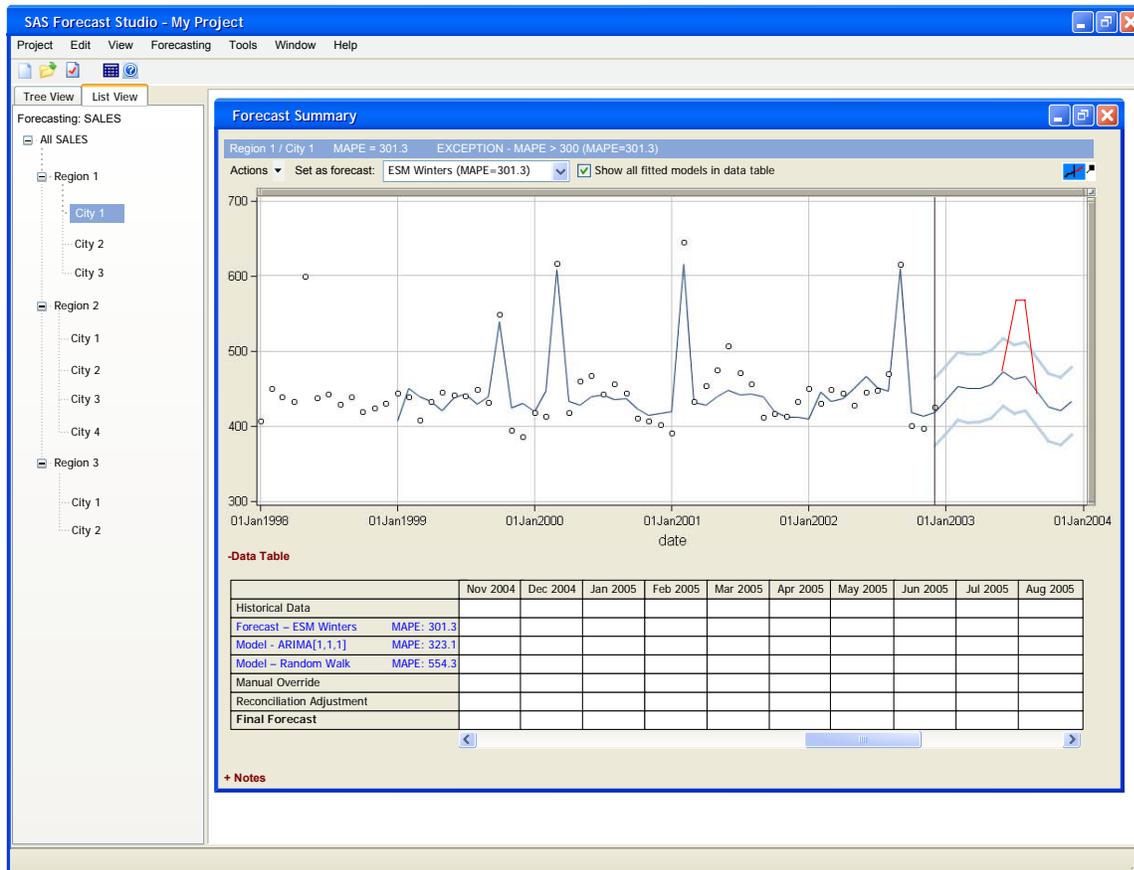
Figure 5: Forecast Summary

IDENTIFYING EXCEPTION FORECASTS

For large-scale forecasting problems, forecasters may not have time to review or work on all of the forecasts. To best utilize their time, forecasters need to quickly identify and address exception forecasts first.

USING THE LIST VIEW

The List View provides an alternate way of viewing the series in a project. The series are presented in a table with the series properties serving as the columns of the table, as shown in Figure 6. The List View can be customized by the forecaster to show additional series properties including goodness-of-fit statistics, the number of missing and nonmissing observations, start dates, end dates, and so on.

The List View provides a quick and easy way of sorting and filtering the series to identify exceptions. Clicking on a column heading sorts the entire project according to that column. Users can identify series with little or no history, or series with poor fitting models, by clicking on the number of observations or goodness-of-fit statistic columns, respectively. When the user selects a row in the table, SAS Forecast Studio navigates to the location of that series in the project so that the user can begin working on fixing problematic forecasts right away.

