

Exploring SAS Forecast Studio



Generating Forecasts Automatically

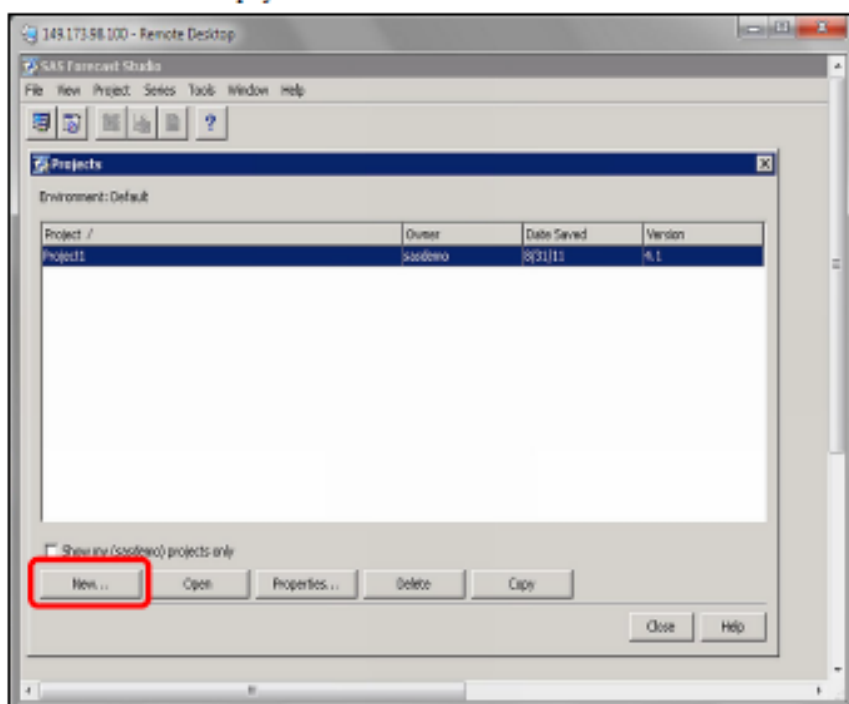
1. Enter your user name and password, and then click **Log On** to start the process of project creation.

The screenshot shows the SAS Forecast Studio 4.1 login window. The window has a title bar that says "SAS Forecast Studio". Inside, there's a blue header with the SAS logo and the text "SAS® Forecast Studio 4.1". Below the header, there's a section with the text "THE POWER TO KNOW." and a copyright notice: "Copyright © 2011, SAS Institute Inc., Cary, NC, USA. All Rights Reserved." The main area contains a form with the following fields: "SAS environment:" with a dropdown menu showing "Default SAS Environment"; "User ID:" with a text box containing "sasdmo"; "Password:" with a text box containing "*****"; and a checkbox labeled "Remember password" which is checked. At the bottom right, there are two buttons: "Log On" and "Cancel".



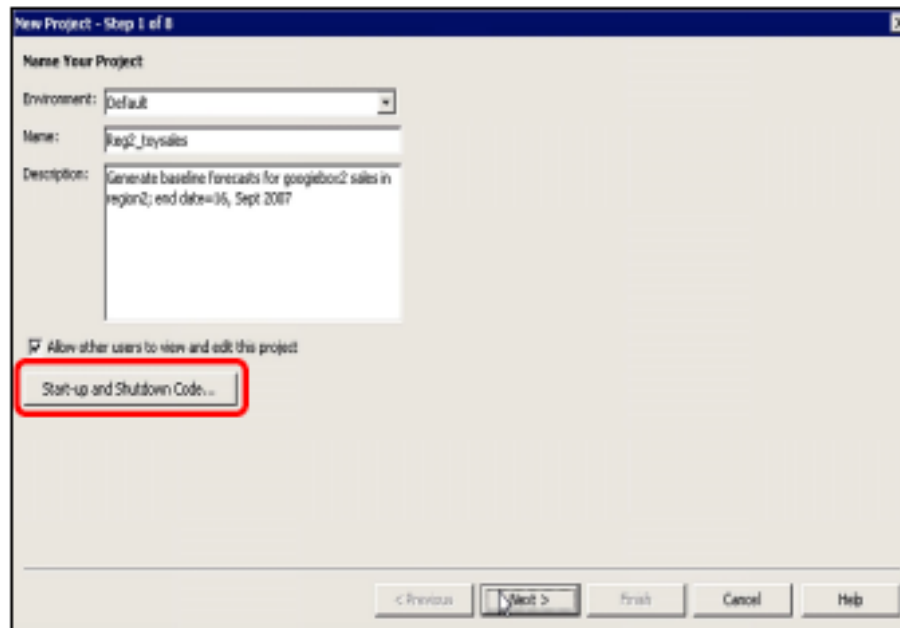
In this course, your user name, password, and server name are populated for you.

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.

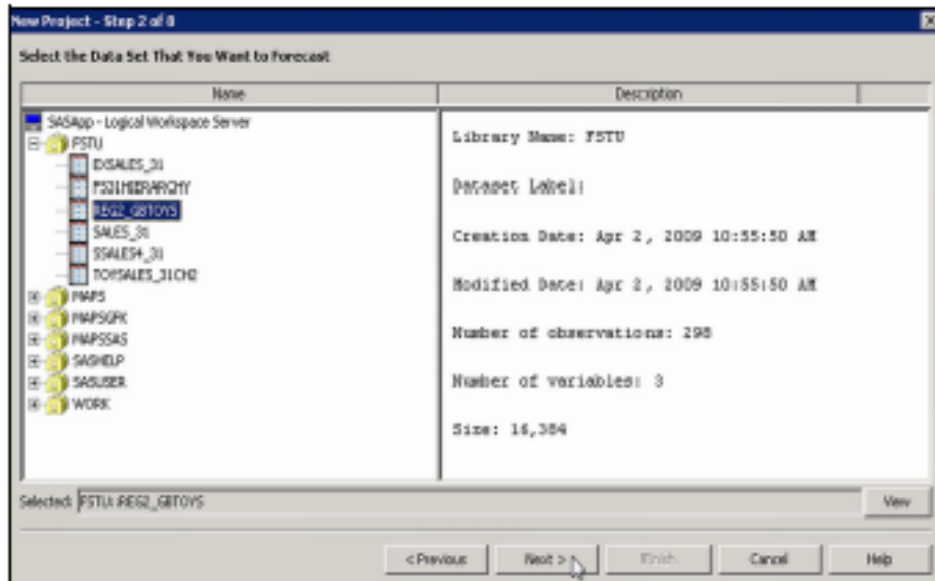
5. Select the **FSTU** library and the **REG2_GBTOYS** data set. Some summary information about the library and data set chosen is given in the panel on the right.



