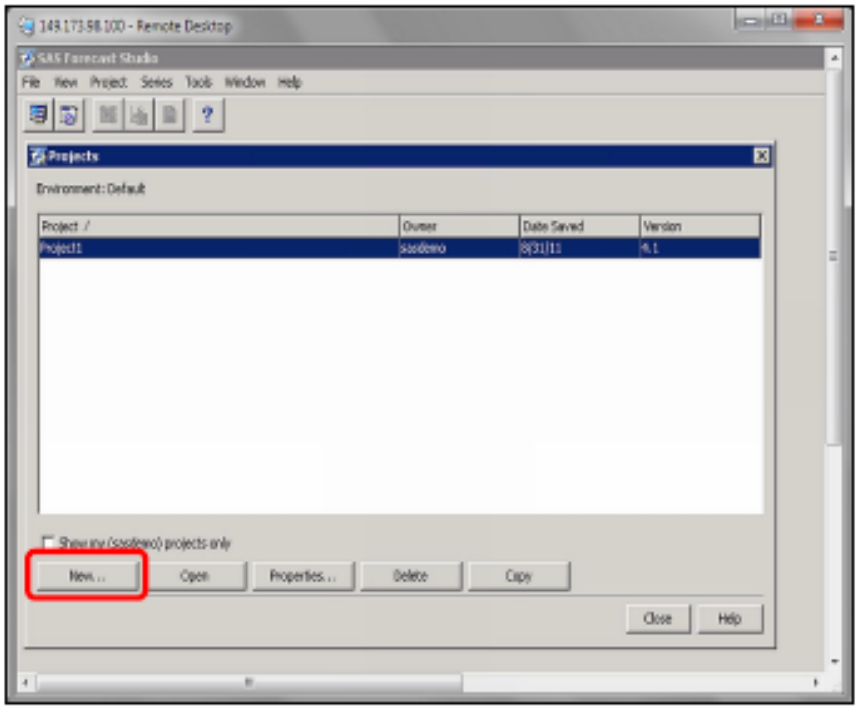
Exploring SAS Forecast Studio

Generating Forecasts Automatically

1. Enter your user name and password, and then click **Log On** to start the process of project creation.

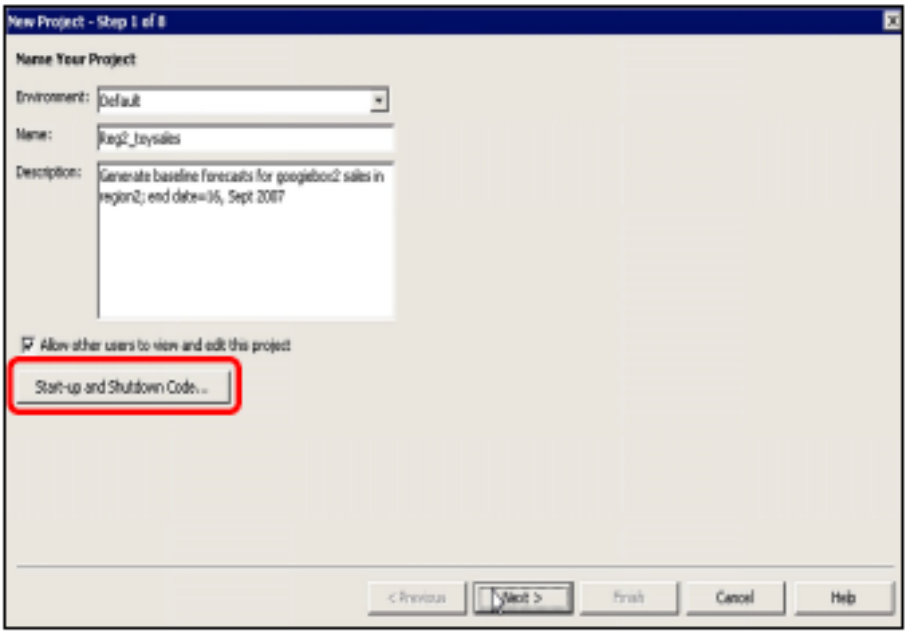
In this course, your user name, password, and server name are populated for you.

2. **Click New to create a new project.**



Your first session has no existing projects unless projects were set up for you.

3. Name your project.



It is easy to create many projects in SAS Forecast Studio. Naming your project something relevant will help you decipher what is in it a few months after it is created.

A reasonably detailed description of the project is considered a best practice.

You can now add custom SAS code that runs under the following circumstances:

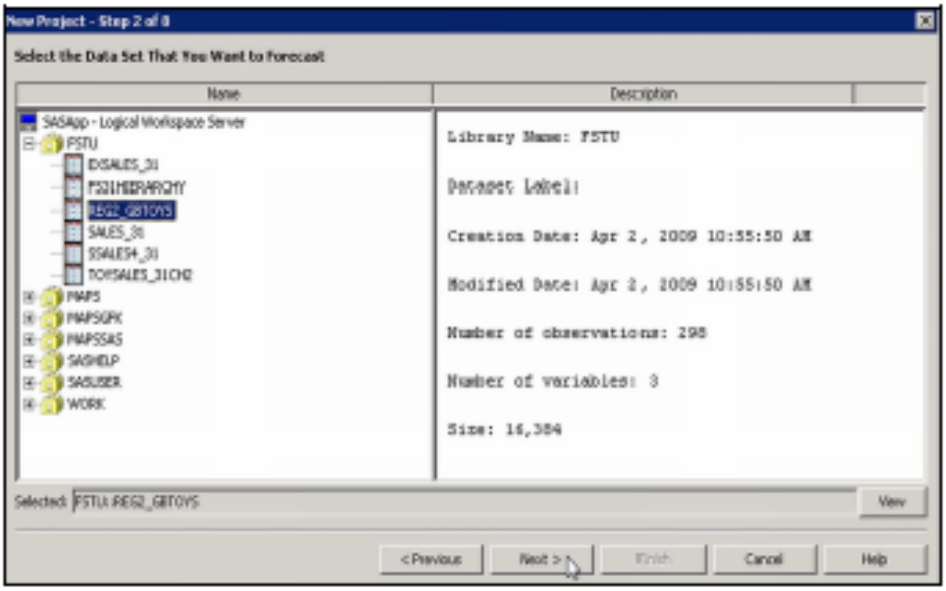
- When an environment is opened or closed. For example, libraries can be assigned by including LIBNAME statements in the start-up code for an environment. These libraries are available only for the associated environment.
- When a project is opened or closed. You can add this code when you create the project in the New Project wizard. After SAS Forecast Studio creates the project, you can add this code by changing the project's properties.

4. Click **Next**.

5. Select the **FSTU** library and the **REG2_GBTOYS** data set. Some summary information about the library and data set chosen is given in the panel on the right.

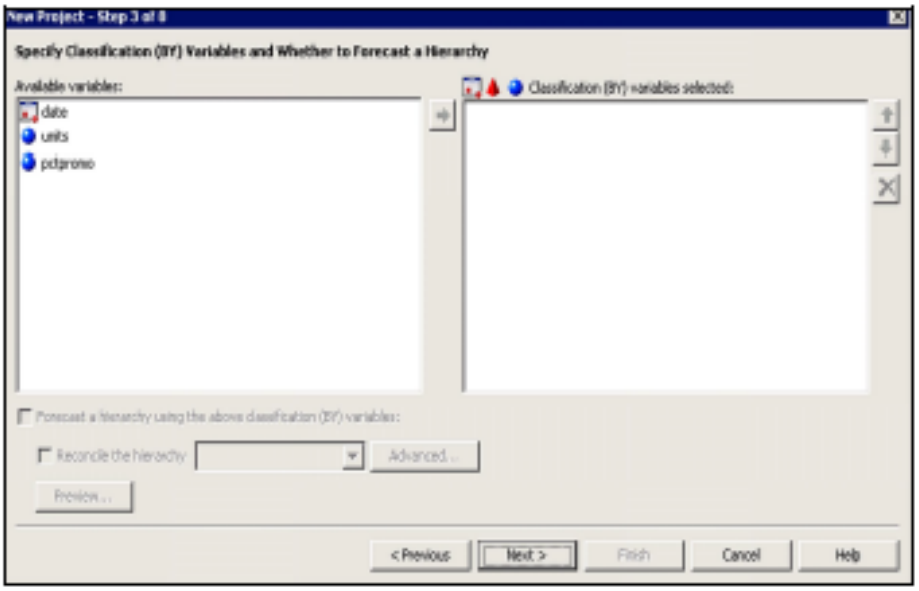
   The library shown above might be different than the library you will use in class.