APPLYING FILTERS

A drop-down menu in the List View enables users to subset the series in the table according to hierarchy levels, providing a convenient method of gauging model fit and series properties on a given level. The drop-down menu also lists exception rule filters defined at the time the project was created. By selecting the exception rule filter from the drop-down menu, only those series that violate the exception rule are shown in the table view. If filters were not created during the project set-up, they can be defined at any time within a project.
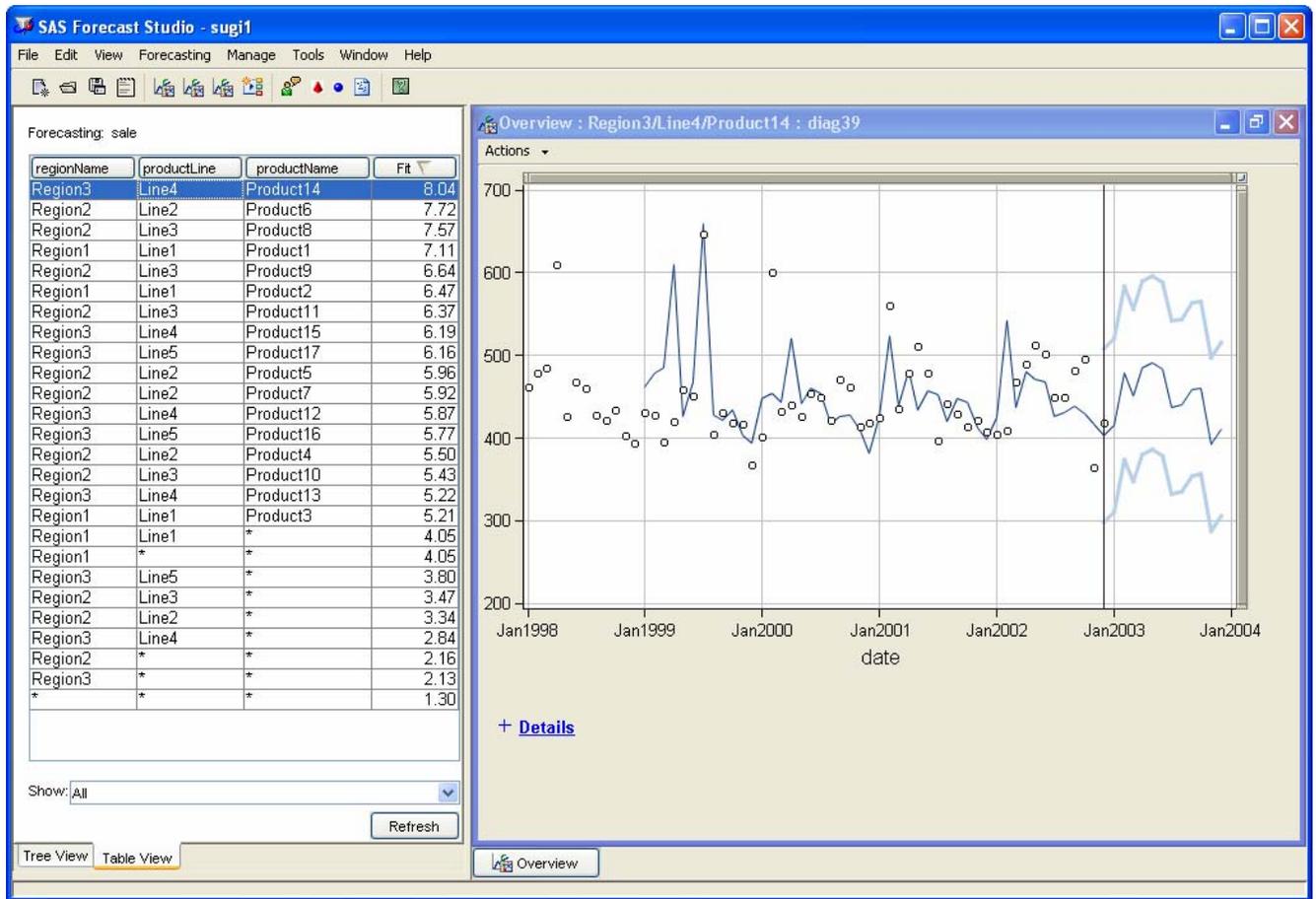
Figure 6: List View

FIXING EXCEPTIONS AND BUILDING MODELS

Exception forecasts may be generated for a variety of reasons. Frequently exceptions occur when little historical data exists, or if additional knowledge relevant to the forecast exists but is not incorporated into the data or modeling process. For these reasons and others, some automatically generated models may not fit the data well or may have violated other exception rules defined by the user. There are multiple approaches for fixing exception forecasts. The approach taken will depend on the modeling skills of the forecaster and the amount of time available to spend in adjusting forecasts.

MANUAL OVERRIDES

If a forecaster has limited modeling experience or if little time is available, forecasters may choose to apply their best judgment and knowledge to adjust the statistical forecasts manually. SAS Forecast Studio provides a line in the data table of the Forecast Summary area for the forecaster to manually override the statistical forecast. User overrides are stored separately from the statistically generated series, enabling the performance of both to be evaluated at a future time when actual data become available.

