

Ten Underused Features That Improve Your SAS® Forecast Server Workflow

Michael Leonard, SAS Institute Inc., Cary, NC, and Evan Anderson

ABSTRACT

SAS® Forecast Server is one of the most feature-rich forecasting products on the market. This paper describes 10 underused features to improve your workflow. First, start with the data: (1) Use SAS® Time Series Studio to become familiar with your data and define a hierarchy, (2) catch data problems through warnings about the time ID variable, (3) look under the covers by using the SAS® log, and (4) use adjustment variables and start-up code to fix data issues. Next, improve the forecasts: (5) Define recurring events that influence the time series, (6) add models by importing from an external list, (7) use rolling simulations to evaluate forecast accuracy over the number of periods you need to forecast, (8) evaluate the effects of independent variables by using scenario analysis, and (9) gain insight into results by comparing models. Finally, put your workflow into production: (10) Run the code in batch.

INTRODUCTION

More information than ever before is being collected with associated timestamps. Computers, cell phones, smart devices, detectors, and other devices record timestamped data. These timestamped data can be modeled, forecasted, or mined (or any combination of these) for better decision making.

SAS Forecast Server software provides a large-scale automatic forecasting system. The software provides for the automatic selection of time series models for use in forecasting timestamped data. Given a timestamped data set, the software provides the following automatic forecasting process: It accumulates the timestamped data to form a fixed-interval time series, diagnoses the time series by using time series analysis techniques, creates a list of candidate model specifications that are based on the diagnostics, fits each candidate model specification to the time series, generates forecasts for each candidate fitted model, uses a model selection criterion to select the most appropriate model specification on the basis of either in-sample or holdout-sample forecast performance (fit statistic comparison), refits the selected model specification to the entire range of the time series, creates a forecast score from the selected fitted model, generates forecasts from the forecast score, and evaluates the forecast by using in-sample analysis. SAS Forecast Server also provides for out-of-sample forecast performance analysis.

There are many capabilities associated with SAS Forecast Server; this paper explores some of the lesser-known features.

DISCOVER FORECAST SERVER CAPABILITIES

The following sections provide details about 10 underused features that you can use to improve your workflow.

FEATURE 1: USE SAS TIME SERIES STUDIO TO BECOME FAMILIAR WITH YOUR DATA AND DEFINE A HIERARCHY

Many organizations collect large amounts of transactional and time series data, such as sales histories, inventory histories, customer transactions, insurance claim histories, and internet data. Analysts often need to structure the time series data into hierarchical time series at particular frequencies in order to enhance their understanding of the data and improve the accuracy of their analyses. In addition, when analysts have large amounts of data, they often need to segment their time series data to support different analyses on different subsets of their data. SAS Time Series Studio enables you to subset your data by using hierarchical queries, graphical queries, parametric queries, or manual selection.

SAS Time Series Studio enables you to do the following:

- Explore multiple time series simultaneously to better understand your data: You can quickly identify anomalies (such as outliers or missing values) and determine which time series does not look like the others. You can explore the effect of transformations of variables and the effect of different time series.
- Identify series that require specialized methods for analysis: You can identify short series (for example, new products or products that have a short lifecycle) and intermittent time series (for example, a series that contains a large number of zero values).
- Identify and group similar time series: You can quickly identify similar time series and group them to allow for hierarchical modeling.
- Export data source that then can be used by SAS® High-Performance Forecasting, SAS® Forecast Studio, SAS® Enterprise Miner™, and other products.

If you are working with new data or considering a new way of organizing your data into hierarchies, try the powerful SAS Time Series Studio application, which is part of SAS Forecast Server. Like SAS Forecast Studio, SAS Time Series Studio has a **New Project** window, which enables you select a data source that has timestamped observations. If you haven't yet identified an appropriate hierarchy for your project, you've come to the right place. Start with a tentative hierarchy by selecting BY variables, or click **Recommend** to select all character variables in the data source, as shown in Figures 1.1 and 1.2. Next, select a time ID variable that contains date or datetime values. If the automatically detected time interval of the observations is undefined or of higher frequency than you need, specify the desired time interval. SAS Time Series Studio will accumulate the data to this interval for you by using the selected formula (the default is the sum of the values). Assign the role of dependent variable to at least one input variable, and assign other roles if appropriate. Distribution plots of each of the assigned variables can then be viewed, as shown in Figures 1.3 and 1.4.

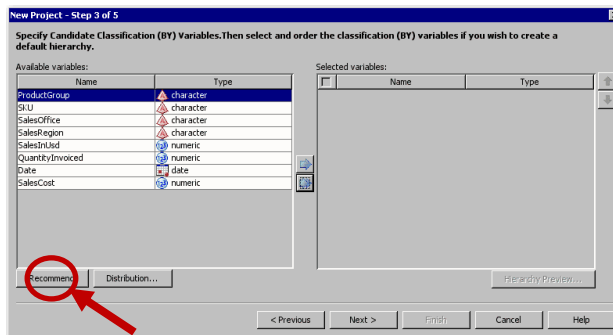


Figure 1.1: New Project, Recommend Button

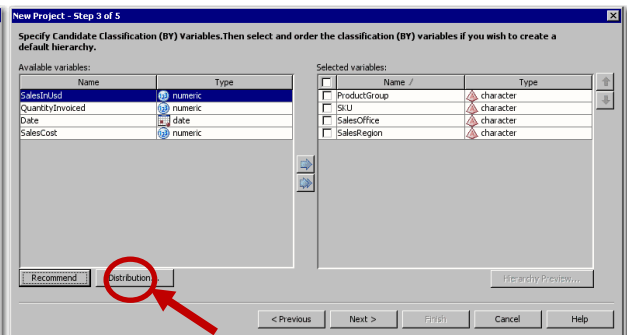


Figure 1.2: Recommended Variables

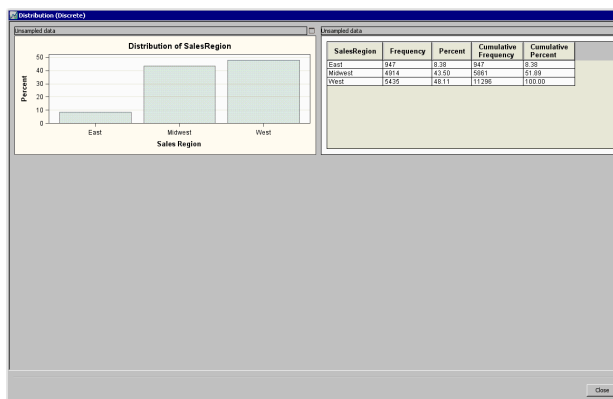


Figure 1.3: Distribution (Discrete)

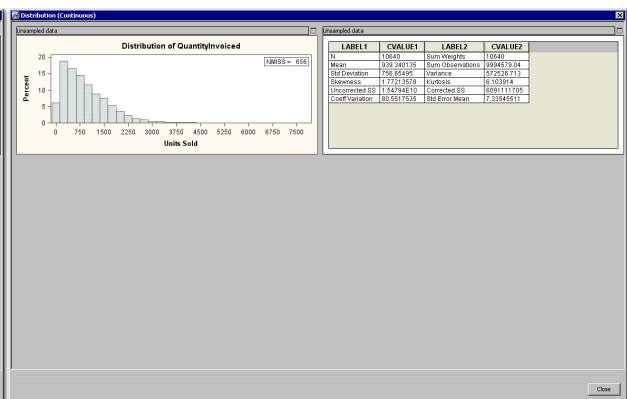


Figure 1.4: Distribution (Continuous)

SAS Time Series Studio provides three views of your project, as shown in Figure 1.5. The **Flow Manager** view shows a tree view of the steps you have taken in constructing the project. The **Selection View**

enables you to select elements of interest. The **Details View** shows graphical and tabular analyses of the selected elements. In the **Details View**, you can view a single series or all series that are contained in a selected node. From the **Display** list, select **Envelope** to see series distribution over time (Figure 1.6), select **Time Series** to see actual values over time (Figure 1.6), or select **Combined** to see both at once (Figure 1.8). The **Series Analysis** tab of the **Details View** provides standard graphs and tables for autocorrelation and seasonal decomposition analysis and enables you to apply transformations (Figure 1.9). The **Multi-Variable Time Series** tab of the **Details View** provides cross-series plots of scatter plots of more than one time series (Figure 1.10 and 1.11).