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

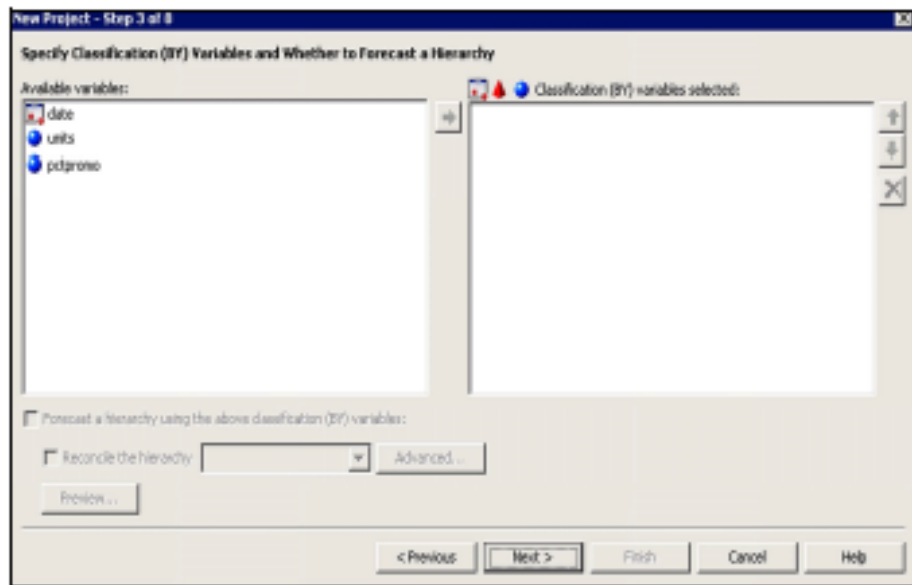


The libraries and data sets shown have been registered in the SAS Management Console

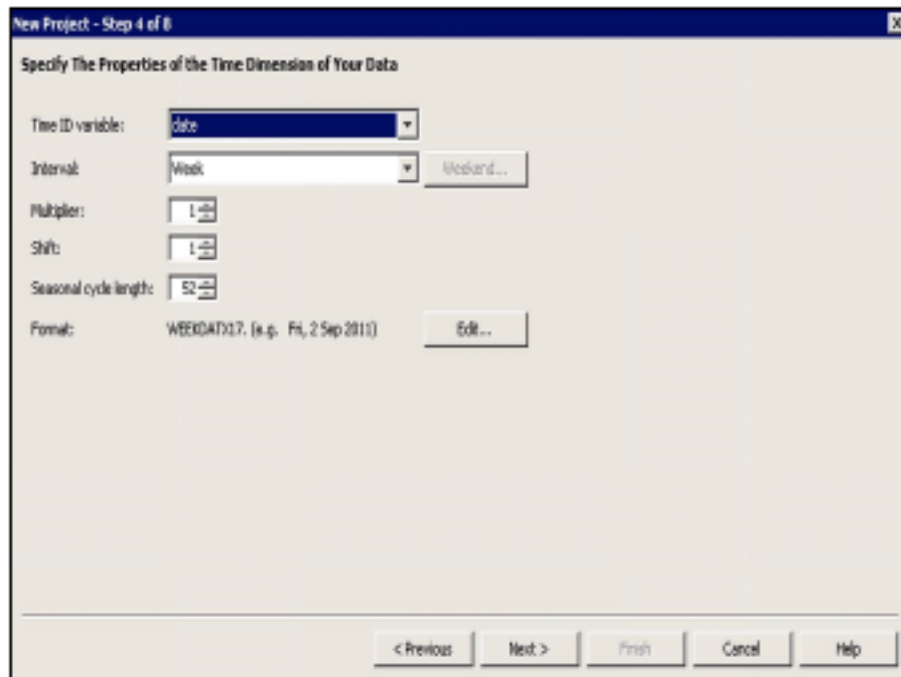


6. Click **Next**.

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.



New Project - Step 4 of 8

Specify The Properties of the Time Dimension of Your Data

Time ID variable:

Interval: Weekend...


Multiplier:

Shift:

Seasonal cycle length:

Format: WEEDATY17. (e.g., Fri, 2 Sep 2011) Edit...

< Previous Next > Finish Cancel Help

 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.

New Project - Step 5 of 8

Assign Roles to Variables in Your Data

You must specify at least one dependent variable.

Variable	Role	Accumulation	Usage in System-Generated Models
units	None		
pctpromo	None		

Adjustments...

< Previous Next > Finish Cancel Help

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

New Project - Step 5 of 8

Assign Roles to Variables in Your Data

You must specify at least one dependent variable.

Variable	Role	Accumulation	Usage in System-Generated Models
units	Dependent	Sum of values	Try to use
pctpromo	Independent	Average of values	Try to use

Adjustments...