   The libraries and data sets shown have been registered in the SAS Management Console

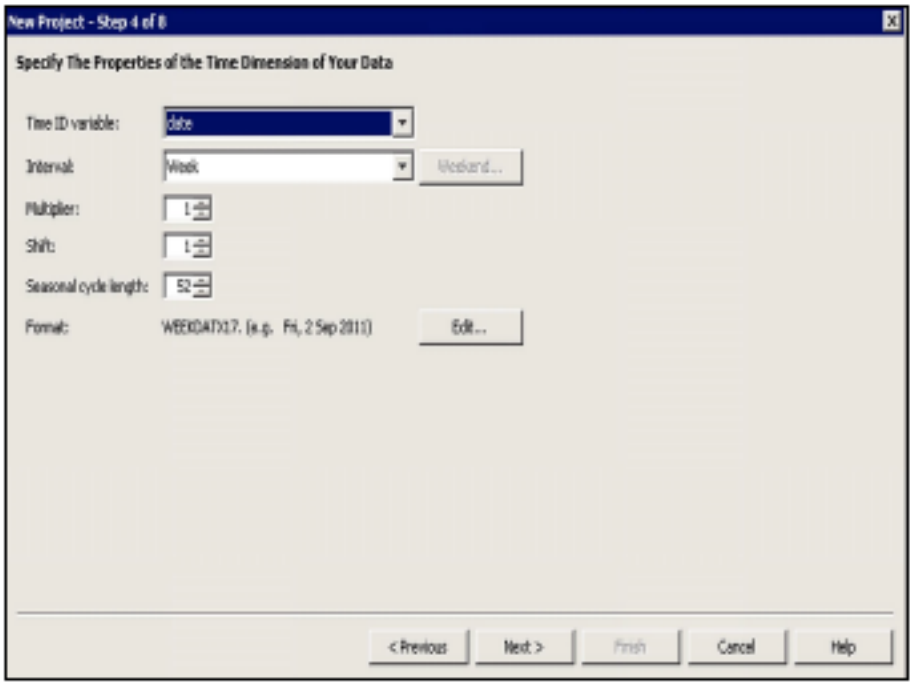


6. Click **Next**.

7. Step 3 assigns BY variables that form the basis of a data hierarchy. Because a single series is to be forecast, there is no associated hierarchy to the data. Click **Next**.

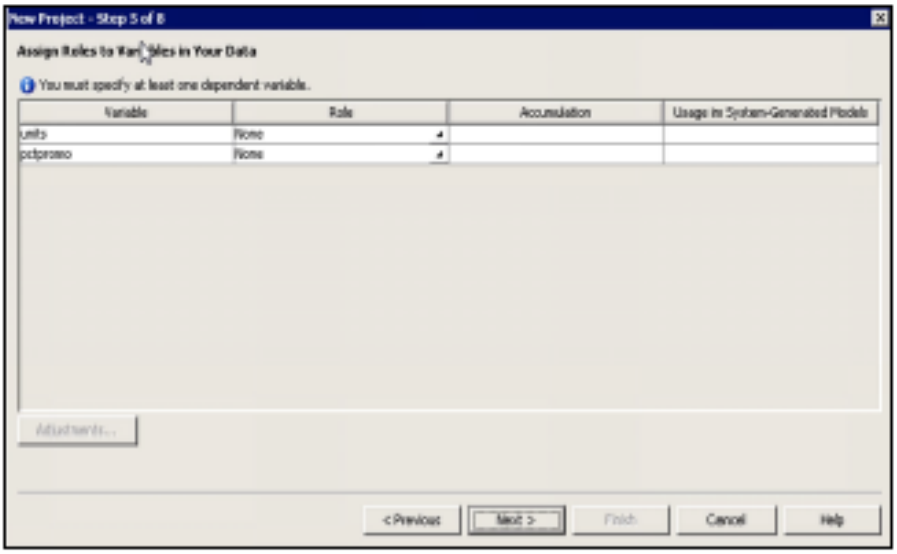


8. Click the down arrow next to **Time ID variable** and select **date** as the time ID.

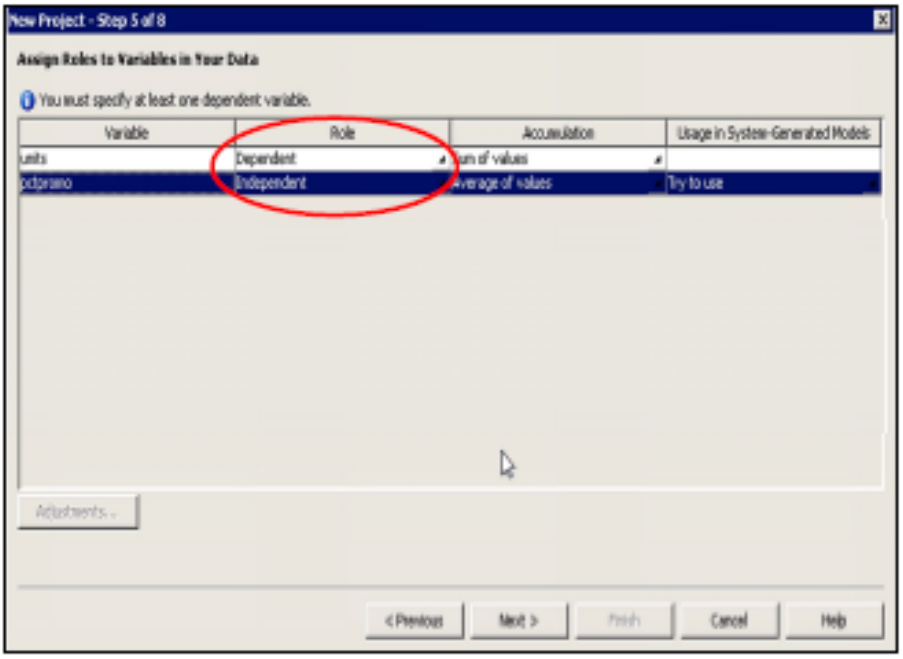


SAS Forecast Studio detects **Week** as the natural interval of the data. The detected interval of the data can be changed directly by choosing another option from the **Interval** list. Selected intervals can be modified using the **Multiplier** and **Shift** options. The Multiplier option can be used to modify the time interval from the default weekly level to biweekly level. The Shift option can be used to shift the default first day of the week from Sunday to Monday.

9. The variable **units** measures per-unit toy sales in region 2. This is the **dependent**, or forecast, variable for the project. **pctpromo** is a binary promotion flag that will be used as a candidate input variable in the project.