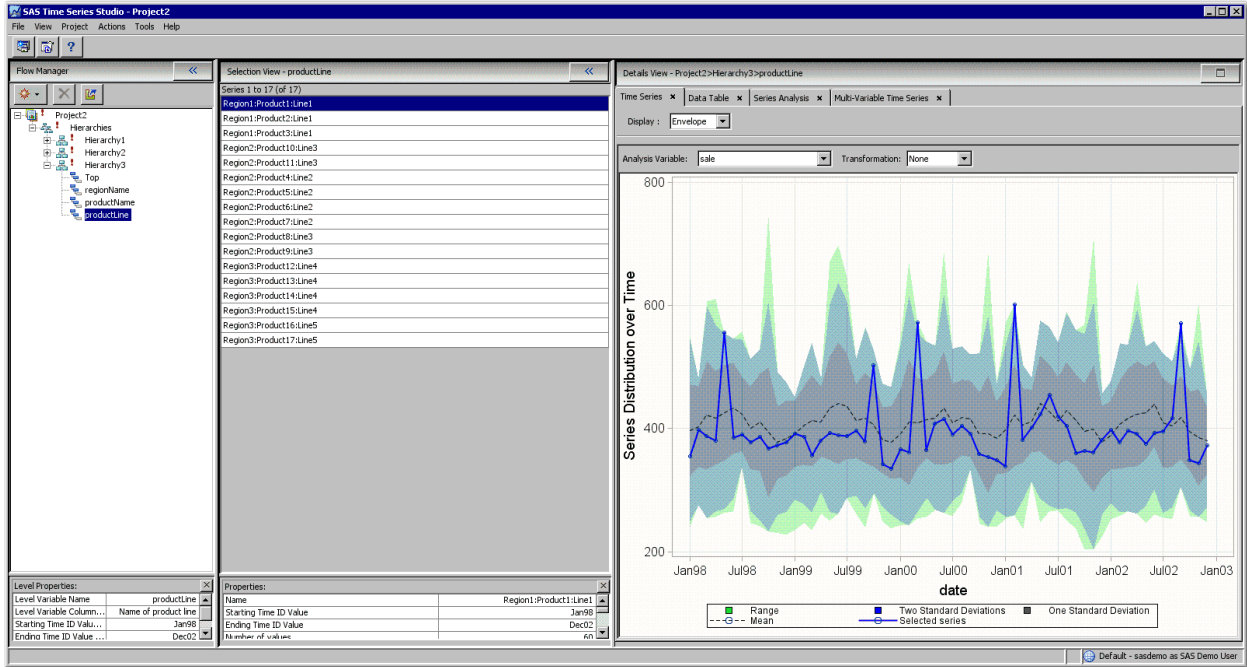


Figure 1.5: Time Series Studio (Flow Manager, Selection View, and Details View)

