Chapter 23

Getting Started with Time Series Forecasting

Chapter Table of Contents

Chapter 23
Getting Started with Time Series Forecasting

This chapter outlines the forecasting process and introduces the major windows of the system through three example sessions.

The first example, beginning with the section "The Time Series Forecasting Window," shows how to use the system for fully automated forecasting of a set of time series. This example also introduces the system's features for viewing data and forecasts through tables and interactive graphs. It also shows how to save and restore forecasting work in SAS catalogs.

The second example, beginning with the section "Develop Models Window," introduces the features for developing the best forecasting models for individual time series. The chapter concludes with an example showing how to create dating variables for your data in the form expected by the system.

After working through the examples in this chapter, you should be able to

- select a data set of time series to work with and specify its periodicity and time ID variable
- use the automatic forecasting model selection feature to create forecasting models for the variables in a data set
- produce and save forecasts of variables in a data set
- examine your data and forecasts as tables of values and through interactive graphs
- save and restore your forecasting models using project files in a SAS catalog and edit project information
- use some of the model development features to fit and select forecasting models for individual time series variables.

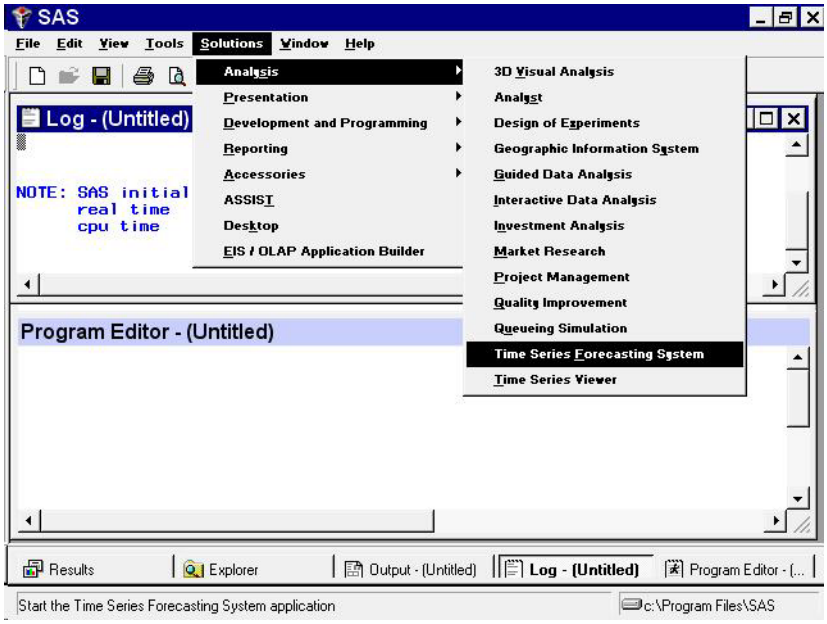
This chapter introduces these topics and will help you get started using the system. Later chapters present these topics in greater detail and document more advanced features and options.

The Time Series Forecasting Window

There are several ways to get to the Time Series Forecasting System. If you prefer to use commands, invoke the system by entering `forecast` on the command line. You can optionally specify additional information on the command line; see Chapter 28, "Command Reference," for details.

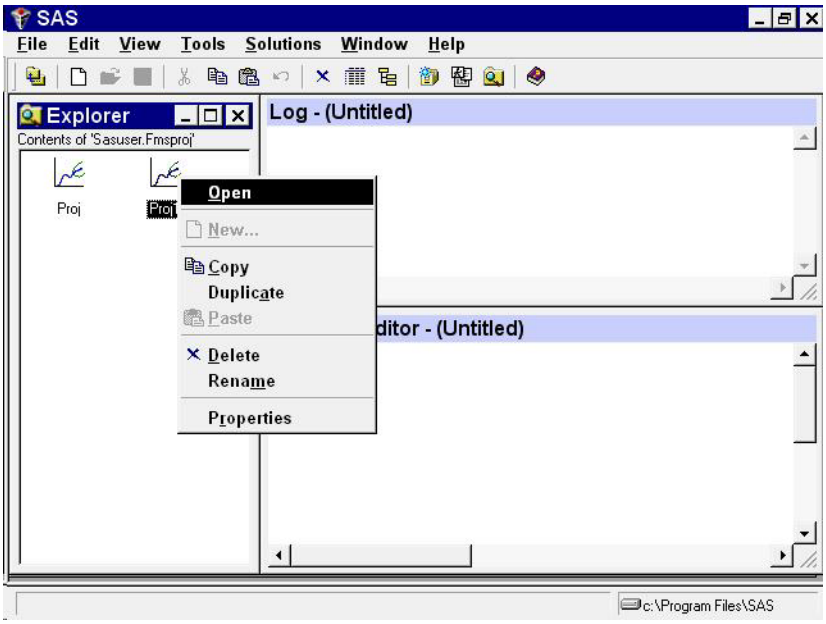
If you are using the SAS Windowing Environment with pull-down menus, select the `Solutions` menu from the menu bar, select the `Analysis` item, and then select `Time Series Forecasting System`, as shown in Display 23.1.