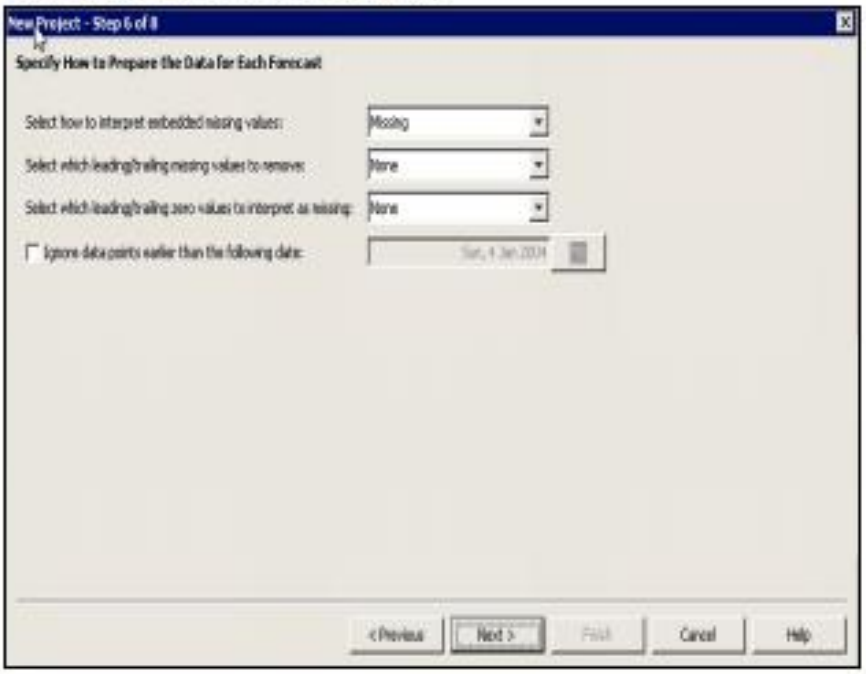


10. Assign roles as indicated. Accumulation and usage options are discussed in Chapter 4.



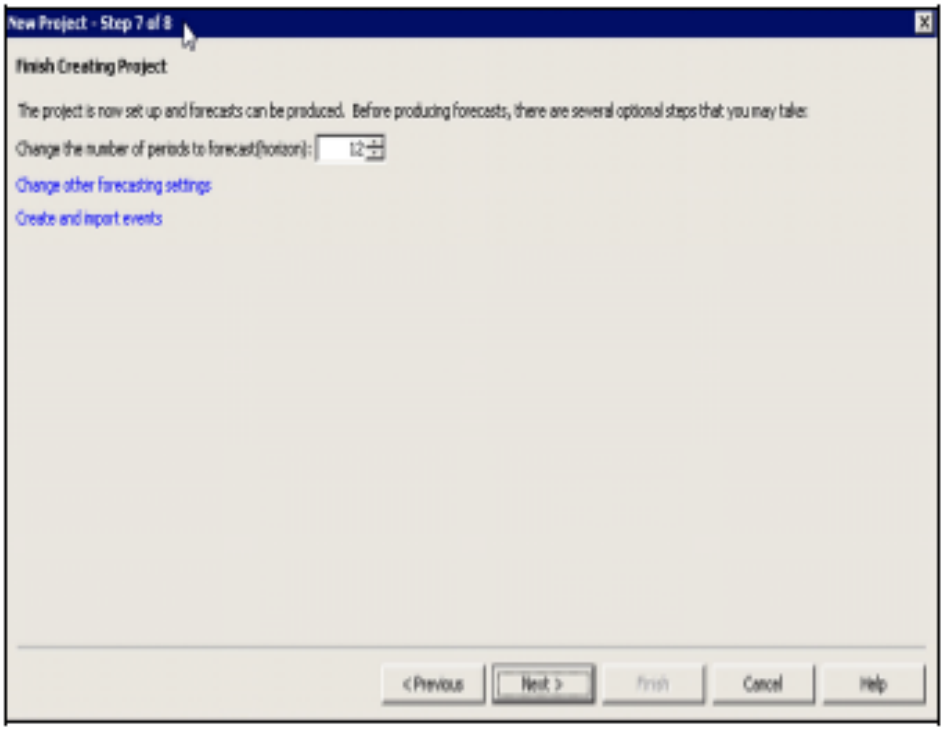
11. This step specifies data preparation options. Click **Next**.

No data preparation is necessary for this series.



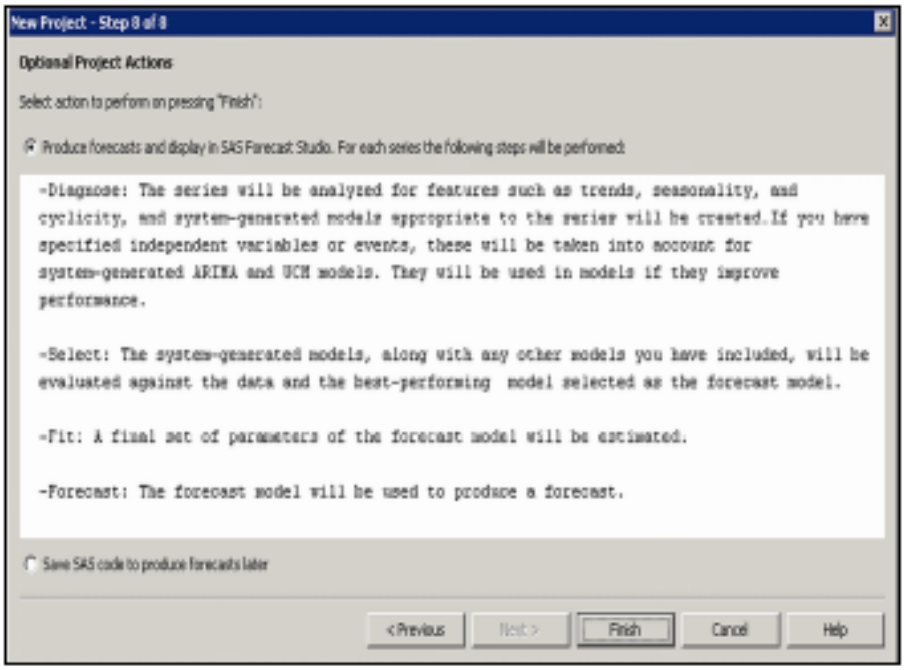
Data preparation options enable the user to interpret, or impute, values for embedded missing values in the series. Data sets that have been padded with zeros can also be modified; leading or trailing zeros can be changed to missing values. Leading or trailing missing values can be eliminated from the data, or trimmed.

12. Our project is now set up, and we are ready to generate forecasts. The lead forecast horizon is set at 12 weeks or, roughly, three months into the future. Click Next.



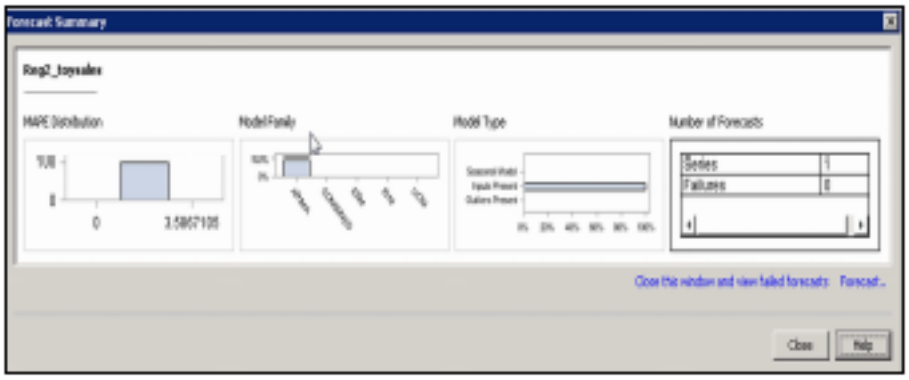
The fields Change other forecasting settings and Create and import events are discussed later in this course.