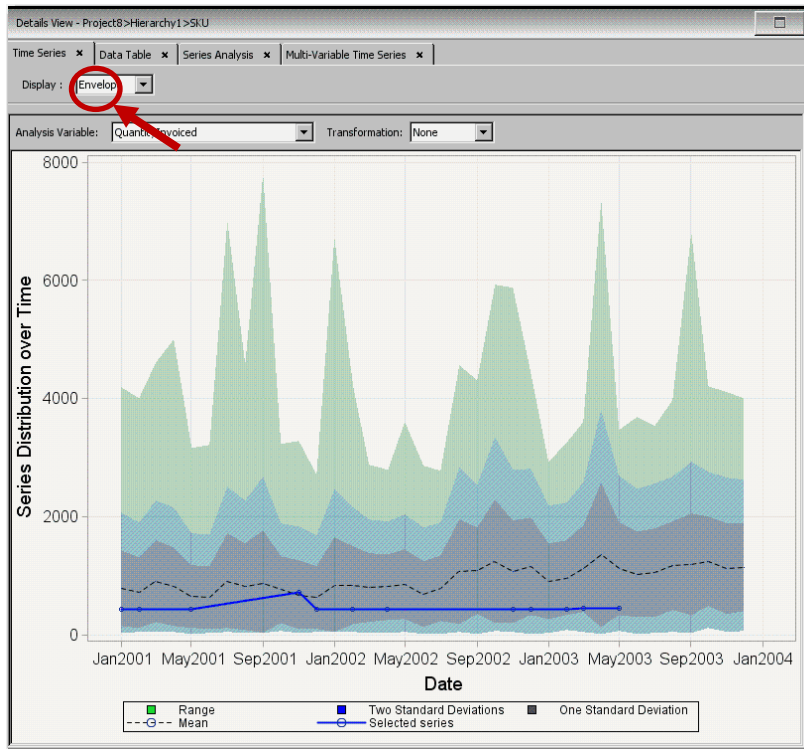


Figure 1.6: Envelope Plots

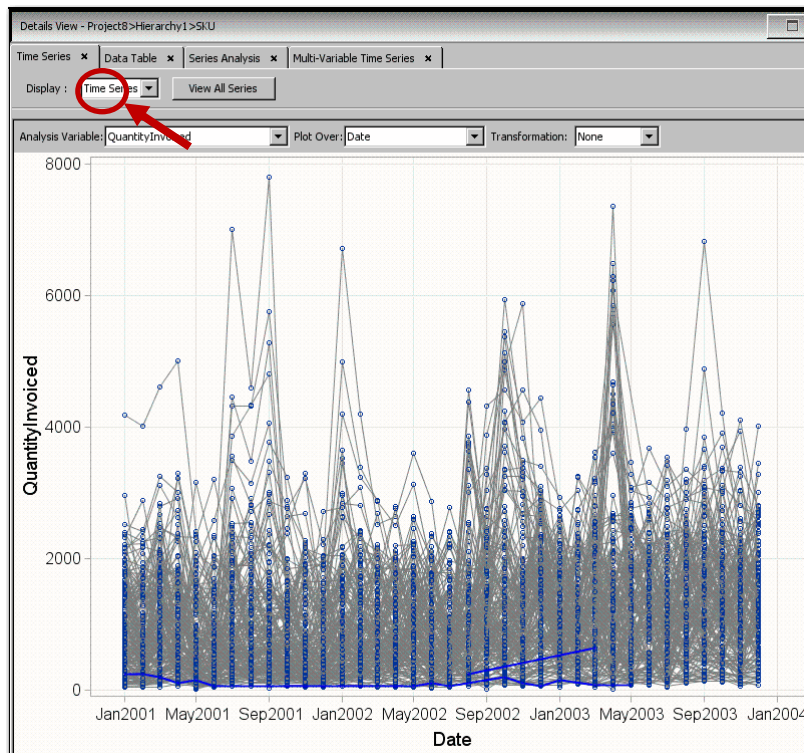


Figure 1.7: Time Series Plots

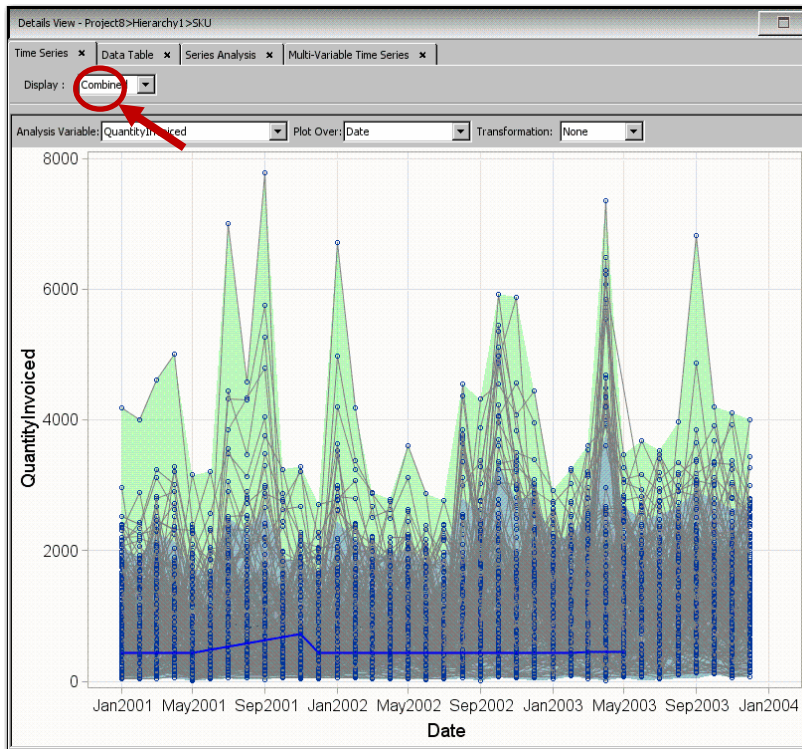


Figure 1.8: Combined Plots

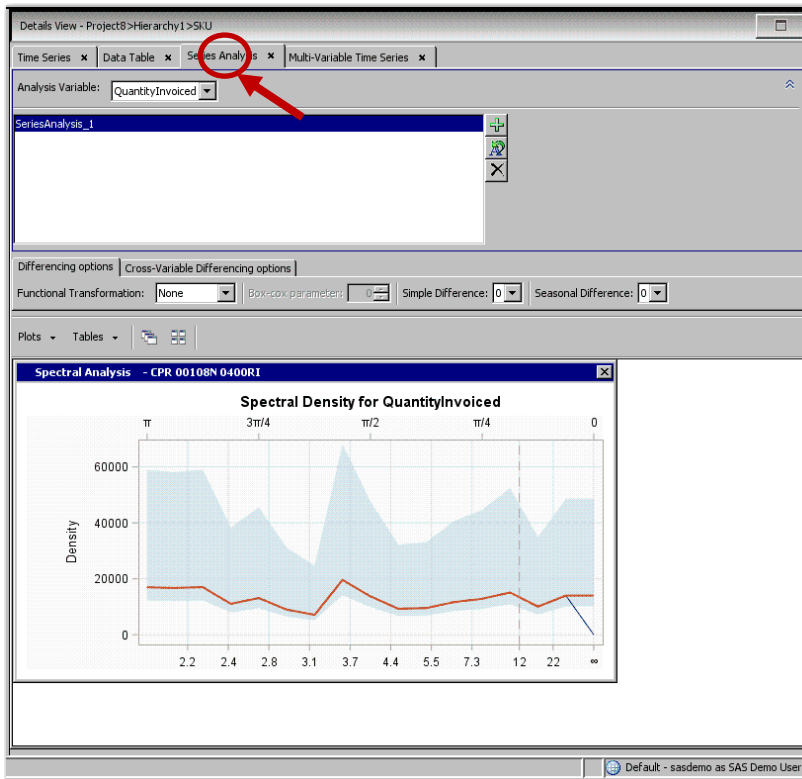


Figure 1.9: Series Analysis Plots

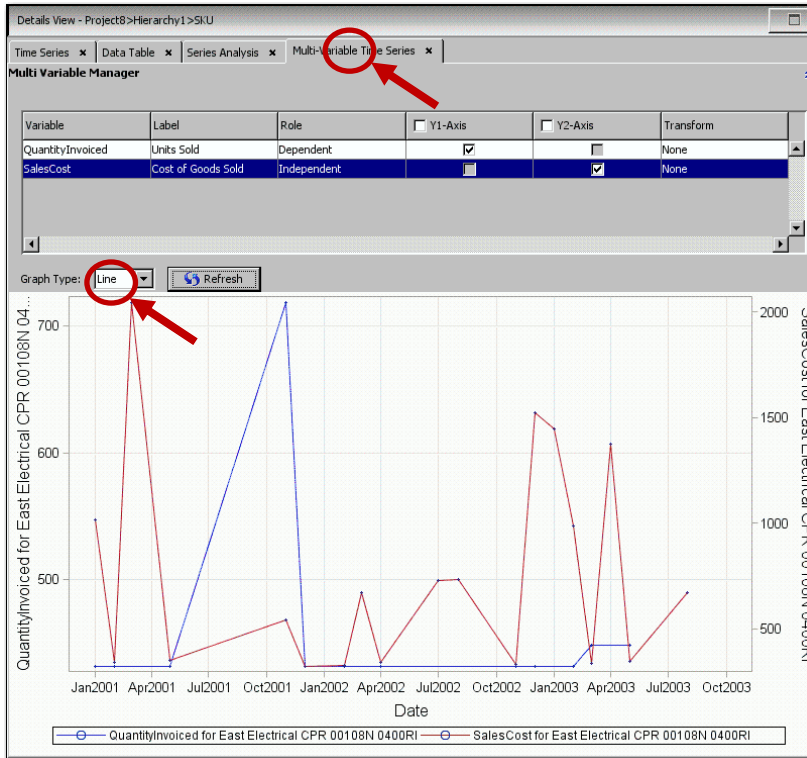


Figure 1.10: Multi-Variable Series, Line Plots

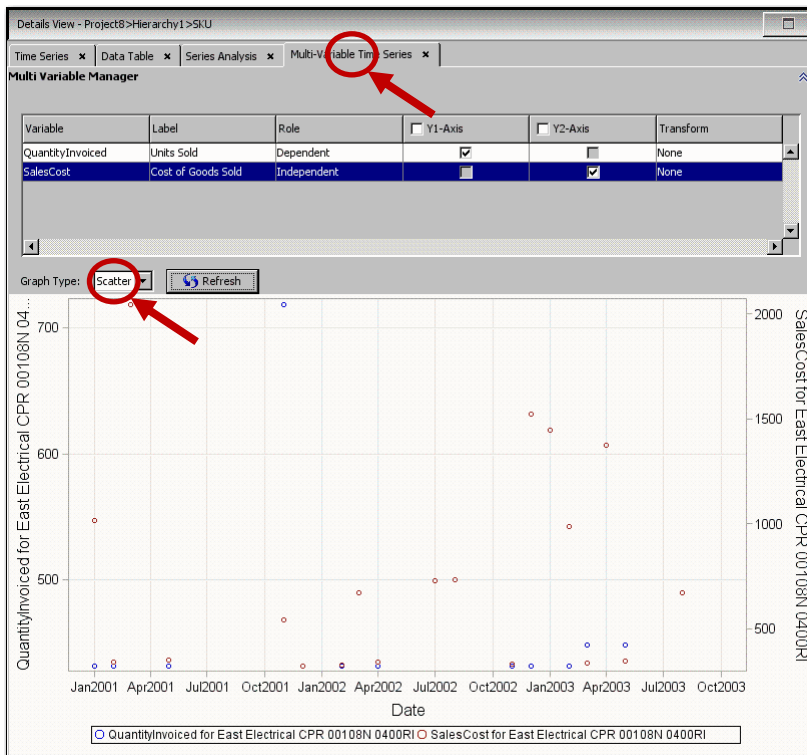


Figure 1.11: Multi-Variable Series, Scatter Plots

One unique feature of SAS Time Series Studio is its ability to structure the data in multiple ways within a project so that you can compare the results within a single view. You set this up by adding one or more

new hierarchies to the **Flow Manager**, selecting BY variables in a particular order, selecting a time interval, and then specifying aggregation and accumulation options. Once you have created the hierarchies you want to explore, the selection view enables you select nodes in any of your hierarchies. The results appear in the **Details View**.

Another unique feature is segmentation, which enables you to identify subsets of the data that behave differently and lend themselves to different forecasting strategies. You can use a graphical query (by drawing a bounding box around a section of a graph) or a parameter query (by entering bounding values on selected variable). Either way, the segments appear in the **Flow Manager**, where you can select them for viewing in the **Details View**. Selected segments can be exported for use in SAS Forecast Studio and other applications.

Once you are happy with your SAS Time Series Studio project, you can use the data to create a Forecast Studio project. You do this by selecting elements in the **Flow Manager** and right-clicking **Export** (which save the results as a SAS data set), and then selecting this data set as input to a new Forecast Studio project. The hierarchy and variable assignments you set up in SAS Time Series Studio will be in effect.

For more information about time series exploration, see *SAS Forecast Studio: User's Guide*.

FEATURE 2: CATCH DATA PROBLEMS THROUGH WARNINGS ABOUT THE TIME ID VARIABLE

Statistical forecasting models assume that the data points in a series occur at evenly spaced points in time. SAS Forecast Server identifies these points in time by the values of the time ID variable that you select. When you select this variable in SAS Forecast Studio, the time values are analyzed to identify the time interval and checked for possible problems in the data. In many cases, SAS Forecast Studio is quite forgiving, but it gives you warnings in some cases so that you can check the data and correct problems if needed. The following warnings are common:

- **Gaps were detected in the values of the Time ID variable.** This warning is issued if the values of the time ID are not evenly spaced. To see more information about this warning, click **Diagnostic graphs** in the warning dialog box. The diagnostics include an offset histogram, which indicates the relative time distance between observations. Any bars that differ from zero indicate an abnormal time ID value. The X axis indicates the position in the data where the problem occurs.
- **The Time ID variable has duplicate values.** This warning is issued if the values of the time ID are not unique. To see more information about this warning, click **Diagnostic graphs** in the warning dialog box. In the diagnostic graphs, one or more bars in the Interval Count histogram will be greater than one.

In most cases, SAS Forecast Studio will produce forecasts even when the time ID is in bad shape. For example, if you use the Sashelp.Workers data and you mistakenly select the variable Electric as the time ID and the variable Date as the dependent series, you will get a forecast, albeit an odd-looking one. But the diagnostic histograms for the time ID will be off-kilter. They should all be flat, and in this case, none are. Figure 2.2 illustrates the **TIMEID Analysis Graphs** dialog box.

One case in which the application is not forgiving is when the time ID values are missing values. The message is "ERROR: Missing or invalid values found." There are no diagnostic graphs and you cannot proceed.