Display 23.1. Time Series Forecasting System Menu Selection



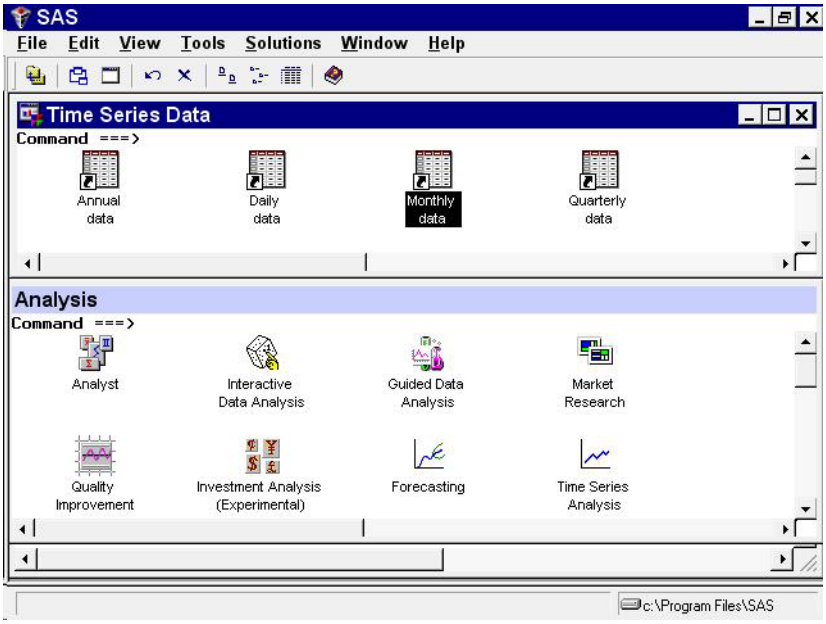
You can invoke the Forecasting System from the SAS Explorer window by opening an existing forecasting project. By default these projects are stored in the Fmsproj catalog in the Sasuser library. Select Sasuser in the Explorer to display its contents. Then select Fmsproj. This catalog is created the first time you use the Forecasting System. If you have saved projects, they appear in the Explorer with the forecasting graph icon, as shown in Display 23.2. Double-click one of the projects, or select it with the right mouse button and then select Open from the pop-up menu, as shown in the figure. This brings up the Forecasting System and opens the selected project.

Display 23.2. Opening a Project from the Explorer



To invoke the Forecasting System in the SAS Desktop environment, select the `Solutions` menu from the menu bar, select `Desktop`, and then open the Analysis folder. You can run the Time Series Forecasting System or the Time Series Viewer directly, or you can drag and drop. Display 23.3 illustrates dragging a data set (known as a table in the Desktop envionment) and dropping it on the Forecasting icon. In this example, the tables reside in a user-defined folder called *Time Series Data*.