< Previous Next > Finish Cancel Help

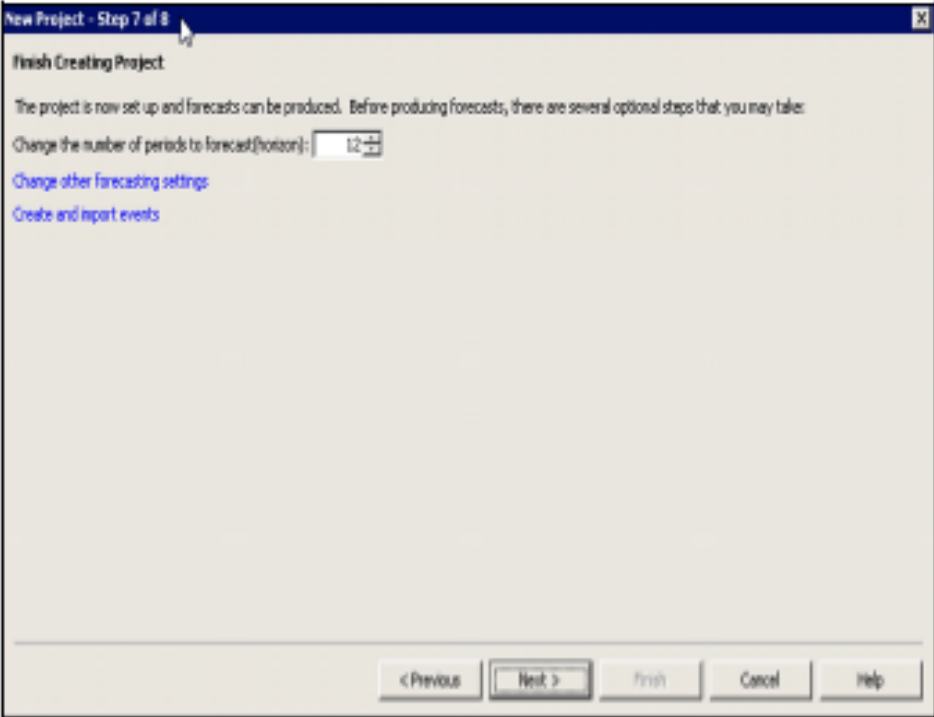
11. This step specifies data preparation options. Click Next.

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.



New Project - Step 7 of 8

Finish Creating Project

The project is now set up and forecasts can be produced. Before producing forecasts, there are several optional steps that you may take:

Change the number of periods to forecast(horizon):

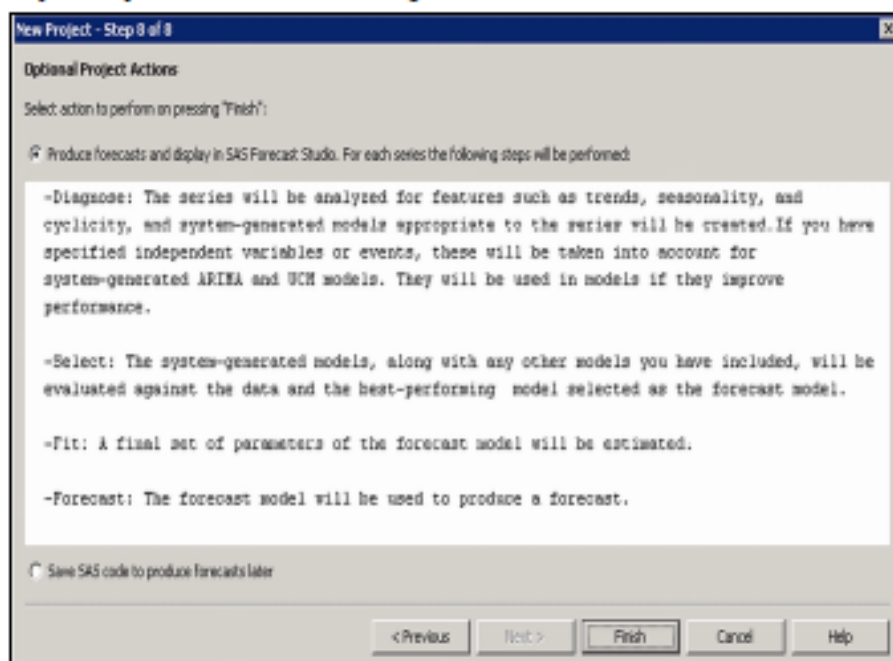
[Change other forecasting settings](#)

[Create and import events](#)

< Previous Next > Finish Cancel Help

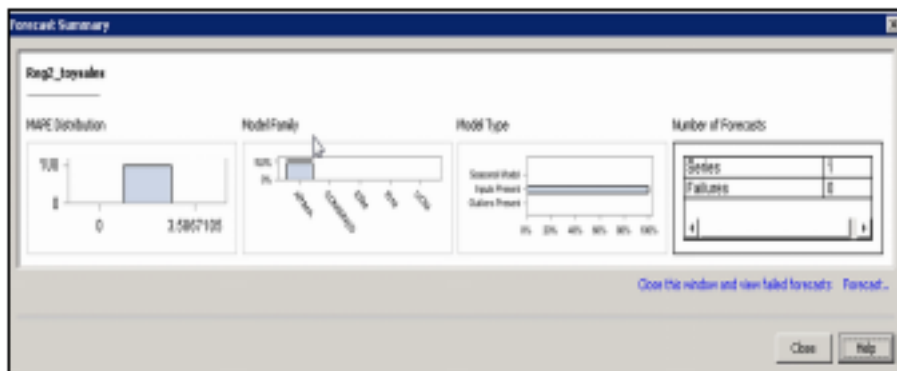
The fields **Change other forecasting settings** and **Create and import events** are discussed later in this course.