EDITING EXISTING MODELS

Users with more experience and time can choose to remedy exceptions by editing the models that were automatically generated by SAS Forecast Studio. Clicking on the model name listed in the data table of the Forecast Summary window opens the appropriate modeling dialog from which users can edit the model. For example, clicking on an ARIMA model automatically generated by the system brings up the ARIMA dialog as shown in Figure 7. From within the modeling dialog, users can edit any of the model parameters, add regressors or events to a model, and so on. The edited model is then fit to the data and added to the summary data table.
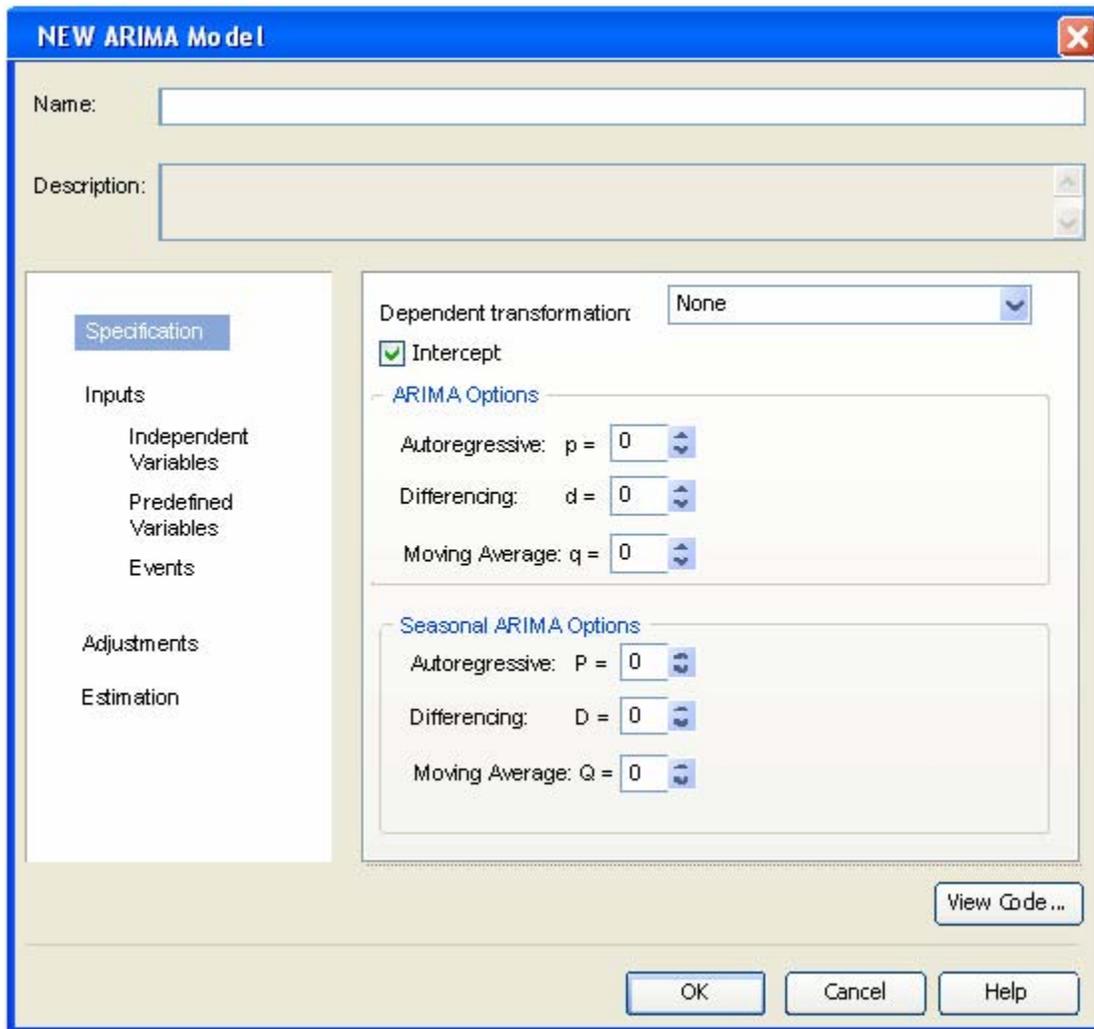
Figure 7: Example of Modeling Dialog

CREATING NEW MODELS

Forecasters may wish to create an entirely new model. When creating a new model, experienced forecasters like to view plots of various aspects of the series in order to know how to begin the modeling process.

The Series Analyzer, shown in Figure 8, provides the following plots:

- Autocorrelations of the series (and standardized autocorrelations)
- Partial autocorrelations of the series (and standardized partial autocorrelations)
- Inverse autocorrelations of the series (and standardized inverse autocorrelations)
- Seasonal decomposition components
- White noise probabilities for the series (also on log scale)
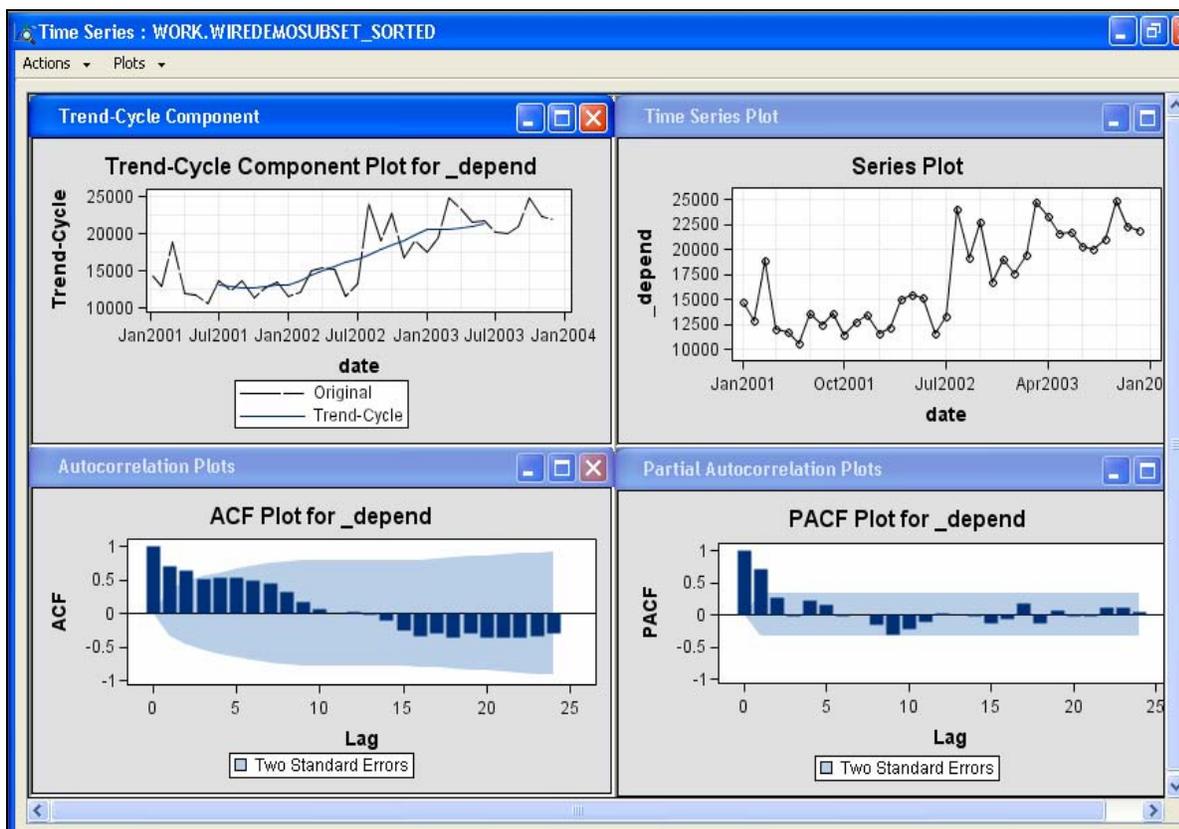
Figure 8: Example Series Analyzer Plots