13. Step 8 lists optional actions. Click **Finish** to generate the forecast.



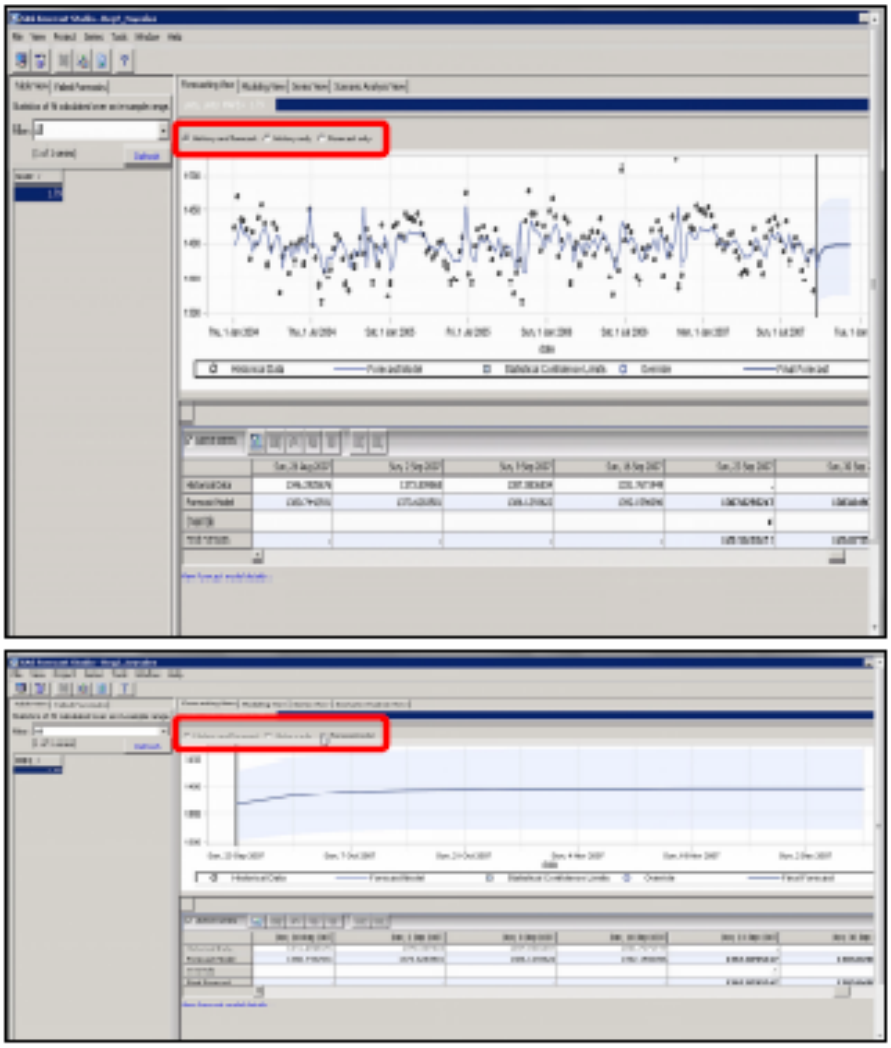
SAS Forecast Studio is a code generator. The default behavior, shown above, immediately submits the code, and results are generated and displayed in the user interface. Alternatively, the code can be saved and run later. Saved code can be modified.

Summary results are displayed first.



- The candidate input variable was chosen and built into the forecast model.
- The forecast model is a system-generated ARIMA model. This model is custom built from scratch to accommodate the systematic variation in our **REG2_GBTOYS** data.
- Generated MAPE is quite low; it is less than 3. Mean absolute percent error (MAPE) measures the difference between the forecast, or predicted, value and the actual observation for each time increment, so smaller is better. Here, it is assumed that this MAPE value is generated by an adequate forecast model; no further refinement is necessary. However, this assumption abstracts from two important ideas.
- First, any time series has a systematic component and a noise component. We want the model to capture all of the systematic component, or signal, and to ignore the noise. The generated MAPE gives us an indication that the model picked up a lot of the systematic variation, and that the proportion of noise to signal is quite low. However, a low MAPE does not guarantee that the model picked up all of the systematic variation. Further refinement might be feasible. In large-scale forecasting, choices about model refinement are usually constrained by the time available for forecasting and the value of the series being forecast.
- Second, because the entire series was used to both estimate parameters and to calculate MAPE, the fit statistic is somewhat suspect. Effectively, MAPE is telling us how well the model reproduced the data, the signal, and the noise, and not how well the model extrapolates onto data it has not seen before. Extrapolation performance is the primary concern in forecasting. Holdout samples are used in later chapters to assess model performance and to select the model that extrapolates best.

The created project opens in the Forecasting view by default.



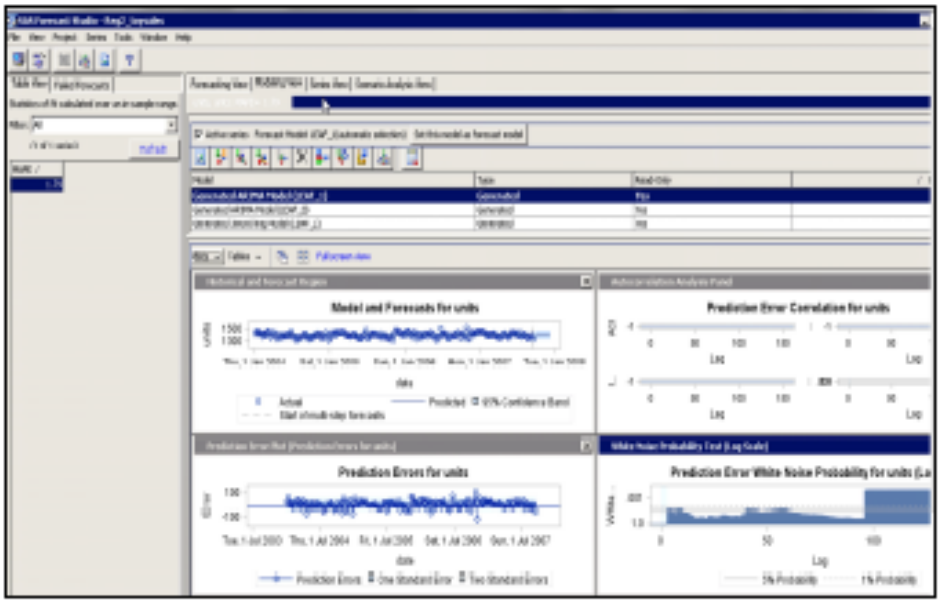
There are two main areas of information contained in the Forecasting view.

- The forecast plot shows the forecast generated by the forecast model. The lead forecast shown in the second plot above is the forecast used for this series by default. The forecast model is the model with the best fit – here, lowest MAPE – among all models considered for this series.
- The table shows forecasted values and historical values for each time increment in the series and in the lead forecast horizon. This table can also be used to apply manual overrides to model-generated forecast values. Overrides are discussed in Chapter 3.

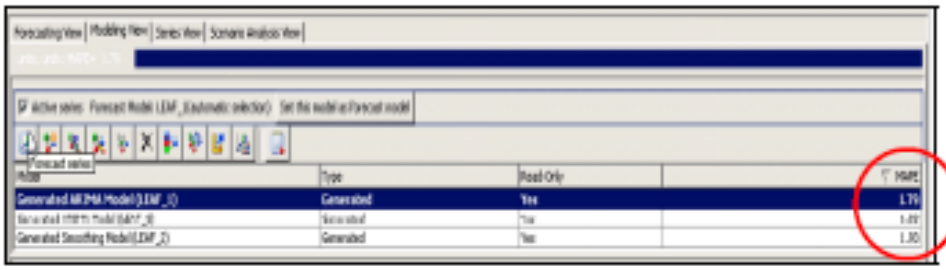
By default, 95% prediction intervals are provided for future forecasts.

14. Click the **Model View** tab.

The Model view contains information about all candidate models considered for the active series, the selected or forecast model, parameter estimates, and model diagnostic information.