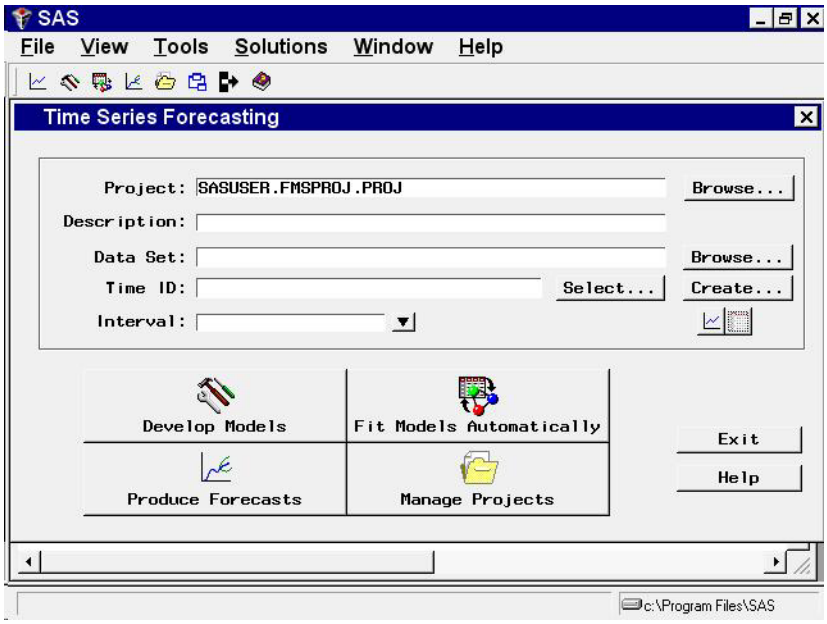
Display 23.3. Drag and Drop Using the SAS Desktop



If you are using SAS/ASSIST software, select the `Planning` button and then select `Forecasting` from the pop-up menu.

Any of these methods takes you to the `Time Series Forecasting` window, as shown in Display 23.4.

Display 23.4. Time Series Forecasting Window



At the top of the window is a data selection area for specifying a project file and the input data set containing historical data (the known past values) for the time series variables that you want to forecast. This area also contains buttons for bringing up viewers to explore your input data either graphically, one series at a time, or as a table, one data set at a time.

The `Project` and `Description` fields are used to specify a project file for saving and restoring forecasting models created by the system. Using project files is discussed later, and we will ignore these fields for now.

The lower part of the window contains six buttons:

Develop Models
: brings up the `Develop Models` window, which you use to develop and fit forecasting models interactively for individual time series.

Fit Models Automatically
: brings up the `Automatic Model Fitting` window, which you use to search automatically for the best forecasting model for multiple series in the input data set.

Produce Forecasts
: brings up the `Produce Forecasts` window, which you use to compute forecasts for all the variables in the input data set for which forecasting models have been fit.

Manage Projects
: brings up the `Manage Forecasting Project` window, which lists the time series for which you have fit forecasting models. You can drill down on a series to see the models that have been fit. You can delete series or models from the project, re-evaluate or refit models, and explore models and forecasts graphically or in tabular form.

Exit
: exits the Forecasting System.

Help
: displays information about the Forecasting System.

Outline of the Forecasting Process

The examples shown in the following sections illustrate the basic process you will use with the Forecasting System.

Specify the Input Data Set

Suppose you have a number of *time series*, variables recorded over time, for which you want to forecast future values. The past values of these time series are stored as variables in a SAS data set or data view. The observations of this data set correspond to regular time periods, such as days, weeks, or months. The first step in the forecasting process is to tell the system to use this data set by setting the `Data Set` field.

If your time series are not in a SAS data set, you must provide a way for the SAS System to access the data. You can use SAS features to read your data into a SAS data set; refer to *SAS Language Reference*. You can use a SAS/ACCESS product to establish a view of data in a database management system; refer to SAS/ACCESS documentation. You can use PROC SQL to create a SAS data view. You can use PROC DATASOURCE to read data from files supplied by supported data vendors; refer to Chapter 10, “The DATASOURCE Procedure,” for more details.

Provide a Valid Time ID Variable

To use the Forecasting System, your data set must be dated: the data set must contain a *time ID variable* that gives the date of each observation. The time ID variable must represent the observation dates with *SAS date values* or with *SAS datetime values* (for hourly data or other frequencies less than a day), or you can use a simple time index.

When SAS date values are used, the ID variable contains dates within the time periods corresponding to the observations. For example, for monthly data, the values for the time ID variable may be the date of the first day of the month corresponding to each observation, or the time ID variable may contain the date of the last day in the month. (Any date within the period will serve as the time ID for the observation.)

If your data set already contains a valid time ID variable with SAS date or datetime values, the next step is to specify this time ID variable in the `Time ID` field. If the time ID variable is named DATE, the system fills in the `Time ID` field automatically.