SAS Forecast Studio provides modeling dialogs to help users create virtually any form of ARIMA, subset and factored ARIMA, unobserved components, exponential smoothing, multiple linear regression, or intermittent demand model (including Croston's method). Users can also fit any type of curve to the data. Selecting the desired model type from the list shown in Figure 9 brings up the appropriate modeling dialog.
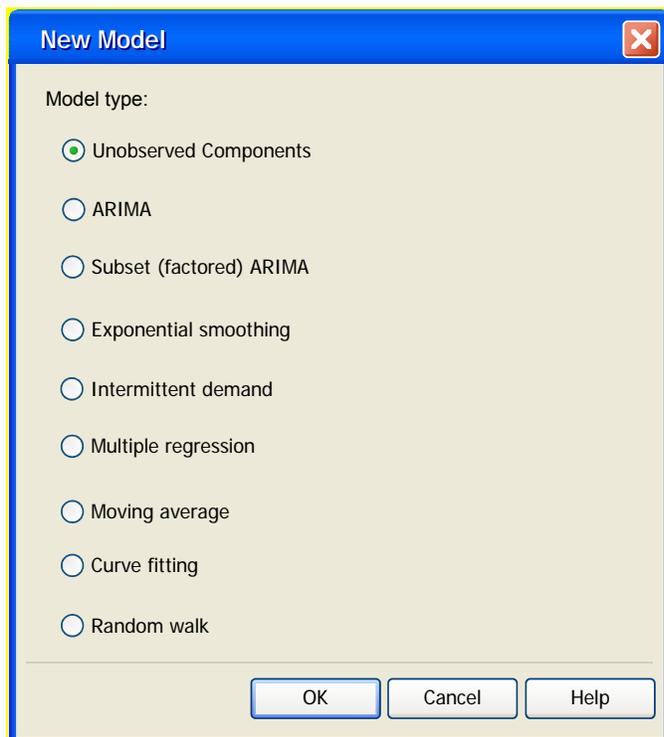
Figure 9: Selecting a New Model Type

DEFINING AND ASSIGNING EVENTS TO SERIES
Some models may be improved by the inclusion of events. An event, also called an "intervention," is defined as anything out of the ordinary that interrupts the time series (promotions, strikes, law changes, storms, etc.). Events can be added to regression, ARIMA, and unobserved components models.

It is not necessary for users to store event information in the source data set, since the events manager in SAS Forecast Studio enables users to create, edit, combine, or delete events. Events defined in SAS Forecast Studio are generated in memory at the time of analysis. This prevents data entry errors, reduces data storage space, and always keeps events up to date even if the input data set changes.