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



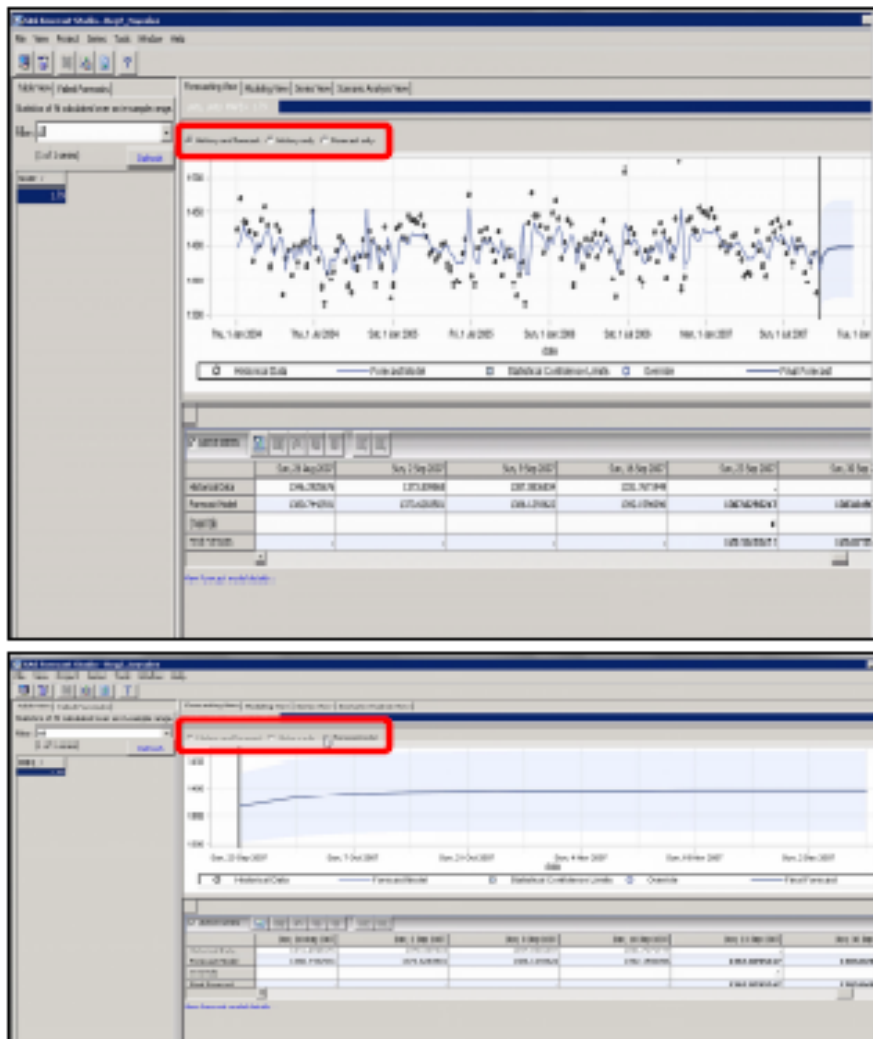
SAS Forecast Studio is a code generator. The default behavior, shown above, immediately submits the code, and results are generated and displayed in the user interface. Alternatively, the code can be saved and run later. Saved code can be modified.

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.



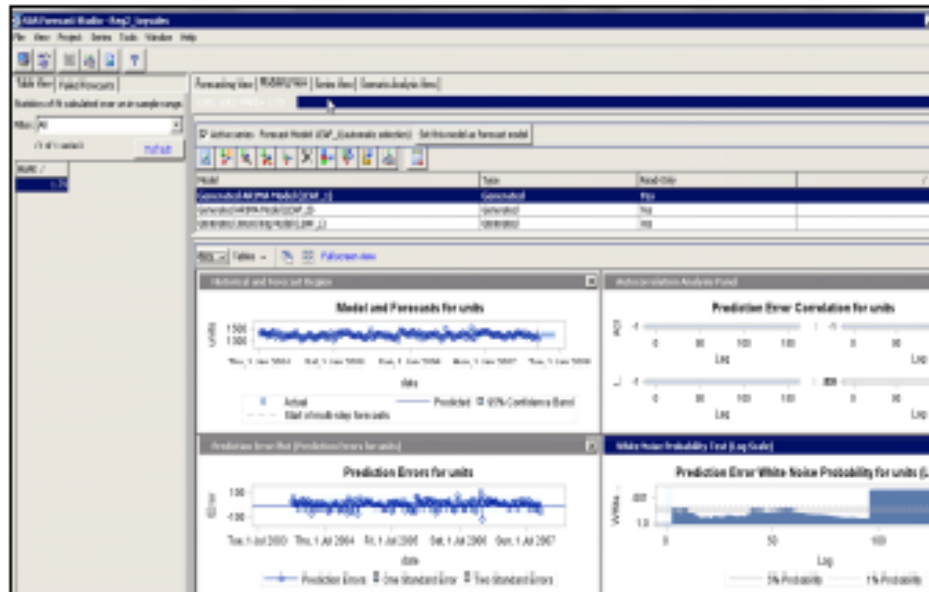
There are two main areas of information contained in the Forecasting view.

- 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.

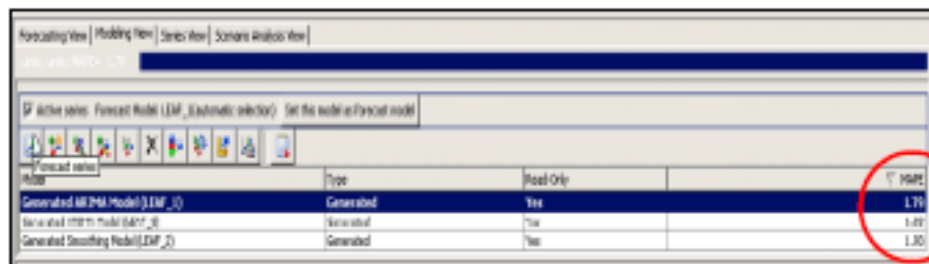
✍ By default, 95% prediction intervals are provided for future forecasts.