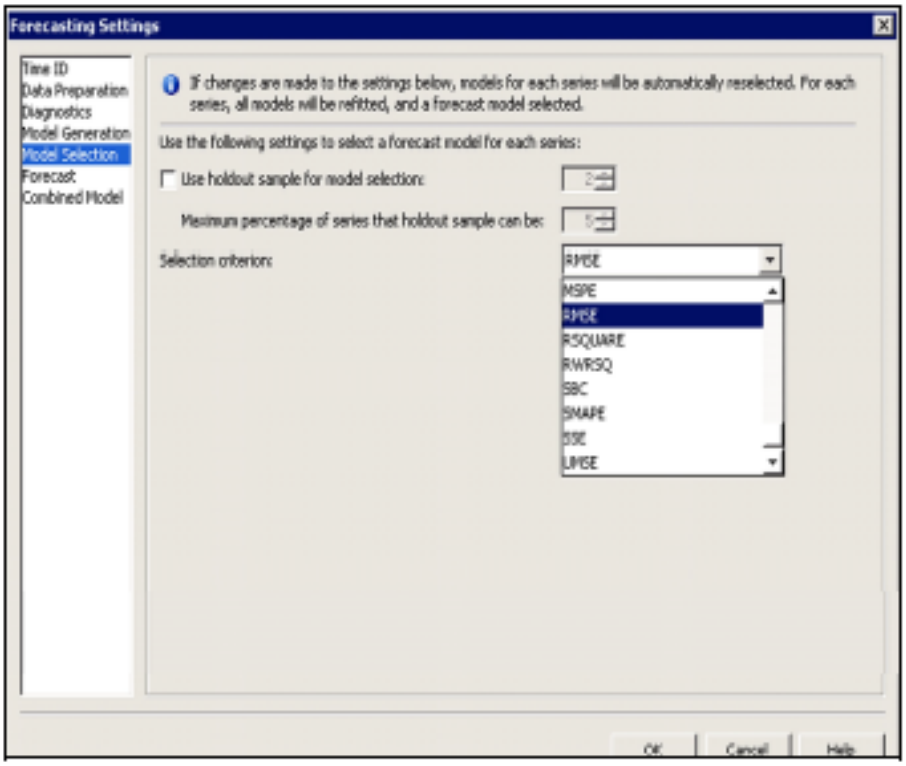
The forecast model for the series is the candidate model with the best fit. Here, this is the generated ARIMA model with the lowest associated MAPE (highlighted).



Setting the Project's Model Selection Statistic

1. **Select Project ➔ Forecasting Settings.**

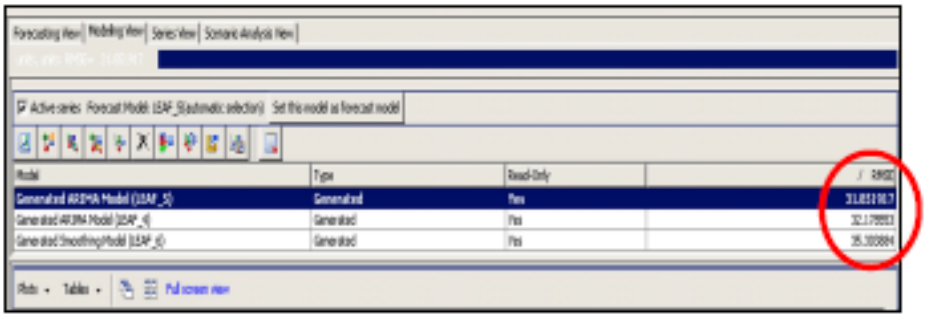
Alternative selection statistics are contained under the Model Selection settings.



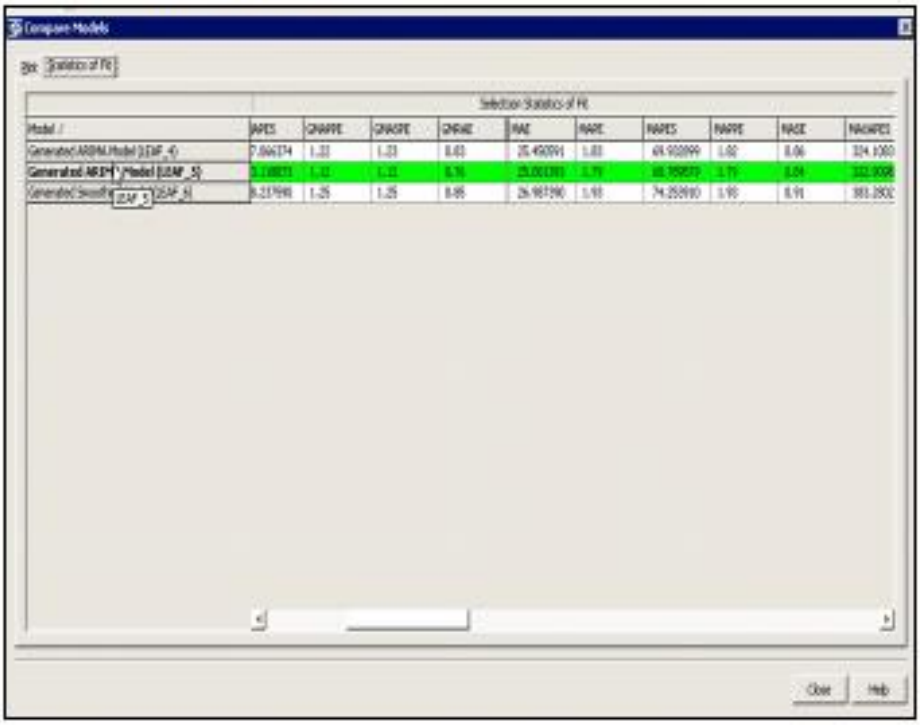
Holdout sample settings can be activated and changed under the Model Selection settings.

Most selected, forecast models are not uniformly best under all selection criteria. That is, changing the fit statistic mid-project can change the model chosen to generate forecasts.

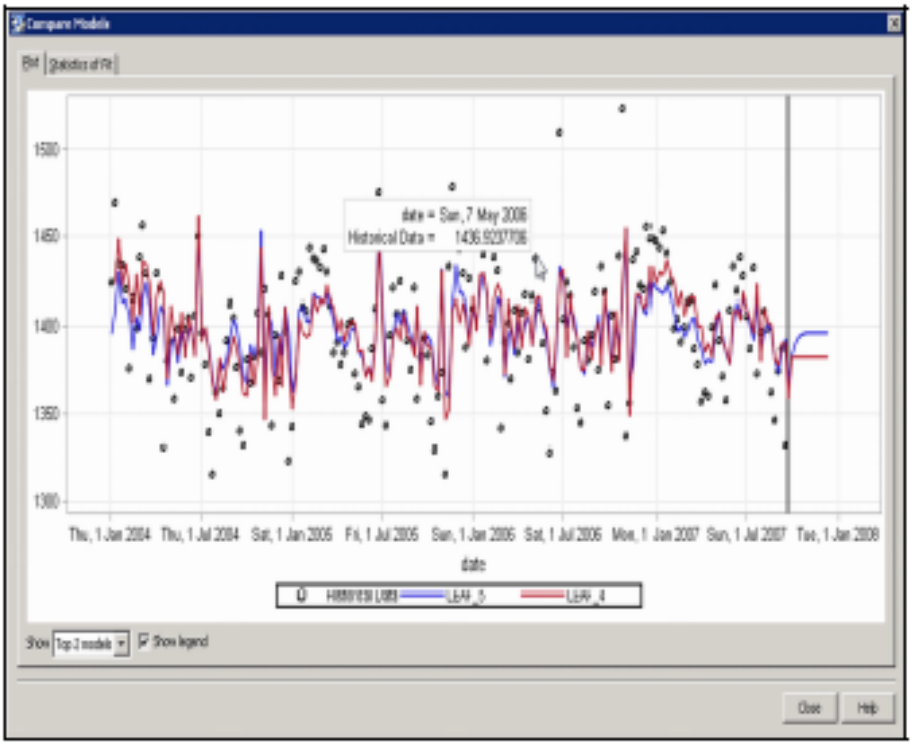
In this case, the change in selections statistic does not change the selected, forecast model.



2. To examine how each candidate model performs across all supported fit statistics, select **Series → Compare Models → Statistics of Fit**. The forecast model, ARIMA (LEAF 5), seems reasonably robust to a change in the selection statistic.