When creating an event, users can select the appropriate shape. For example, an event may be very short-lived (such as a power outage) and can be modeled using a pulse of limited duration, or an event may represent an ongoing change (such as a law change), and a permanent level shift would be more appropriate. Figure 10 illustrates the choices of intervention shapes.
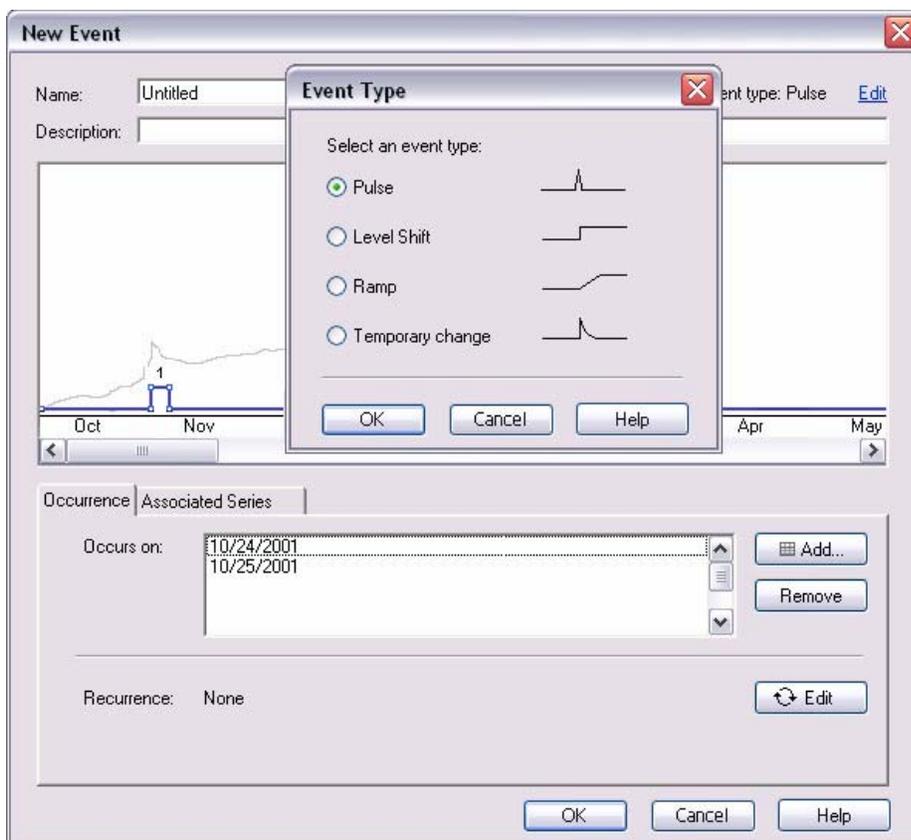
Figure 10: Choosing the Event Shape

The event building dialog displays the appropriate event selection interval to correspond with the frequency of the series in the input data set. If the series are hourly, users can select the hours that correspond to the event; likewise for daily, weekly, or monthly data, and so on. As shown in Figure 11, SAS Forecast Studio provides several predefined holidays so that users do not need to define these themselves. This is especially useful for moving holidays such as Easter.
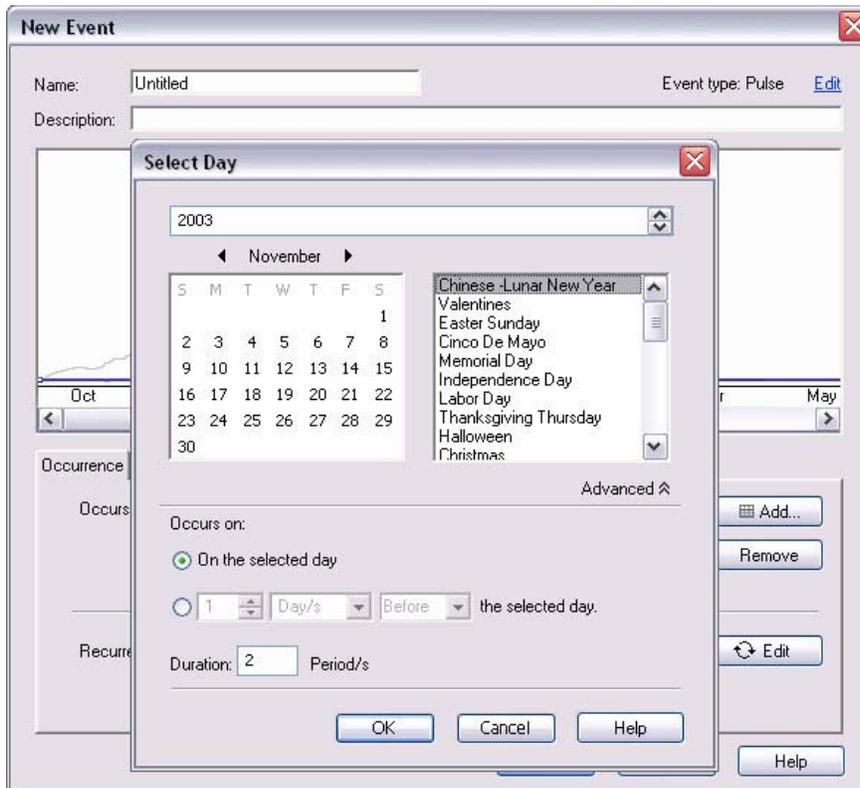
Figure 11: Event Date Selection