If your data set does not contain a time ID, you must add a valid time ID variable before beginning the forecasting process. The Forecasting System provides features that make this easy to do. See Chapter 24, "Creating Time ID Variables," for details.

Select and Fit a Forecasting Model for each Series

If you are using the automated model selection feature, the system performs this step for you and chooses a forecasting model for each series automatically. All you need to do is select the `Fit Models Automatically` button and then select the variables to fit models for.

If you want more control over forecasting model selection, you can select the `Develop Models` button, select the series you want to forecast, and use the Develop Models window to specify a forecasting model. As part of this process, you may use the `Time Series Viewer` and `Model Viewer` graphical tools. Once you have selected a model for the first series, you can select a different series to work with and repeat the model development process until you have created forecasting models for all the series you want to forecast.

The system provides many features to help you choose the best forecasting model for each series. The features of the Develop Models window and graphical viewer tools are introduced in later sections.

Produce the Forecasts

Once a forecasting model has been fit for each series, select the `Produce Forecasts` button and use the Produce Forecasts window to compute forecast values and store them in a SAS data set.

Save Your Work

If you want only a single forecast, your task is now complete. But you may want to produce updated forecasts later, as more data becomes available. In this case, you want to save the forecasting models you have created, so that you will not need to repeat the model selection and fitting process.

To save your work, fill in the `Project` field with the name of a SAS catalog member in which the system will store the model information when you exit the system. Later, you will select the same catalog member name when you first enter the Forecasting System, and the model information will be reloaded.

Note that any number of people can work with the same project file. If you are working on a forecasting project as part of a team, you should take care to avoid conflicting updates to the project file by different team members.

Summary

This is the basic outline of how the Forecasting System works. The system offers many other features and options that you may need to use (for example, the time range of the data used to fit models and how far into the future to forecast). These options will become apparent as you work with the Forecasting System.

As an introductory example, the following sections use the `Automatic Model Fitting` and `Produce Forecasts` windows to perform automated forecasting of the series in an example data set.

The Input Data Set

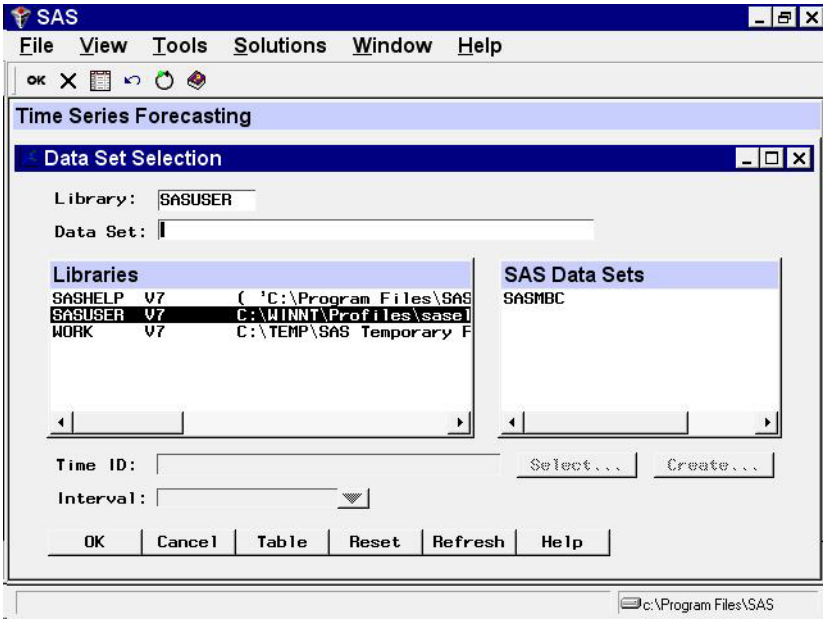
As the first step, you must specify the input data set.

The `Data Set` field in the Time Series Forecasting window gives the name of the input data set containing the time series to forecast. Initially, this field is blank. You can specify the input data set by typing the data set name in this field. Alternatively, you can select the Browse button at the right of the `Data Set` field to select the data set from a list, as shown in the following section.

The Data Set Selection Window

Select the Browse button to the right of the Data Set field. This brings up the `Data Set Selection` window, as shown in Display 23.5.