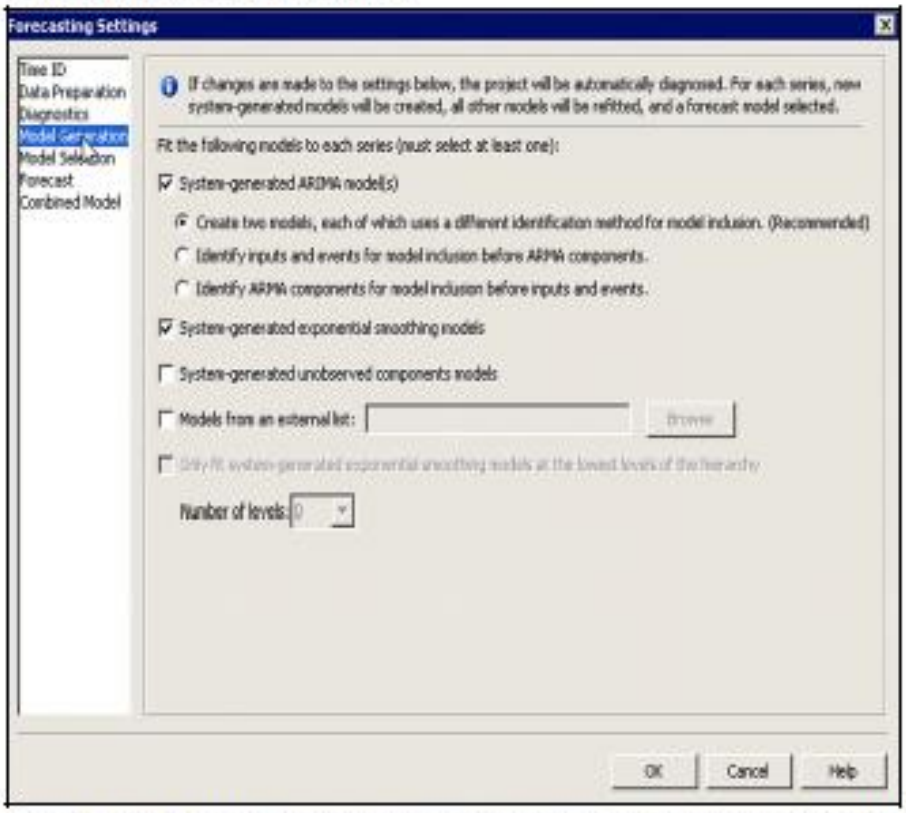
3. To directly compare the forecast plots from two or more models, select **Series ➡ Compare ➡ Plot** (or, click the **Plot** tab from above).



An overlay plot of the forecasts for the top two models is shown by default. Above, the project has been modified so that an overlay plot of the generated exponential smoothing model and the forecast model, ARIMA (LEAF 5), are shown.

The selection statistic is changed back to MAPE for the remainder of the slides in this chapter. Because the change in selection statistic did not change the selected forecast model, the analysis and outcomes are not materially affected by changing back to MAPE. That is, students can change back to MAPE if they want to, but it is not necessary.

You can view and modify the default settings for model generation mid-project by selecting **Project ➡ Forecasting Settings ➡ Model Generation**.



By default, SAS Forecast Studio generates two types of custom or system-generated models: ARIMA and exponential smoothing (ESM).

Potential default changes include the following:

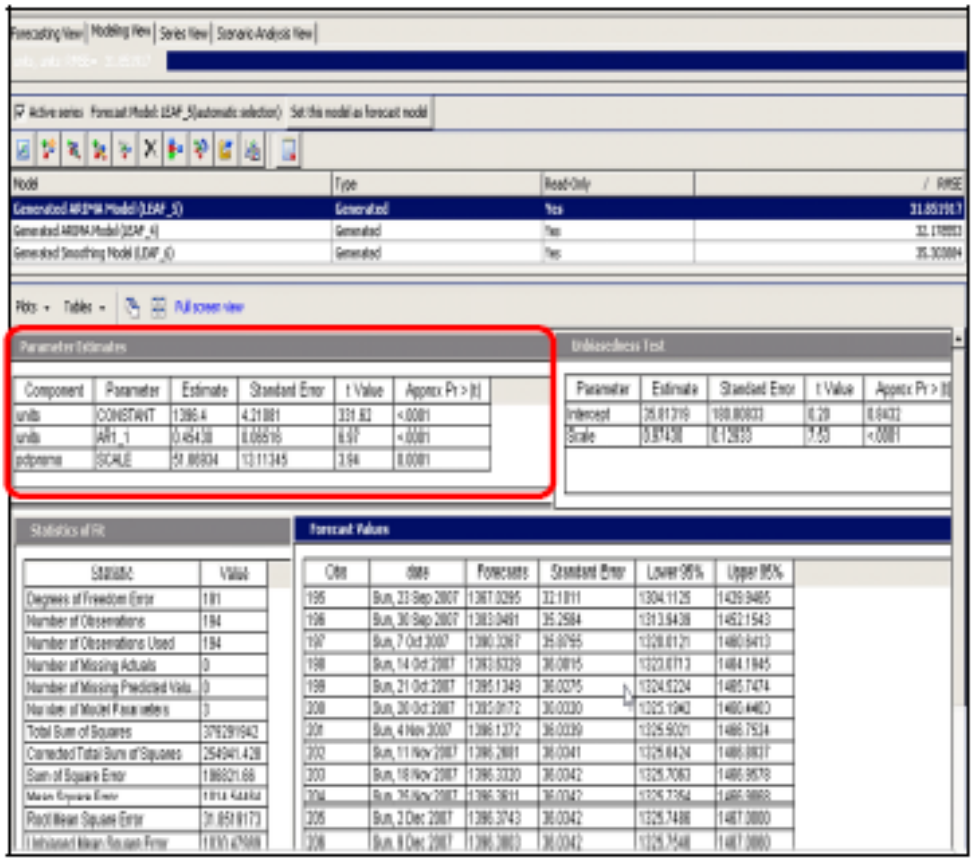
- Generating a custom, unobserved components model for each series.
- A model selection list containing user-defined models can be used in the project. Each model on the list is fit to each series in the project and is considered in the model selection process.

There are three options associated with generated ARIMA models that concern exogenous (input and event) variable handling. There are no similar options associated with ESM. *Why is this?*

The software creates generated models that are customized to accommodate the systematic variation in each series in the project. This relieves the user from having to manually specify an appropriate model for each series in the data. That is, a SAS Forecast Studio user does not need a background in time series model building to generate good time series models (ARIMA, ESM, and/or UCM) for his or her data.

However, there are functionality constraints embedded in model type choices.

To view the forecast model (ARIMA) more closely, select **Modeling View → Tables → Parameter Estimates**.

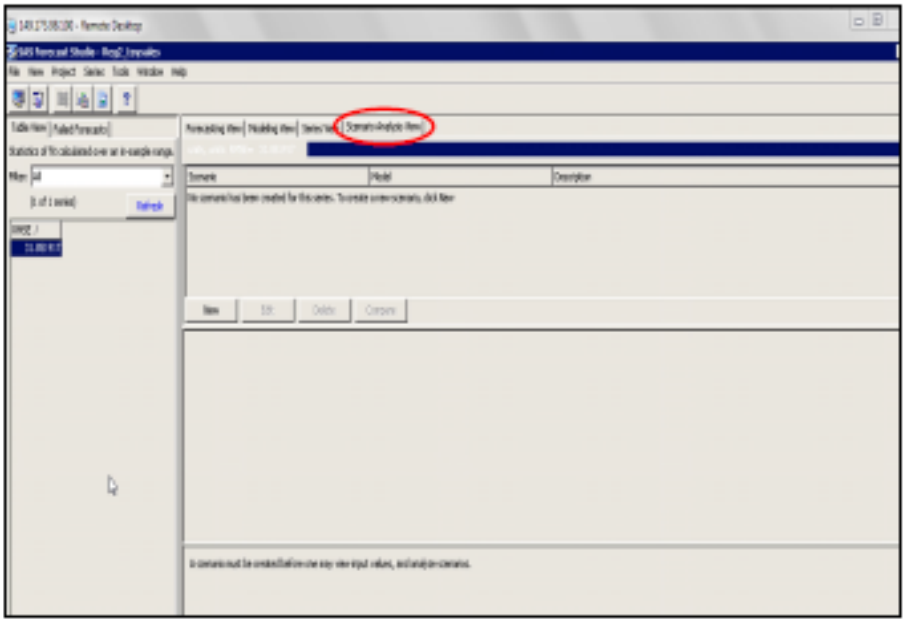


Two estimates are of primary interest to the majority of forecast analysts: the estimated intercept for the series is approximately 1400 units, and the lift associated with the binary, promotion input variable is about 52 units.