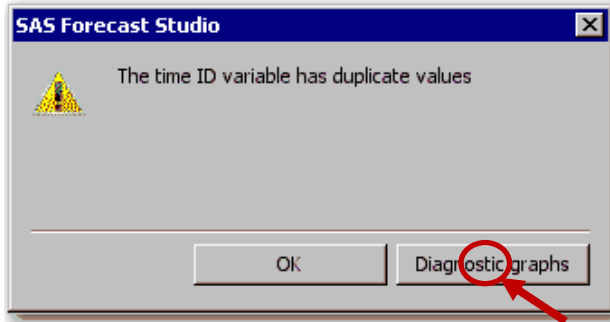


Figure 2.1: Warning Dialog Box

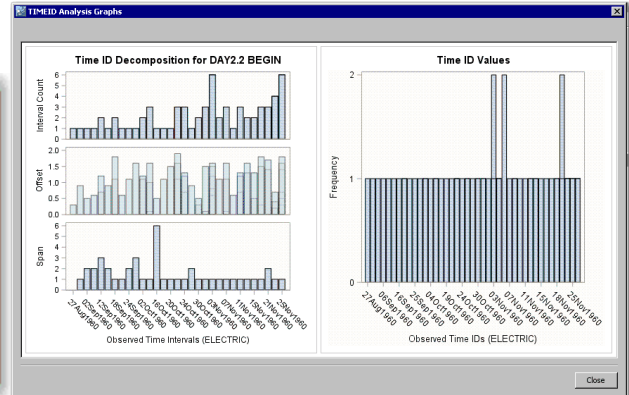


Figure 2.2: TIMEID Analysis Graphs Dialog Box

FEATURE 3: LOOK UNDER THE COVERS BY USING THE SAS LOG

You can view the SAS code that was used to execute the project by selecting **Project ► SAS Code** from the main menu. The **SAS Code** dialog box appears as shown in Figure 3.1. You can save the SAS code to a file from this dialog box.

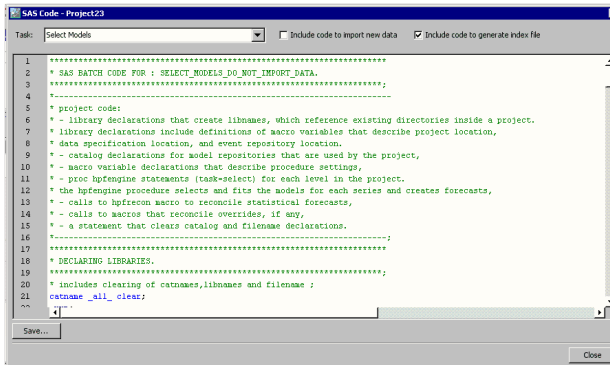


Figure 3.1: SAS Code Dialog Box

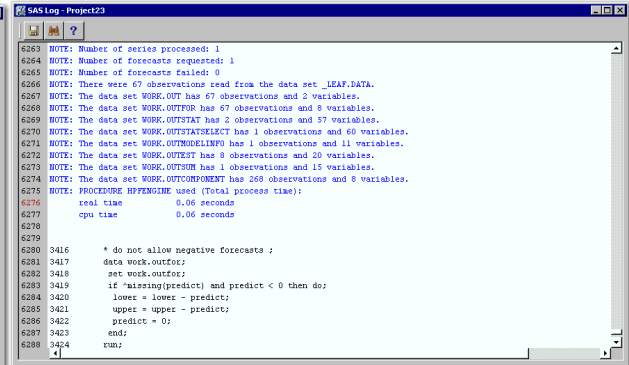


Figure 3.2: SAS Log Dialog Box

You can view the SAS log for an opened project by selecting **Tools ► SAS Log** from the main menu. The **SAS Log** dialog box contains information about the SAS execution, including the resources that were used, notes, and warnings, as illustrated in Figure 3.2.

If an error occurs during the SAS execution, the  icon usually appears on the status bar. Clicking this icon opens the log, which is searchable. For example, search for “ERROR:” to find error messages.

FEATURE 4: USE ADJUSTMENT VARIABLES AND START-UP CODE TO FIX DATA ISSUES

Sometimes it is necessary to forecast a time series that is not expressed as desired in the input data. Adjusting the time series for known systematic variations or deterministic components enables you to more readily identify and model the underlying stochastic (unknown) time series process. Examples of systematic adjustments are currency-unit conversions, exchange rates, trading days, and other known systematic variations. Examples of deterministic adjustments are advanced bookings and reservations, contractual agreements, and other known contributions or deterministic components.

By providing suitable adjustment variables in the same data set, you can transform the time series values without changing the original series. In the **New Project** dialog box, assign your dependent variable and then click **Adjustments**, as illustrated in Figure 4.1. Alternatively, for an existing project, select **Project ► Hierarchy and Variable Settings** from the main menu.

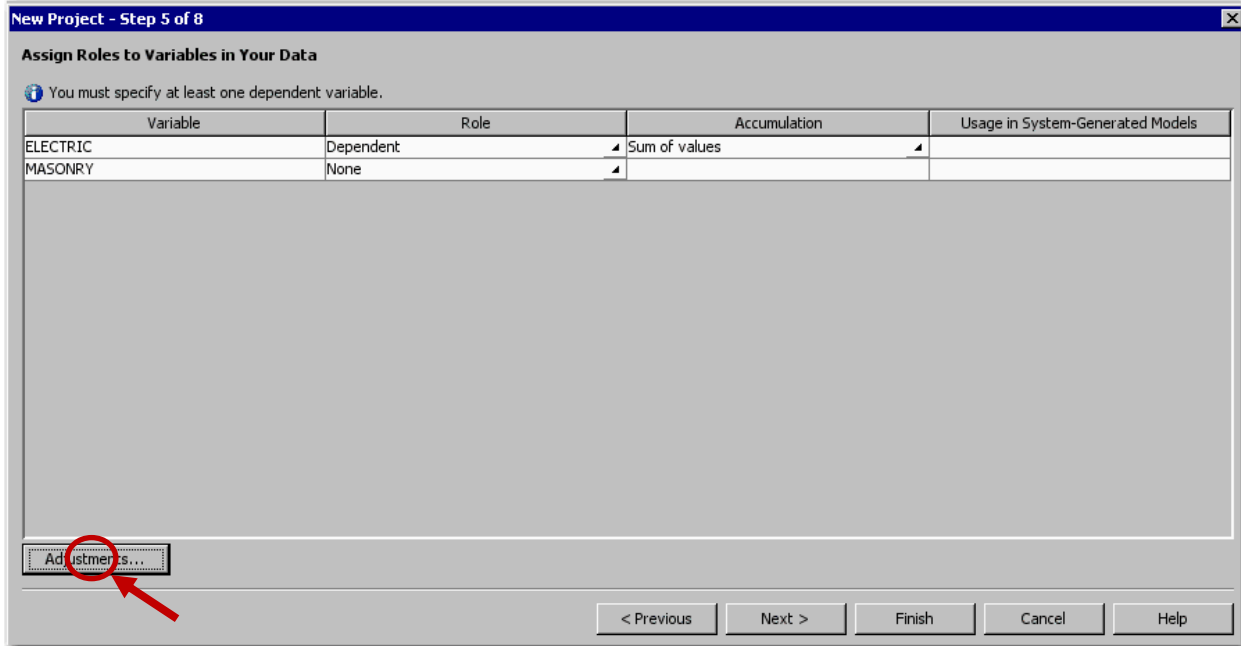


Figure 4.1: Assign Roles to Variables in Your Data: Adjustments Button

After you click **Adjustments**, the **Adjustments** dialog box appears, as shown in Figure 4.2; it lists the existing adjustments that are being applied to the project. Click **New** to open the **Adjustment Properties** dialog box, which is shown in Figure 4.3. In the **Adjustment Properties** dialog box, you can specify how to accumulate the adjustment variable values, the pre-operation (add, subtract, multiply, or divide), and the post-operation (add, subtract, multiply, or divide). Typically, the post-operation is the inverse of the pre-operation. So, if the pre-operation is add, the post-operation is subtract.