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.



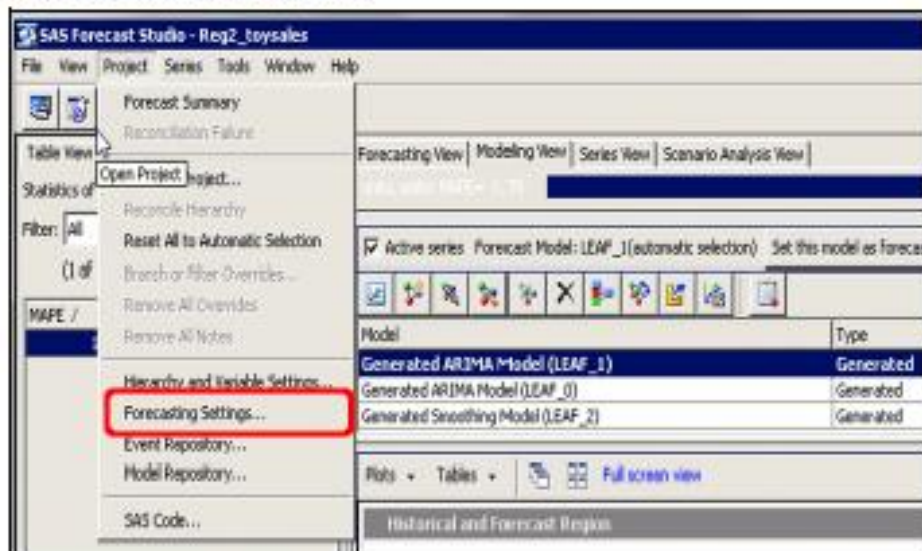
The forecast model for the series is the candidate model with the best fit. Here, this is the generated ARIMA model with the lowest associated MAPE (highlighted).



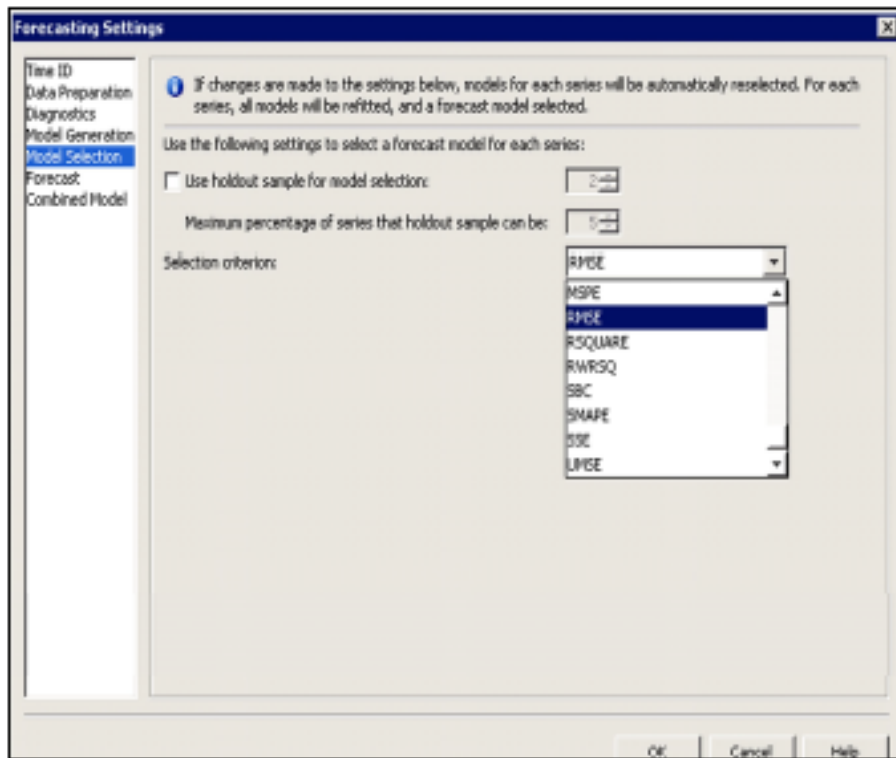


Setting the Project's Model Selection Statistic

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.

Forecasting View Modeling View Series View Smart Analysis View			
Active series: Forecast Model (SAP_Statistical selection) Set the model as forecast model			
Model	Type	Read-Only	RMSE
Generated ARIMA Model (SAP_S)	Generated	Yes	31.611167
Generated ARIMA Model (SAP_S)	Generated	Yes	32.179913
Generated Smoothing Model (SAP_S)	Generated	Yes	35.300894

- 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.