The estimate for the **pctpromo** variable represents the average effect of the promotion over all of the occurrences of the promotion in the range of the data. Chapter 4 discusses parameter estimate interpretation for binary variables in more detail.

Creating a Scenario Analysis

1. To explore this, begin by going to the Scenario Analysis view.

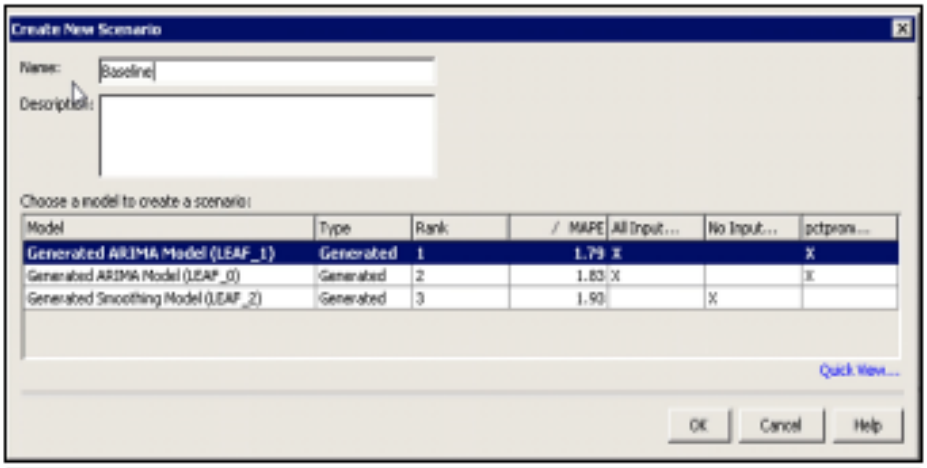


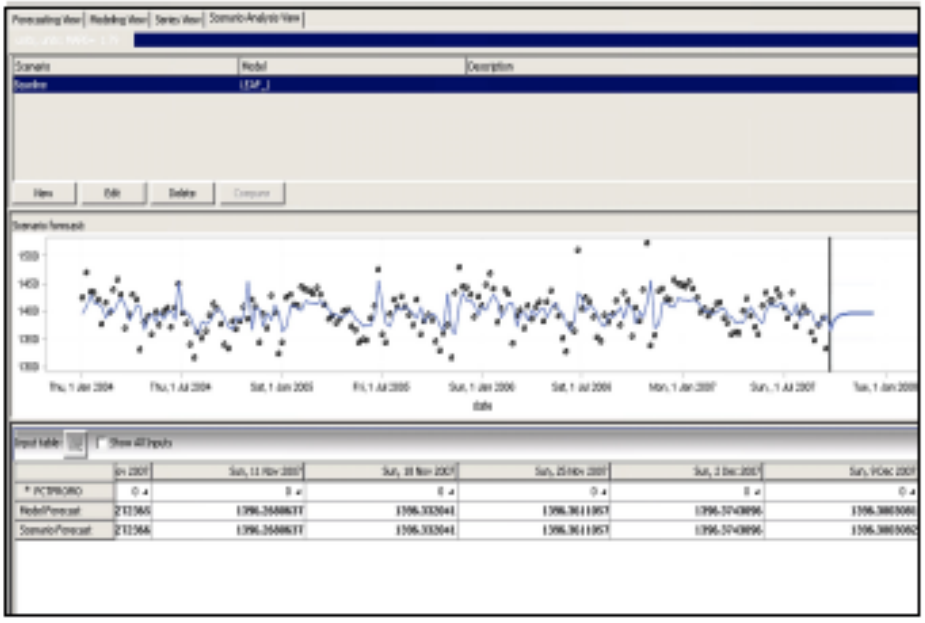
Currently, there are no scenarios defined.

One approach to using the Scenario Analysis view is to perform a sensitivity analysis. The first step in this analysis is to create a baseline scenario using the current promotion plan. Proposed changes to the current plan can be represented in alternative scenarios.

The baseline scenario is created based on the current promotion plan. Here, no changes are made to the lead values for the **pctpromo** variable.

2. Select **New**.

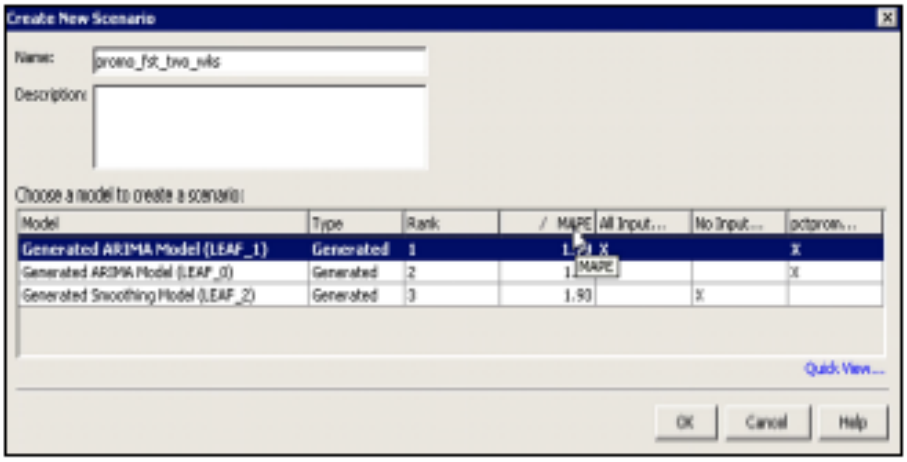




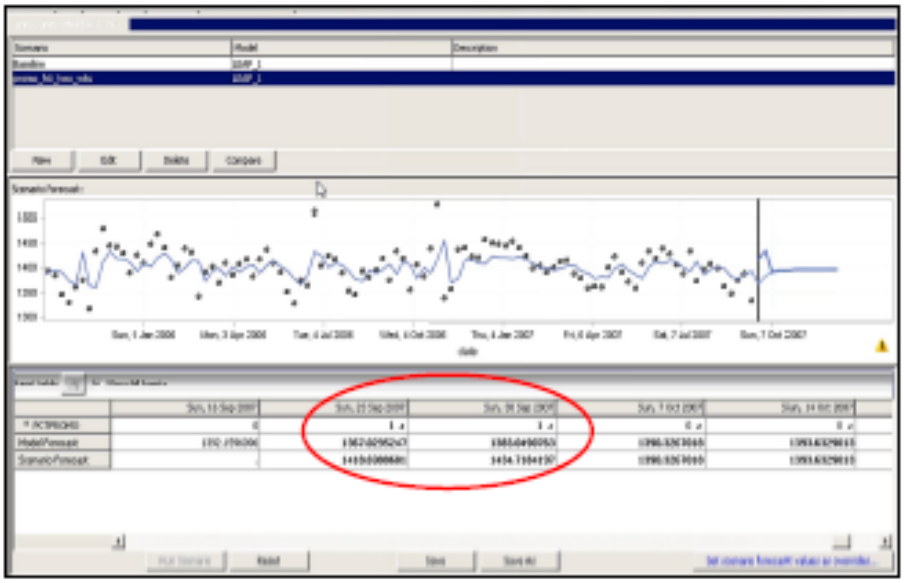
No promotions are currently set to run in the ***lead forecast horizon***.