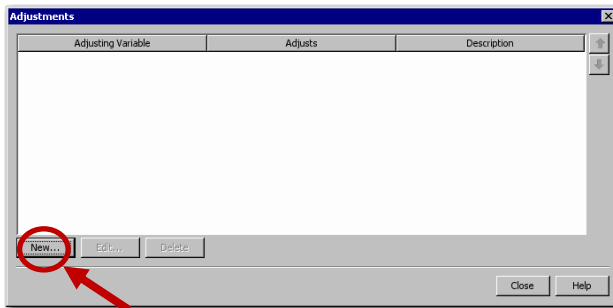


Figure 4.2: Adjustments Dialog Box

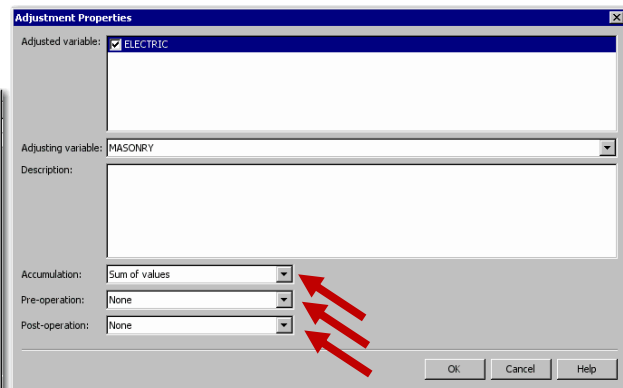


Figure 4.3: Adjustment Properties Dialog Box

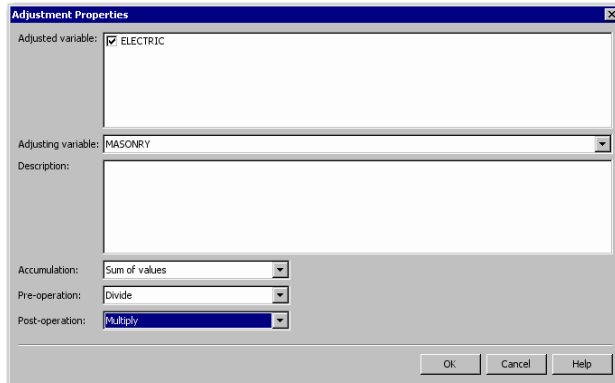


Figure 4.4: Adjustment Properties Dialog Box

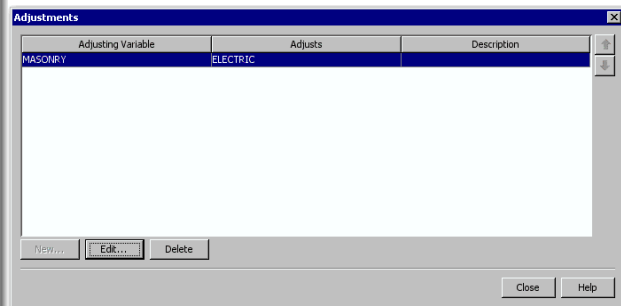


Figure 4.5: Adjustments Dialog Box

You can apply multiple adjustments. For example, you can select a variable whose values contain adjustments for estimated loss of data that occurred on certain dates by selecting **Add** from the **Pre-operation** list. The actual series will be modified according to the first adjustment added, and the results will be modified according to the second adjustment.

You can use pre-adjustment variables to adjust the dependent series prior to model parameter estimation, evaluation, and forecasting. After the predictions of the adjusted dependent series are obtained from the forecasting mode, you can use the post-adjustment variables to adjust these forecasts to obtain predictions that more closely match the original dependent series.

A typical example is adjusting a monetary-valued series for exchange rates. Suppose that the dependent series is product sales in Europe, expressed in Euros, and you want to convert currency to US dollars. Select the exchange rate as the adjustment variable and select **Multiply** from the **Pre-operation** list. The forecast graph will show the original data values, but the forecast will use the values that have been adjusted for the exchange rate. If you want to convert these values back to the original metric for display, select **Divide** from the **Post-operation** list.

As another example, suppose that your dependent series has missing data in some cases and you want to use values of a proxy series in these cases. Select the proxy series as the adjustment variable, and select **Max** from the **Pre-operation** list. Any nonmissing value of the proxy series will be greater than a missing value. The forecast graph will show the original actual values, with missing values, but the forecast will use the adjusted series with no missing values.

FEATURE 5: DEFINE RECURRING EVENTS THAT INFLUENCE THE TIME SERIES

An event repository stores information about calendar events with a brief description of each event. Calendar events can be represented by indicator variables that could be stored in the time series data. However, because the influential calendar events can vary from series to series, there might be too many to store efficiently and many calendar events will be redundant, making updates difficult. Therefore, to allow the reuse and update of the calendar events, it is better to store a brief description of the calendar event, reproduce the indicator variable in the computer's memory when needed, and store the calendar events independently of the time series data.

To create a new event:

1. You can create new events in the **New Project – Step 7 of 8** dialog box (as shown in Figure 5.1), or you can create new events for an existing project by selecting **Project ► Event Repository** from the main menu. The **Event Repository** dialog box appears, as shown in Figure 5.2.

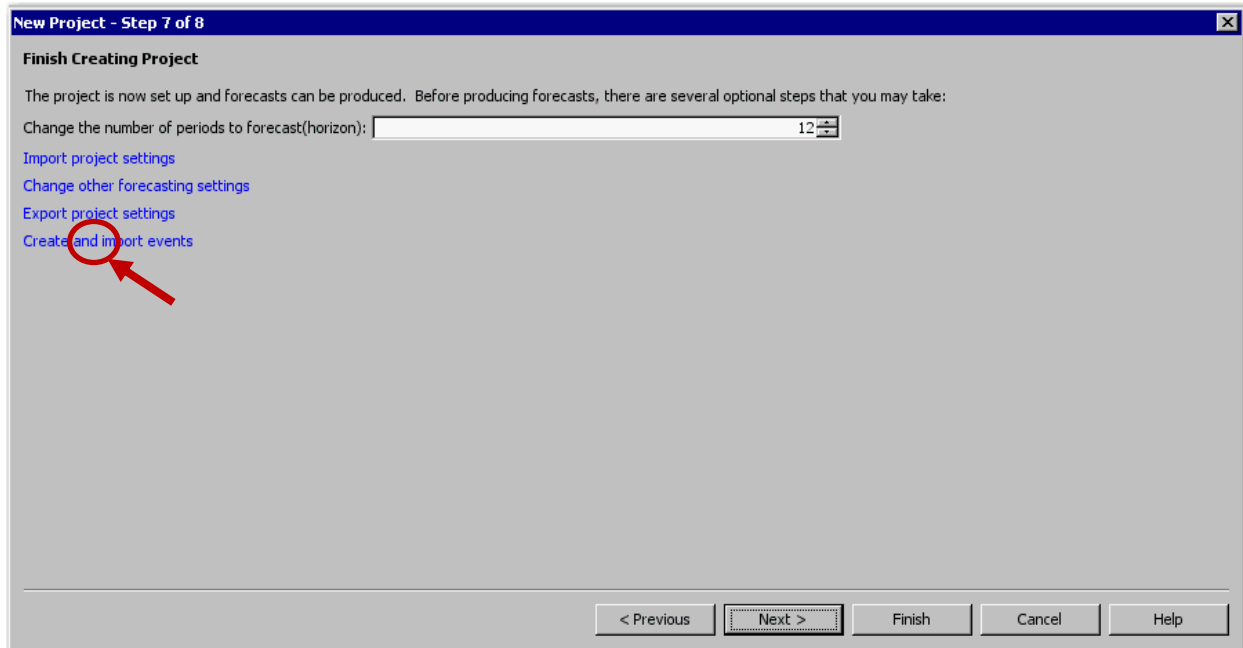


Figure 5.1: New Project Window

2. Click **New**. The **Event Properties** dialog box appears, as shown in Figure 5.3.
3. In the **Name** box, specify a valid SAS name for the event. You can also specify a description for the event in the **Description** box.
4. Click **Change** to specify an event type, and select from **Pulse**, **Level Shift**, **Ramp**, or **Temporary Change**.
5. Click **Add** to add an occurrence. The **Select Occurrences** dialog box appears, as shown in Figure 5.4. Select the time period for the occurrence and click **OK**. The **Occurrence** box in the **Event Properties** dialog box now displays a list of the occurrences for the event, as illustrated in Figure 5.4.
6. To specify the options for the occurrence, click **Edit** next to the **Options** box. The **Event Options** dialog box appears (not shown). In this dialog box, you can specify the duration of the occurrences and the time of the occurrences.
7. To specify recurrence, click **Edit** next to the **Recurrence** box. The **Event Recurrence** dialog box appears (not shown). In this dialog box, you can specify how frequently the event recurs and how long the recurrence will last.
Note: Before you can specify a recurrence, you must specify an occurrence. However, you cannot specify a recurrence if you have selected two or more occurrences or if you have selected a holiday for the event occurrence.
8. Click **OK** to save the event. The new event appears in the **Event Repository** dialog box.