Once defined, events can be assigned to any of the series in the project. A filter dialog is available for quickly specifying which series in a project should be assigned a particular event. Figure 12 shows how users can choose the appropriate series from a list.
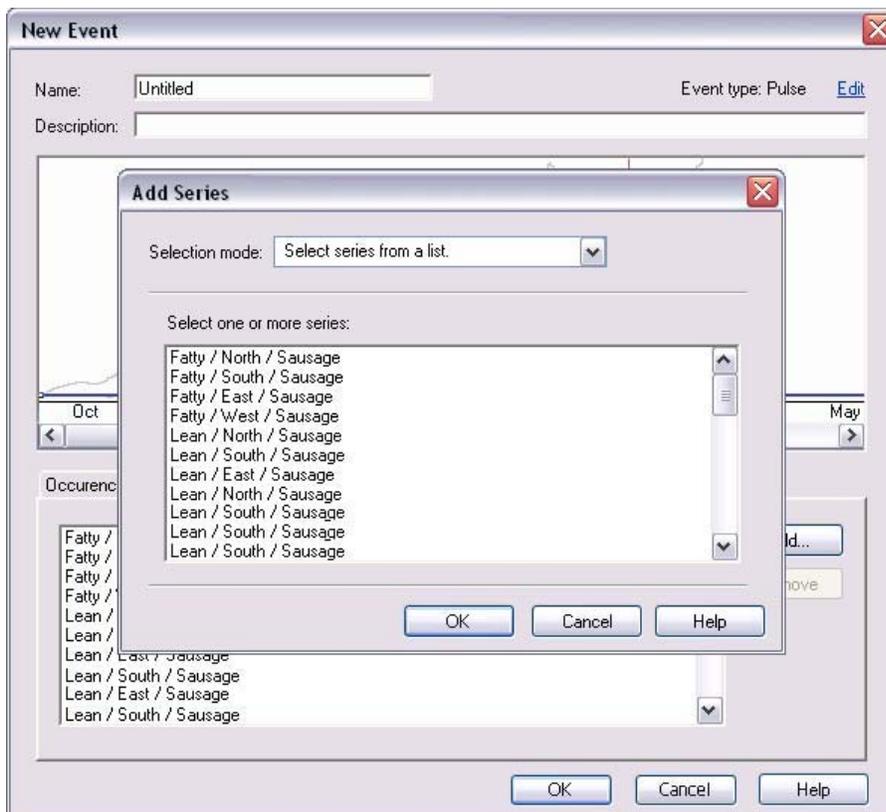
Figure 12: Assigning Events to Series

EVALUATING MODELS

The Model Analyzer provides several diagnostic plots to help forecasters determine how well a given model is fitting the data.

Users can view the following plots from within the Model Viewer window:

- Prediction errors
- Prediction error distribution
- Autocorrelations of the prediction errors (and standardized autocorrelations)
- Partial autocorrelations of the prediction errors (and standardized partial autocorrelations)
- Inverse autocorrelations of the prediction errors (and standardized inverse autocorrelations)
- White noise probabilities for the prediction errors (also on log scale)
- Components of the series (smoothed trend, season, and level states)

When the model is changed, the plots in the Model Analyzer automatically update, enabling users to quickly and easily evaluate the changes made to the model.
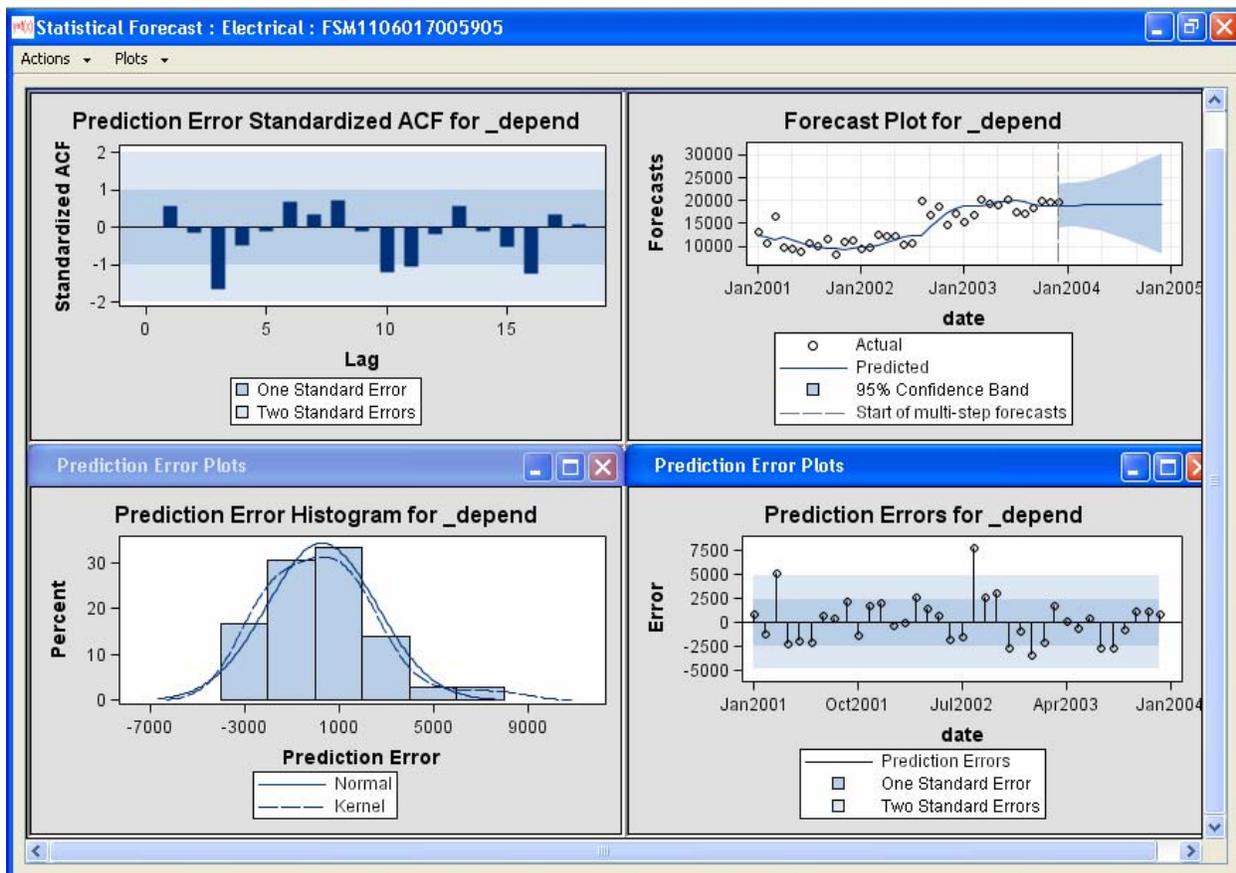
Figure 13: Model Analyzer