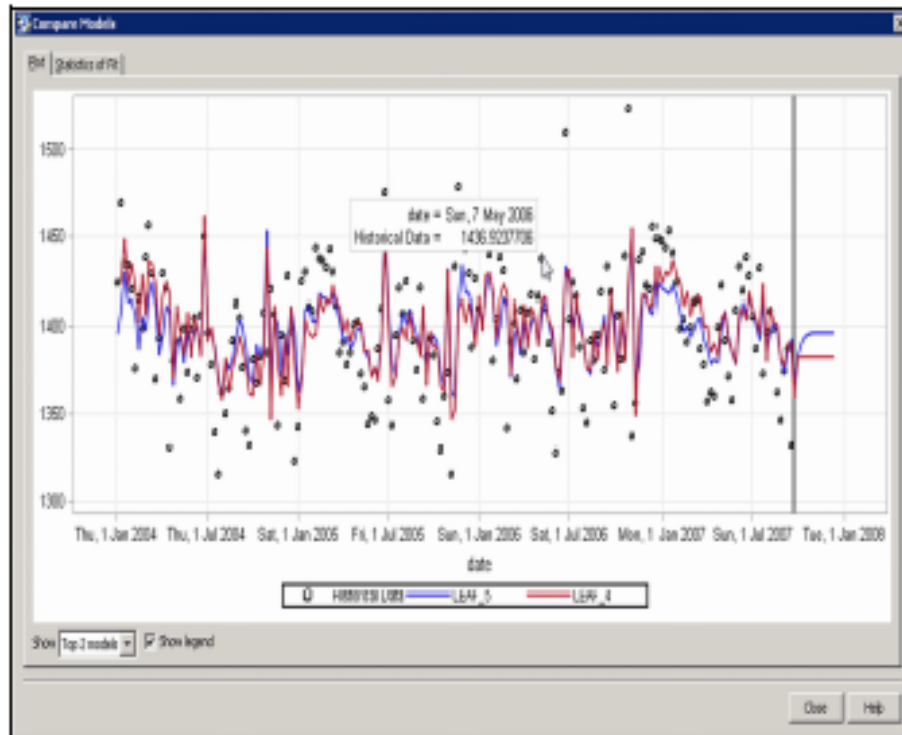
Compare Models

go: [Statistics of Fit]

	Selection Statistics of Fit									
Model /	APE	GMPE	GMSE	GMRE	RAE	RAPE	RMSE	RMPE	RMSE	RMPE
Generated ARIMA Model (LEAF 4)	7.86174	1.33	1.33	8.83	25.49091	1.83	68.92899	1.82	8.06	324.1000
Generated ARIMA Model (LEAF 5)	5.18821	1.32	1.32	8.76	25.08185	1.79	66.79879	1.79	8.04	322.9000
Generated Seasonal ARIMA (LEAF 6)	6.21798	1.29	1.29	8.66	26.96196	1.93	74.25093	1.93	8.91	383.2000

Close Help

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

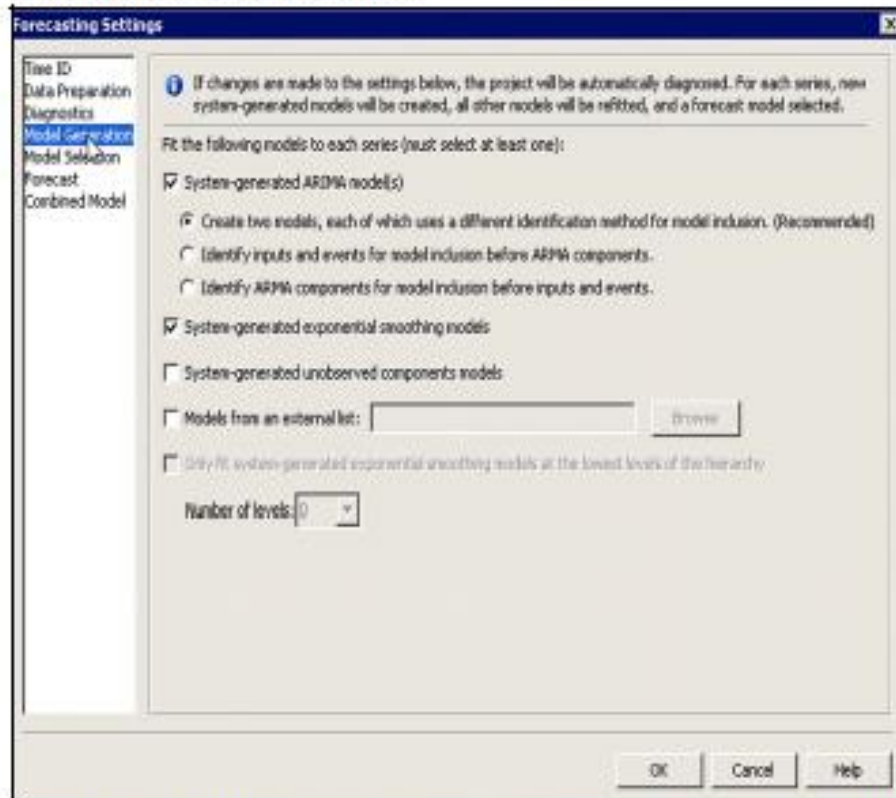


An overlay plot of the forecasts for the top two models is shown by default. Above, the project has been modified so that an overlay plot of the generated exponential smoothing model and the forecast model, ARIMA (LEAF 5), are shown.



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.

Forecasting View Modeling View Series View Scenario Analysis View					
File Edit View Options Help					
Active series: Forecast Model (LJAF_5) (automatic selection) Set this model as forecast model					
Model		Type	Read-Only	/ RMSE	
Generated ARIMA Model (LJAF_5)		Generated	Yes	31.851161	
Generated ARIMA Model (LJAF_4)		Generated	Yes	32.170953	
Generated Smoothing Model (LJAF_6)		Generated	Yes	35.300804	
File Edit View Options Help					
Parameter Estimates					
Component	Parameter	Estimate	Standard Error	t Value Approx Pr > t	
units	CONSTANT	1386.4	4.21881	331.83	<.0001
units	AR1_1	0.45438	0.08516	5.37	<.0001
pctpromo	SCALE	51.86904	12.11345	3.94	0.0001
Unbiasedness Test					
	Parameter	Estimate	Standard Error	t Value Approx Pr > t	
	Intercept	35.81379	108.80833	0.29	0.8432
	Scale	0.87438	0.12835	7.55	<.0001
Statistics of Fit					
Statistic	Value				
Degrees of Freedom Error	181				
Number of Observations	194				
Number of Observations Used	194				
Number of Missing Actuals	0				
Number of Missing Predicted Values	0				
Number of Model Parameters	3				
Total Sum of Squares	376291942				
Corrected Total Sum of Squares	254341.428				
Sum of Squares Error	186521.68				
Mean Square Error	1011.44864				
Root Mean Square Error	31.851161				
Unbiased Mean Square Error	1835.47588				
Forecast Values					
Obs	date	Forecast	Standard Error	Lower 95%	Upper 95%
195	Sun, 23 Sep 2007	1387.0295	32.1811	1304.1125	1429.9465
196	Sun, 30 Sep 2007	1383.0481	35.2584	1313.9438	1452.1543
197	Sun, 7 Oct 2007	1380.3367	35.9795	1328.8121	1480.9413
198	Sun, 14 Oct 2007	1383.8339	36.0815	1323.8113	1484.1845
199	Sun, 21 Oct 2007	1385.1349	36.0375	1324.8224	1485.7474
200	Sun, 28 Oct 2007	1383.9172	36.0330	1325.1940	1486.4403
201	Sun, 4 Nov 2007	1386.1372	36.0330	1325.9021	1486.7534
202	Sun, 11 Nov 2007	1386.2881	36.0341	1325.8424	1486.9837
203	Sun, 18 Nov 2007	1386.3330	36.0342	1325.7063	1486.9578
204	Sun, 25 Nov 2007	1386.3811	36.0342	1325.7364	1486.9868
205	Sun, 2 Dec 2007	1386.3743	36.0342	1325.7488	1487.0000
206	Sun, 9 Dec 2007	1386.3803	36.0342	1325.7548	1487.0000