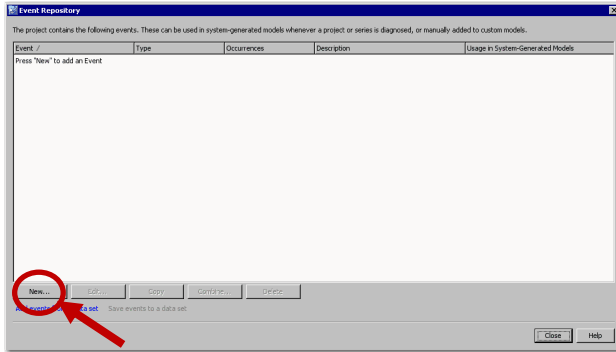


Figure 5.2: Event Repository Dialog Box

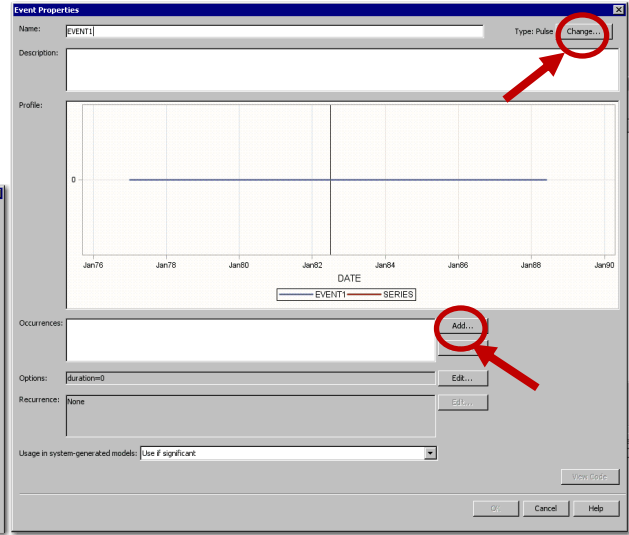


Figure 5.3: Event Properties Dialog Box

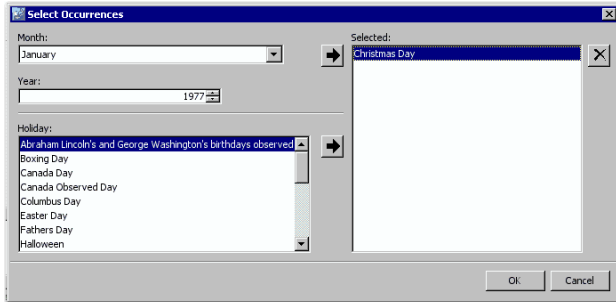


Figure 5.4: Select Occurrences Dialog Box

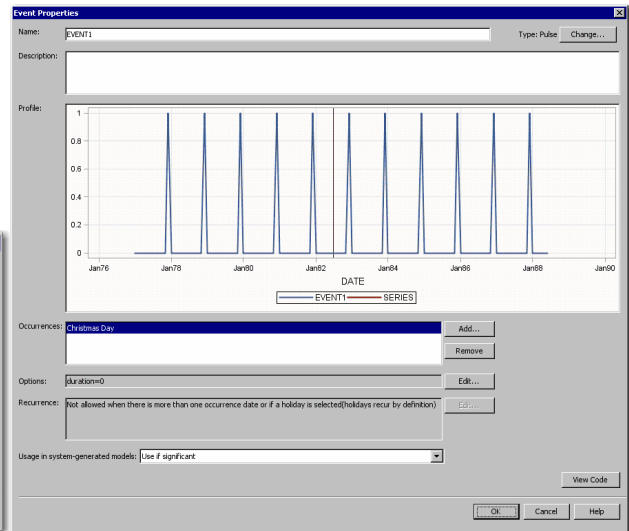


Figure 5.5: Event Properties Dialog Box

FEATURE 6: ADD MODELS BY IMPORTING FROM AN EXTERNAL LIST

SAS Forecast Studio supports the following types of models:

- *Default models* are included in SAS Forecast Studio. You cannot edit or delete these models. However, you can copy a default model and customize it to meet your needs.
- *System-generated models* are automatically generated by the diagnostics in SAS Forecast Studio. You cannot edit these models, but you can delete them and you can copy a system-generated model and customize it to meet your needs.
- *Custom models* are created in SAS Forecast Studio. When you create a model for a particular time series, that model is automatically added to the project model repository. After the model is added, you can apply it to other series in the project. Because these custom models are user-defined, you can add, edit, copy, or delete them.

- *Models from an external list* originate from the external list that is specified in the **Model Generation** panel of the **Forecasting Setting** dialog box. In the model repository, models from an external list are listed as having type **Custom (Read-Only)**. You cannot edit or delete these models. However, you can copy a model from an external list and customize it to meet your needs.

To select models to fit to each series from an external model selection list:

1. In the **New Project - Step 7 of 8** dialog box, click **Change other forecast settings**, as shown in Figure 6.1. For a previously created project, you can select **Project ► Forecasting Settings** from the main menu. The **Forecasting Settings** dialog box appears, as shown in Figure 6.2.
2. Select **Model Generation**, as highlighted in Figure 6.2.
3. Select the **Models from an external list** check box, and click **Browse**. The **Select Model Selection List** dialog box appears (not shown here), and you can select the model selection list that you want to use.
4. Click **OK**.

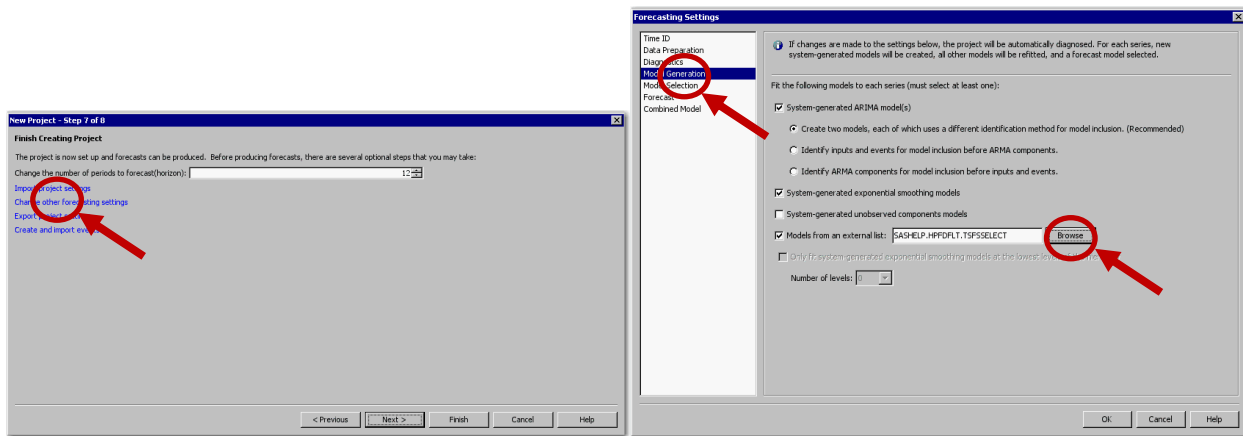


Figure 6.1: New Project – Step 7 of 8

Figure 6.2: Forecasting Settings

FEATURE 7: USE ROLLING SIMULATIONS TO EVALUATE FORECAST ACCURACY OVER THE NUMBER OF PERIODS YOU NEED TO FORECAST

For a particular time series data set, you can use automatic time series modeling software to select an appropriate time series model. You can use various statistics to judge how well each candidate model fits the data (in-sample). Likewise, you can use various statistics to select an appropriate model from a list of candidate models (in-sample or out-of-sample or both). Finally, you can use rolling simulations to evaluate ex-ante forecast performance over several forecast origins. Leonard et al. (2014) demonstrate how you can use SAS® Forecast Server Procedures and SAS Forecast Studio software to perform the statistical analyses that are related to rolling simulations.

On the **Model View** tab, click the rolling simulation button (circled in red in Figure 7.1) or select **Series ► Rolling Simulation** from the main menu.