CUSTOMIZING THE PROJECT MODEL REPOSITORY

Models constructed or edited by the user can be added to the project model repository for future use throughout the project. In this way, the model repository can become increasingly tailored to the business or forecasting problem over time. When creating projects, users can bypass the automatic model building process and require the system to use a specific model repository for best-pick model selection.

UPDATING THE FORECASTS

The first time SAS Forecast Studio produces forecasts for a project, it must go through three steps: model selection, parameter estimation, and forecast generation. Of the three steps, model selection takes the majority of the processing time. To speed the process, users can choose to skip the model selection step upon subsequent re-forecasting of the project. Users may also choose to keep both the existing models and the model parameter estimates and simply reforecast the project or individual series, as shown in Figure 14.
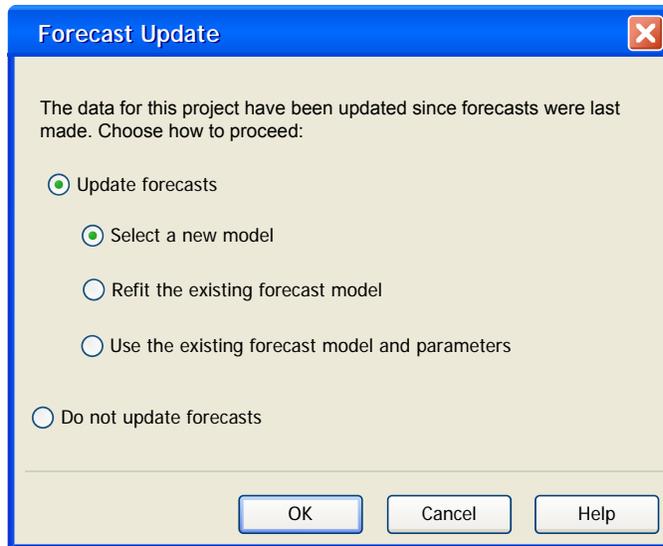
Figure 14: Forecast Update Options

PUBLISHING RESULTS

In addition to the forecasting information surfaced within the application, SAS Forecast Studio produces several output data sets that can be used outside the system for querying and reporting, or can be fed into planning systems.

The following output is available:

- Forecasts
- Prediction errors (and standard errors)
- Forecast confidence intervals
- Selected model
- Parameter estimates, estimate standard errors, and significance tests
- Variable transformations
- Numerous statistics of fit
- Summary information about the series (series min, max, standard deviation, number of observations, etc.)
- Project notes

REPORTING

SAS Forecast Studio provides several default reports to help the forecaster evaluate the forecasting process and present the results to others.   The reporting in SAS Forecast Studio is based on SAS stored process capabilities, making it an open and extensible system. Users can author their own reporting stored processes and add them to the system. Reports provided by SAS Forecast Studio or authored by the user show up in the report selection dialog displayed in Figure 15.

14
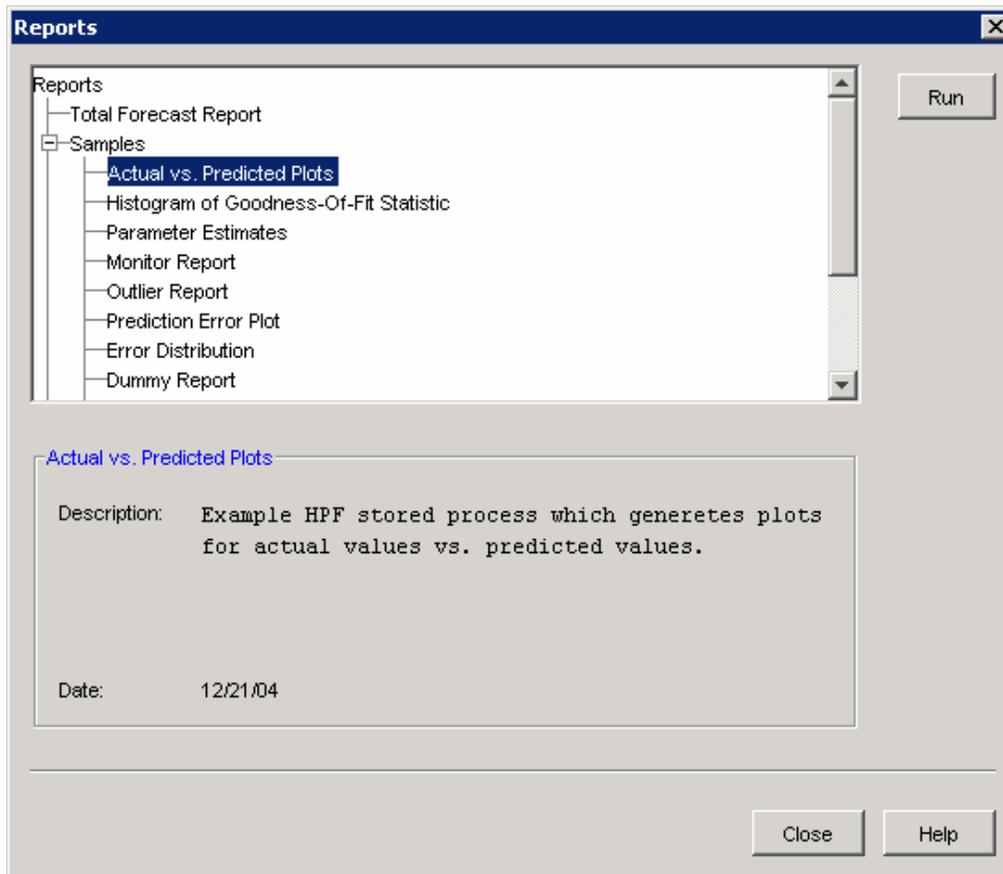
Figure 15: Example of SAS Forecast Studio Reports