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.

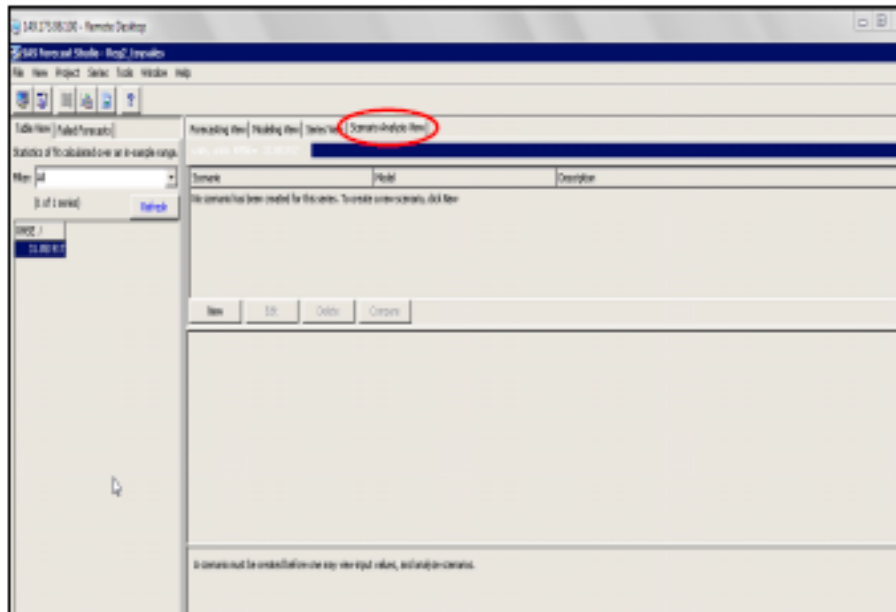


The estimate for the pctpromo variable represents the average effect of the promotion over all of the occurrences of the promotion in the range of the data. Chapter 4 discusses parameter estimate interpretation for binary variables in more detail.



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.

Create New Scenario

Name:

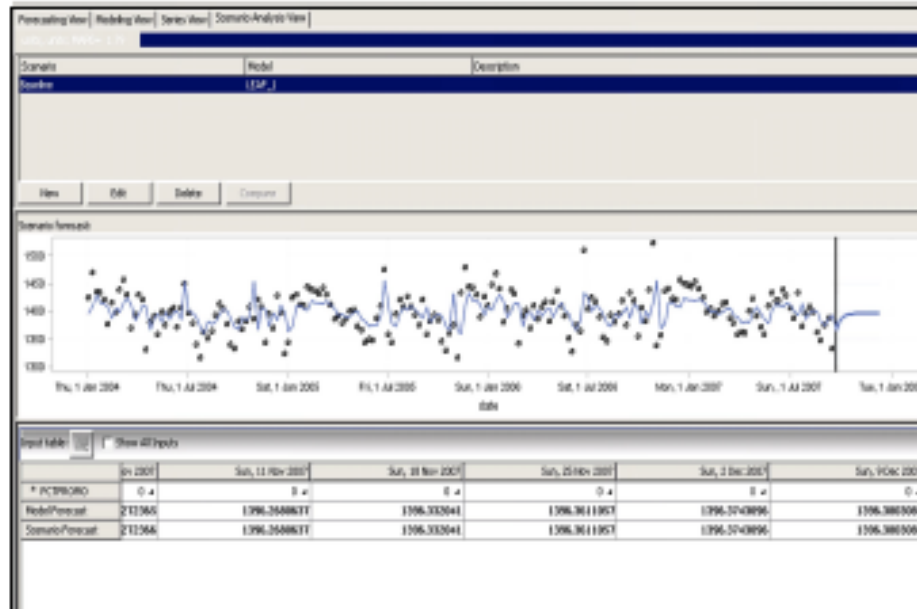
Description:

Choose a model to create a scenario:

Model	Type	Rank	/ MAPE	All Input...	No Input...	pctpromo...
Generated ARIMA Model (LEAF_1)	Generated	1	1.79	X		X
Generated ARIMA Model (LEAF_0)	Generated	2	1.83	X		X
Generated Smoothing Model (LEAF_2)	Generated	3	1.93		X	

[Quick View...](#)

OK Cancel Help



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**.