Explore the effect of running the promotion for the first two weeks of the forecast horizon on forecast units of toy sales.

3. Select **New** again.
4. Name the alternative scenario **promo_fst_two_wks**.
5. Click **OK**.



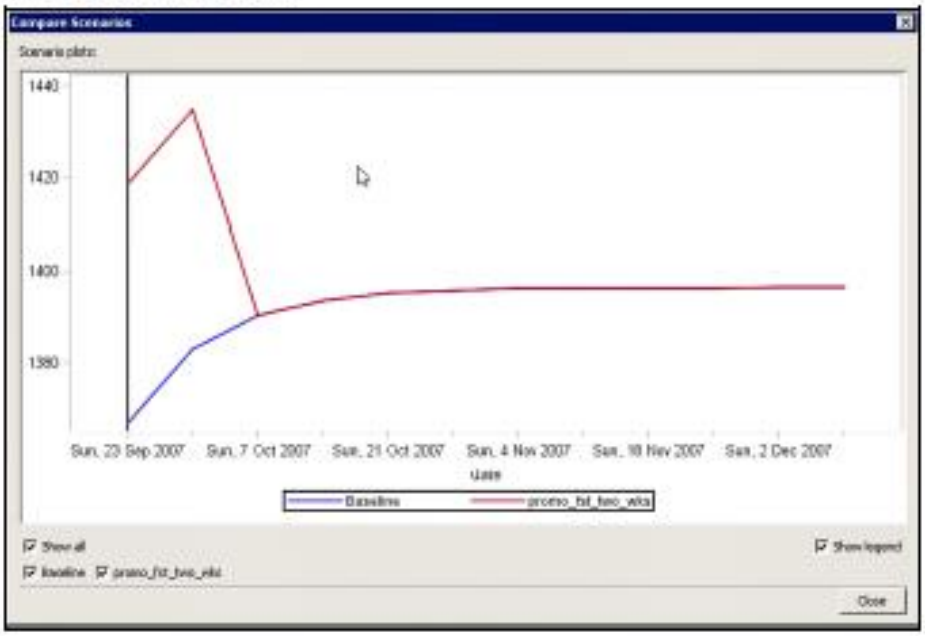
6. Change the table values for **pctpromo** from 0 to 1 for the first two weeks of the forecast horizon, September 23, 2007, and September 30, 2007.



7. Select **Run Scenario**.

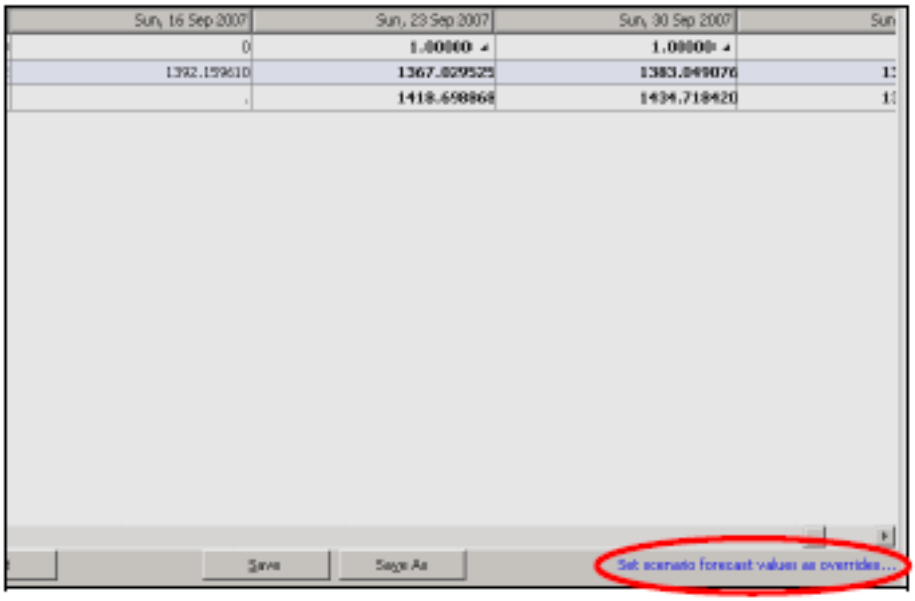
The **promo_fst_two_wks** scenario is active. The table compares the original forecast with the selected scenario forecast. Adjusting the baseline promotion plan to promote the product in the first two weeks of the forecast horizon increases sales of the product by approximately 3.5% in those two weeks.

8. Select **Compare Scenarios** to get an overlay plot that illustrates the effect of the promotion plan change in forecast units of sales.



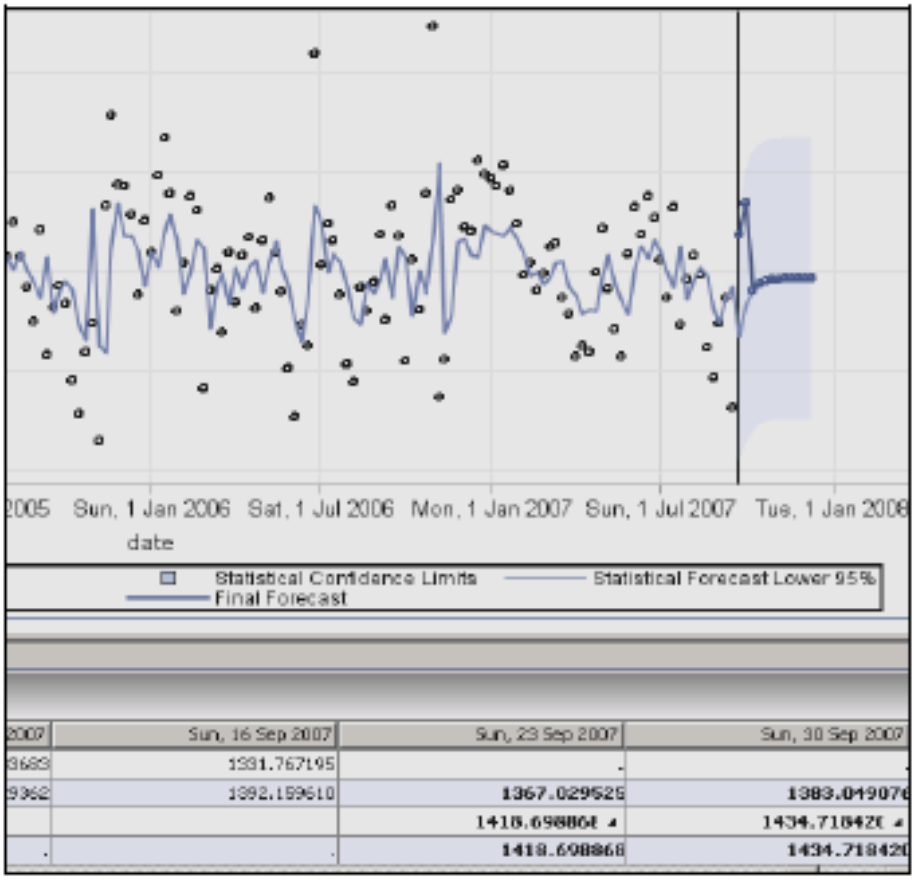
Now, assume that management has seen and approved the Scenario Analysis results. It is now necessary to change the lead values of **pctpromo**, and to regenerate forecasts to implement the new promotion plan. SAS Forecast Studio provides a way to do this automatically.

9. Select **Set scenario forecast values as overrides**.



Details of overrides on generated forecasts are presented in the next chapter.

Switching back to the Forecasting view confirms that the final forecast for the project accommodates the new promotion plan.



2007	Sun, 16 Sep 2007	Sun, 23 Sep 2007	Sun, 30 Sep 2007
3683	1391.767195	.	.
9362	1392.159610	1367.029525	1383.049076
		1418.698868	1434.718420
.	.	1418.698868	1434.718420