EXPORTING CODE
SAS Forecast Studio is a code-generating user interface. Figure 16 shows the code window in SAS Forecast Studio with an example of SAS code generated automatically by the system. All SAS code generated by SAS Forecast Studio can be saved and run in a batch environment. Users can use the SAS code generated by the user interface in any manner that they can use any SAS program. For example, users may wish to schedule the project to run on a regular basis or to create a stored process that can be accessed from SAS® Enterprise Guide® or Microsoft Office products.
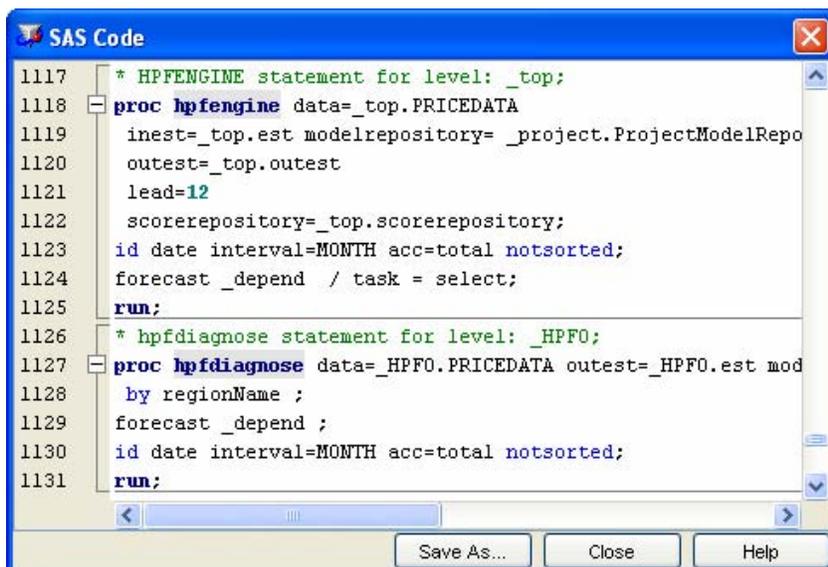
Figure 16: Example of SAS Code Generated by Forecast Studio

CONCLUSION

SAS Forecast Studio surfaces the sophisticated time series exploration, model building, and large-scale automatic forecasting capabilities of the SAS System in an easy-to-use interface. The application supports the needs of both novice forecasters who need to move through the production forecasting process quickly as well as more experienced forecast analysts who can use the application as a productivity tool. SAS Forecast Studio is extensible and flexible. As a code-generating user interface, projects can run in batch, and reporting in the application can be extended using stored processes. Several output data sets are available for querying and reporting or publishing to planning systems. SAS Forecast Studio can be used in concert with other technologies from SAS, such as SAS® ETL Studio, SAS® Add-In for Microsoft Office, and SAS® Enterprise Guide to create a forecasting solution that can serve the needs of the entire organization.

REFERENCES

Leonard, Michael (2002), "Large-Scale Automatic Forecasting: Millions of Forecasts," Cary, North Carolina: SAS Institute, Inc.

Leonard, Michael (2004), "Large-Scale Automatic Forecasting with Inputs and Calendar Events," Cary, NC: SAS Institute Inc.

Leonard, M. J. (2003), "Mining Transactional and Time Series Data," International Symposium of Forecasting.

Leonard, M. J. (2004), "Predictive Modeling Markup Language for Time Series Models," Cary, NC: SAS Institute Inc.

Leonard, Michael (2000), "Promotional Analysis and Forecasting for Demand Planning: A Practical Time Series Approach," Cary, North Carolina: SAS Institute Inc.

RECOMMENDED READING

Box, G. E. P, Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control,* Englewood Cliffs, NJ: Prentice Hall, Inc.

Brocklebank, John C. and Dickey, David A. (2003), *SAS for Forecasting Time Series*, Second Edition, Cary, NC: SAS Institute Inc.

Makridakis, S. G., Wheelwright, S. C., and Hyndman, R. J. (1997), *Forecasting: Methods and Applications,* New York: John Wiley & Sons, Inc.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Brenda Wolfe
SAS Institute
SAS Campus Drive
Cary, NC 7513

(919) 531-6326
brenda.wolfe@sas.com
www.sas.com