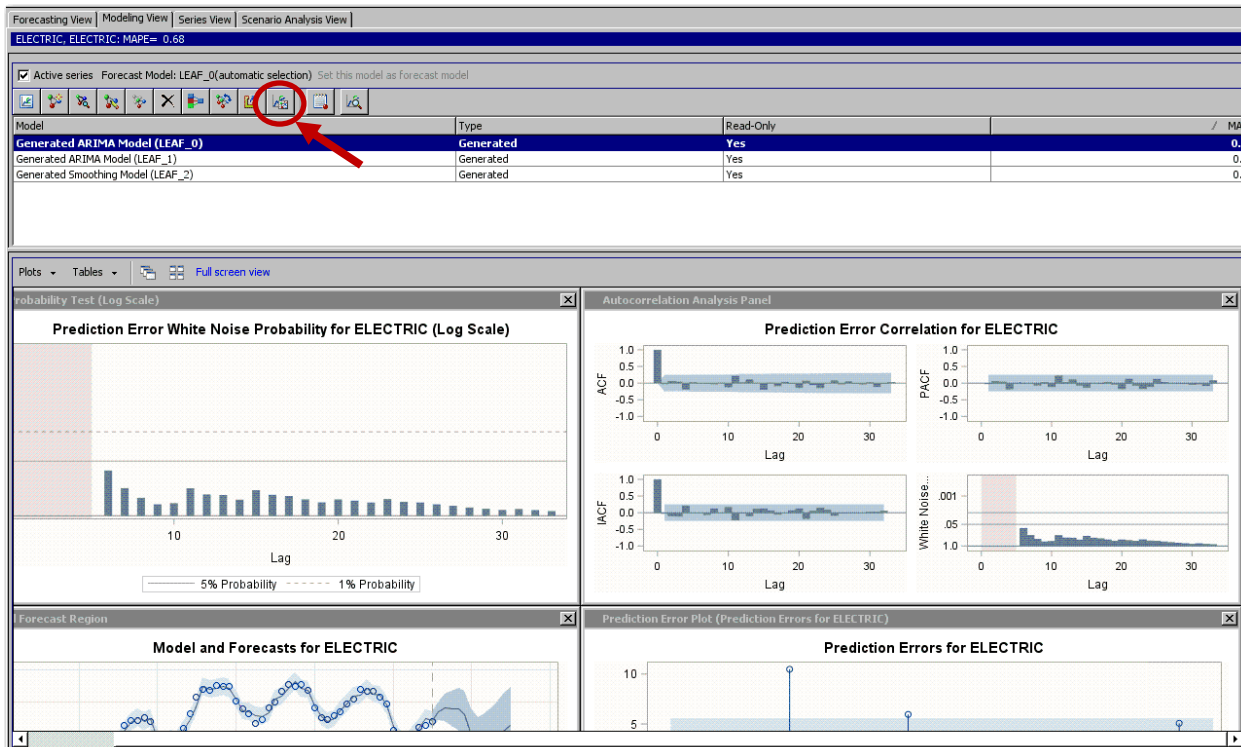


Figure 7.1: Modeling View: Rolling Simulations Button

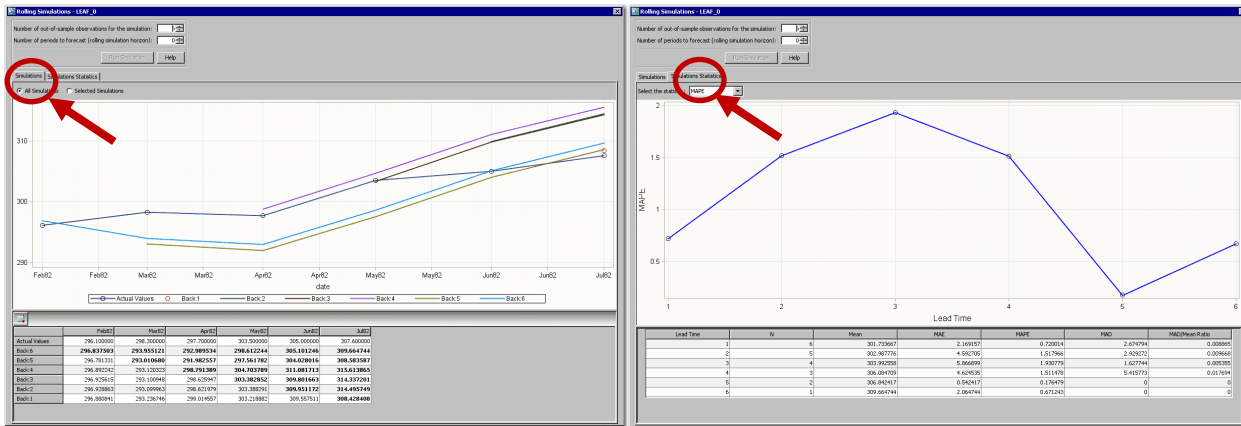


Figure 7.2: Rolling Simulation: Prediction Plots Figure 7.3: Rolling Simulation: Statistics of Fit

Rolling simulations (also called rolling horizon) mimic how the forecast model performs over time. Figure 7.2 illustrates a rolling horizon of six periods. The actual time series values are the black circles, which are connected by the black lines. The colored lines represent the six predictions, each beginning at a different origin. Figure 7.3 illustrates a rolling statistic of fit for six periods.

FEATURE 8: EVALUATE THE EFFECTS OF INDEPENDENT VARIABLES

Organizations often want to make decisions on the basis of time series forecasts that depend on future, controllable causal factors. Examples of future causal factors are pricing and advertising expenditures for products and services. To help your organization make better decisions, you can vary the future values of the controllable causal factors to help determine the best decision (a what-if analysis). For example, if you are forecasting the profit of your company and material cost is an underlying factor, then you could use scenario analysis to determine how the forecasted profit would change if the material cost increased by 10%. Leonard (2000) thoroughly demonstrates how you can use SAS/ETS® software for this purpose and

describes the detailed statistical analyses that are related to scenario analysis. To perform scenario analyses, click the **Scenario Analysis View** tab on the **Modeling View** tab, as shown in Figure 8.1.

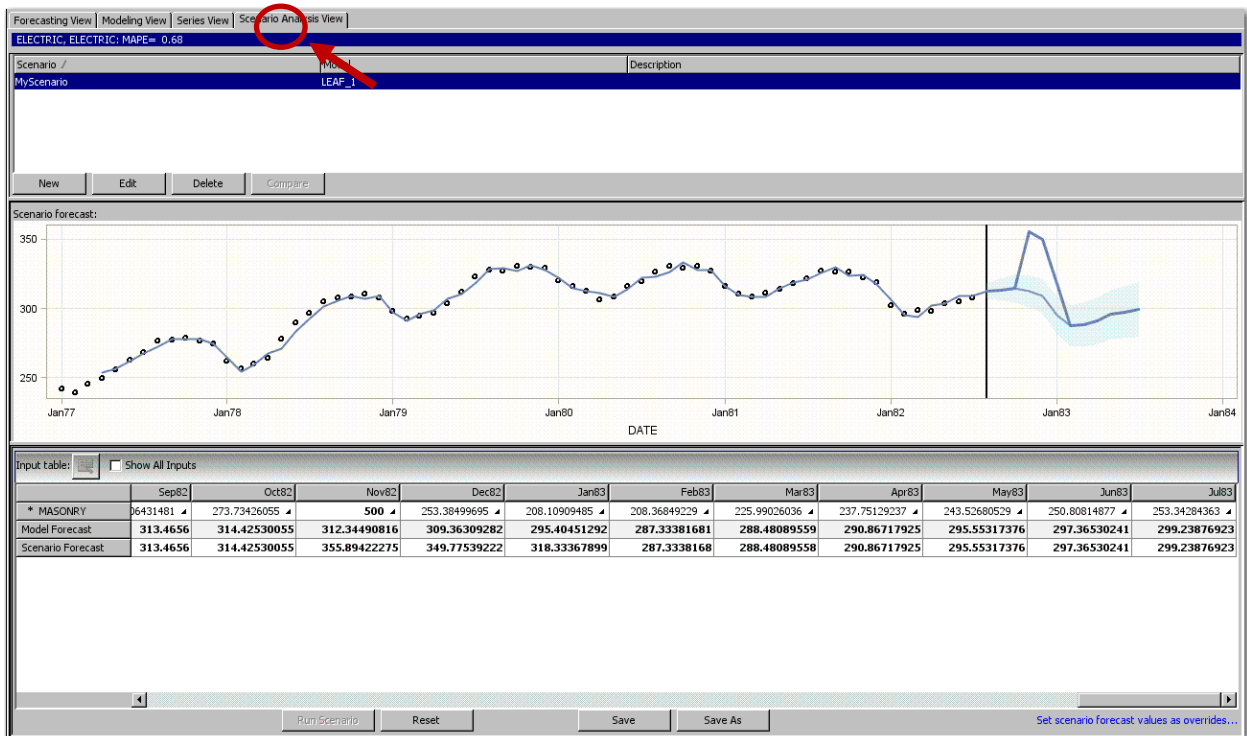


Figure 8.1: Scenario Analysis View

FEATURE 9: GAIN INSIGHT INTO RESULTS BY COMPARING MODELS

If you have created several models for a time series, then you might want to compare the statistics of fit for the models that were fitted to that series. This comparison can help you determine which model you want to use to generate forecasts.

From the **Modeling View**, click the compare models button (circled in red in Figure 9.1) or select **Series ► Compare Models** from the main menu. The **Compare Models** dialog box appears.