Create New Scenario

Name: promo_fst_two_wks

Description:

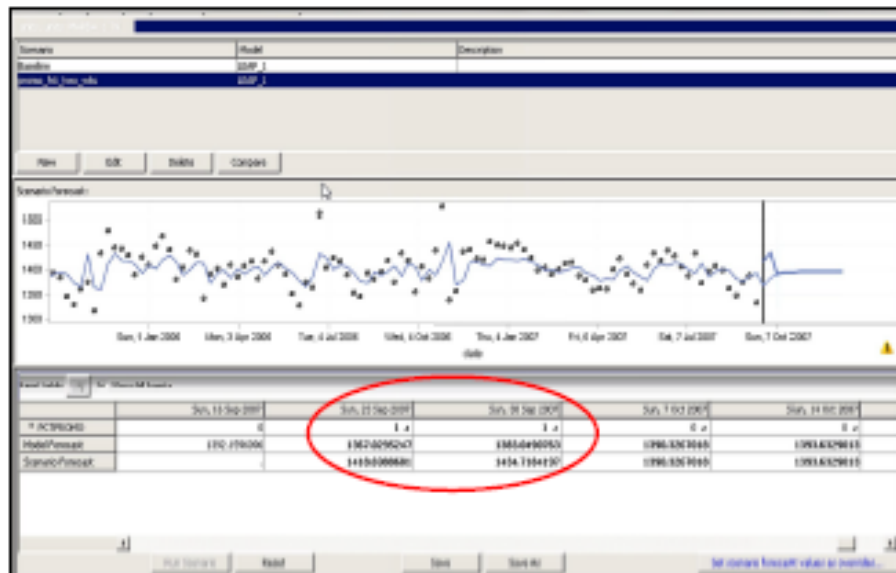
Choose a model to create a scenario:

Model	Type	Rank	/ MAPE	All Input...	No Input...	pctprom...
Generated ARIMA Model (LEAF_1)	Generated	1	1.91	X		X
Generated ARIMA Model (LEAF_0)	Generated	2	1.91	X		X
Generated Smoothing Model (LEAF_2)	Generated	3	1.93		X	

Quick View...

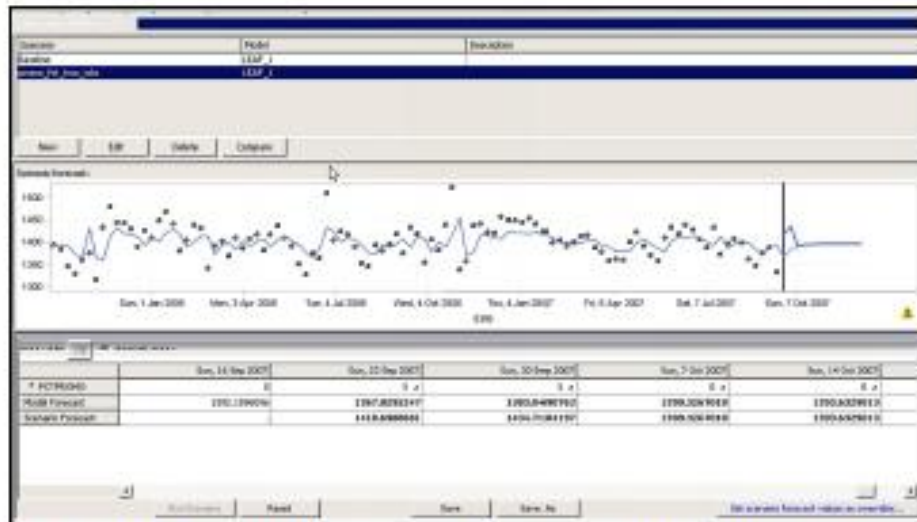
OK Cancel Help

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

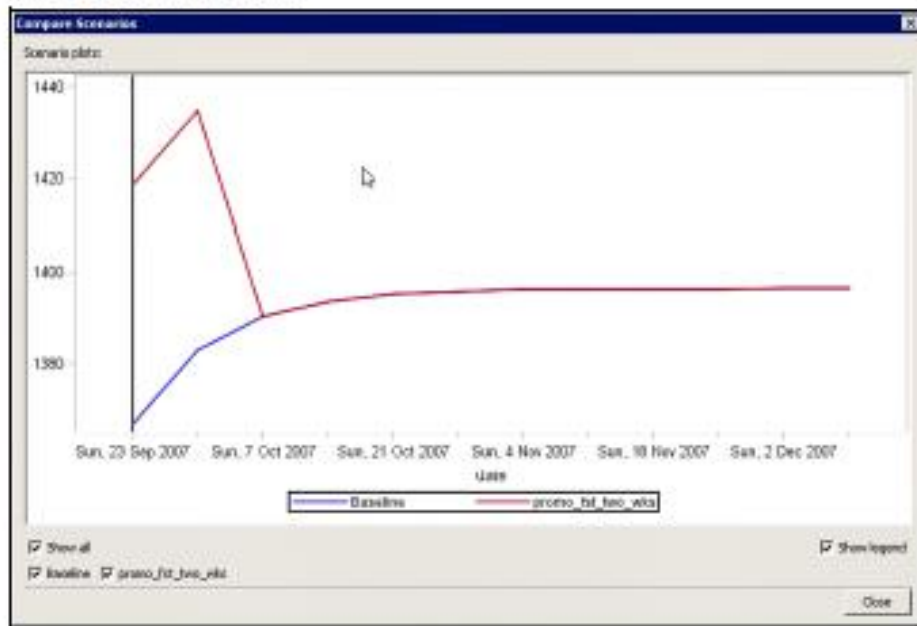


7. Select **Run Scenario**.

The `promo_fit_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 petpromote, 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.

Sun, 16 Sep 2007	Sun, 23 Sep 2007	Sun, 30 Sep 2007	Sun
0	1,00000	1,00000	
1392.159610	1367.029525	1383.049076	13
.	1418.698868	1434.718420	14



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.