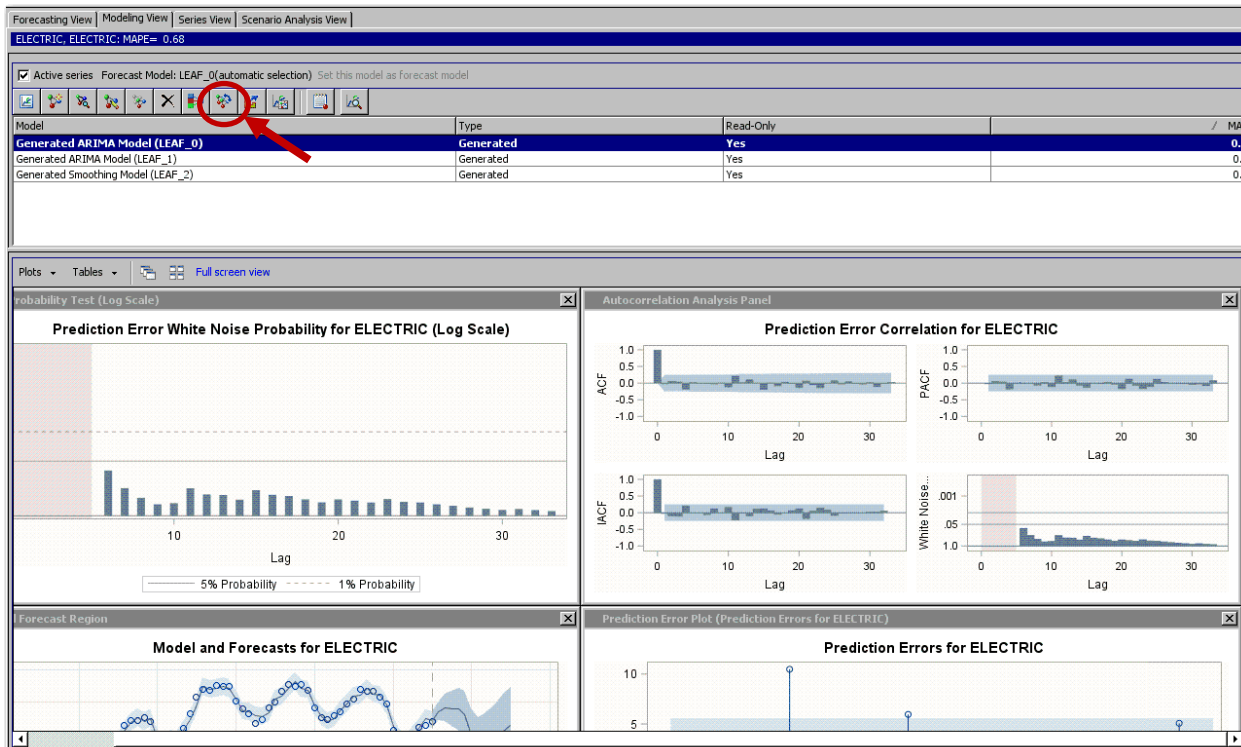


Figure 9.1: Modeling View: Compare Models Button

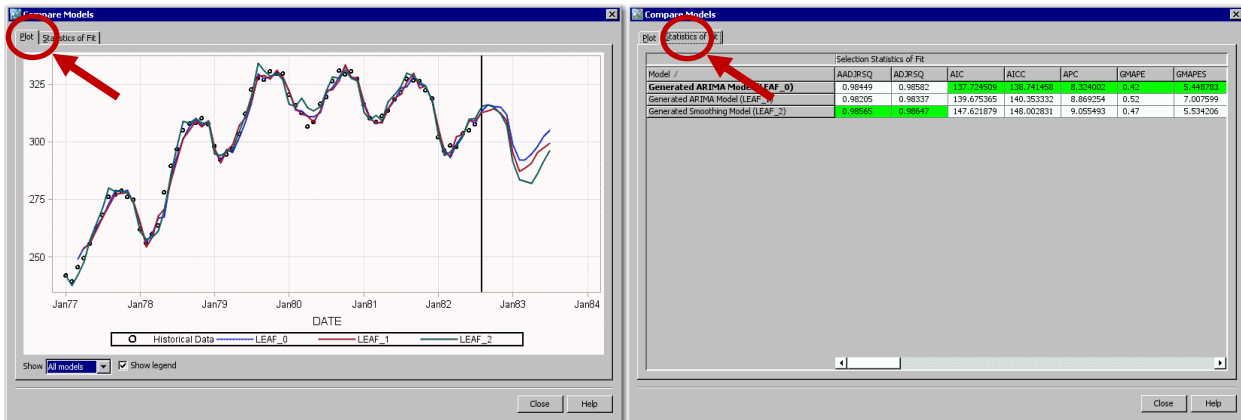


Figure 9.2: Compare Models: Prediction Plots Figure 9.3: Compare Models: Statistics of Fit

Figure 9.2 illustrates the model comparison plot for three models that are associated with a single time series. The actual time series values are represented by the black circles. The colored lines represent the model predictions from three different time series models. For this time series, the models are in close agreement. This is a good situation.

To view the model comparison statistics for the three time series models, click the **Statistics of Fit** tab, as shown in Figure 9.3. The champion model for each statistic is highlighted in green. For this time series, the models are in close agreement. This is a good situation.

FEATURE 10: RUN THE CODE IN BATCH

Forecasts can be generated either interactively or in batch. When you interact with SAS Forecast Studio, it generates SAS code for execution. This code is stored in the system directory of your Forecast Studio

projects. These SAS code files implement the forecasting process steps that are associated with SAS Forecast Server. Table 1 lists the files and describes their contents.

File Names	Contains SAS Code to:
CREATE_PROJECT_DO_NOT_IMPORT_DATA and CREATE_PROJECT_IMPORT_DATA	Re-create the project. This code includes all the data preparation and forecasting steps.
DIAGNOSE_DESTRUCTIVE_DO_NOT_IMPORT_DATA and DIAGNOSE_DESTRUCTIVE_IMPORT_DATA	Re-diagnose the project. This code does not include the data preparation steps.
SELECT_MODELS_DO_NOT_IMPORT_DATA and SELECT_MODELS_IMPORT_DATA	Perform model selection, estimate the parameters of the selected model, and produce forecasts. This code does not include the data preparation, diagnose, fit, and forecasting steps.
FIT_MODELS_DO_NOT_IMPORT_DATA and FIT_MODELS_IMPORT_DATA	Estimate parameters by using the model that you specified and then create a forecast. No model selection is performed. This code does not include the data preparation and diagnose steps.
FORECAST_MODELS_DO_NOT_IMPORT_DATA and FORECAST_MODELS_IMPORT_DATA	Re-forecast the model and parameters. When refitting the model parameters, SAS Forecast Studio uses the estimate of the previous parameter as a starting point for re-estimation. This code does not include the data preparation, diagnose, and fit model steps.
RECONCILE_FORECASTS_AND_OVERRIDES_DO_NOT_IMPORT_DATA	Reconcile model forecasts and overrides. This code does not include the data preparation, diagnose, select, fit, and forecast steps. This code can be run only if those steps have been executed at least once.
RECONCILE_FORECASTS_DO_NOT_IMPORT_DATA	Reconcile model forecasts. This code does not include the data preparation, diagnose, select, fit, forecast, and override reconciliation steps. This code can be run only if those steps have been executed at least once.
RECONCILE_OVERRIDES_DO_NOT_IMPORT_DATA	Reconcile overrides. This code does not include the data preparation, diagnose, select, fit, forecast, and reconciliation of the forecast steps. This code can be run only if those steps have been executed at least once.

Table 1: Batch Execution Files

CONCLUSION

SAS Forecast Server is one of the most feature-rich forecasting products on the market. This paper described 10 underused features that can improve your workflow.

REFERENCES

Leonard, Michael. 2000. "Promotional Analysis and Forecasting for Demand Planning: A Practical Time Series Approach." Presented at the 2000 International Symposium of Forecasting. Lisbon, Portugal.

Available <https://support.sas.com/rnd/app/ets/papers/PromotionalAnalysis.pdf>

Leonard, Michael. 2002. "Large-Scale Automatic Forecasting: Millions of Forecasts." Presented at the 2002 International Symposium of Forecasting. Dublin, Ireland.

Leonard, Michael. 2004. "Large-Scale Automatic Forecasting with Calendar Events and Inputs." Presented at the 2004 International Symposium of Forecasting. Sydney, Australia.

Leonard, Michael, Ashwini Dixit, and Udo Sglavo. 2014. "Ex-Ante Forecast Model Performance with Rolling Simulations." *Proceedings of the SAS Global Forum 2014 Conference*. Cary, NC: SAS Institute Inc.

Available <https://support.sas.com/resources/papers/proceedings14/SAS213-2014.pdf>

ACKNOWLEDGMENTS

The authors would like to thank Anne Baxter from SAS for her editing.

CONTACT INFORMATION

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

Michael Leonard
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
919-531-6967
Michael.Leonard@sas.com

Evan Anderson
evanlewisanderson@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.