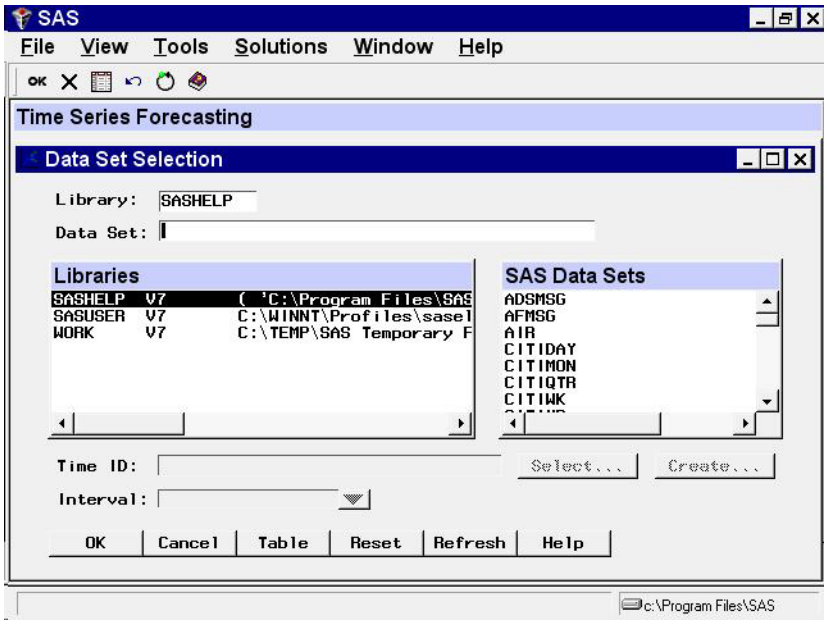
Display 23.5. Data Set Selection Window



The `Libraries` list shows the SAS librefs that are currently allocated in your SAS session. Initially, the SASUSER library is selected, and the `SAS Data Sets` list shows the data sets available in your SASUSER library.

In the `Libraries` list, select the row that starts with SASHELP. The Data Set Selection window now lists the data sets in the SASHELP library, as shown in Display 23.6.

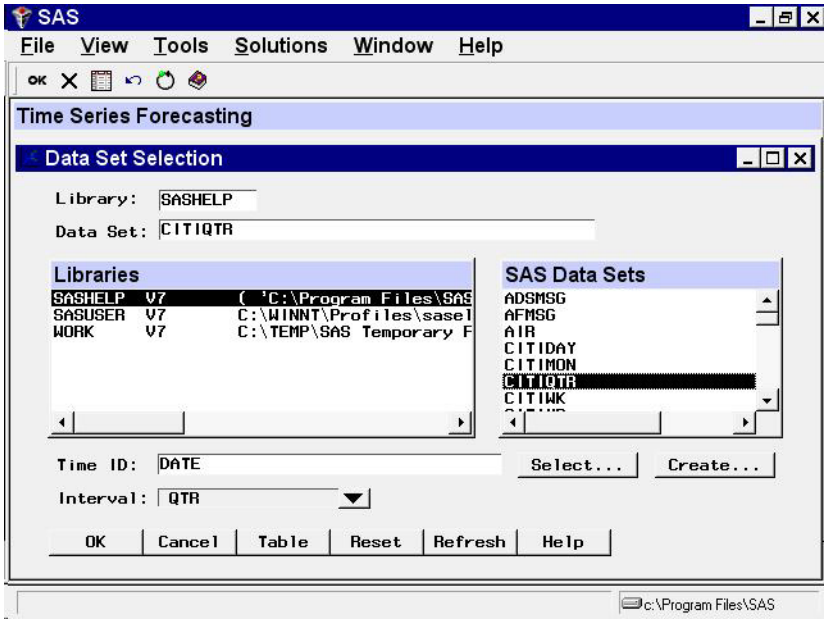
Display 23.6. SASHELP Library



Use the vertical scroll bar on the `SAS Data Sets` list to scroll down the list until the data set CITIQTR appears. Then select the CITIQTR row. This selects the data set SASHELP.CITIQTR as the input data set.

Display 23.7 shows the Data Set Selection window after selection of CITIQTR from the SAS Data Sets list.

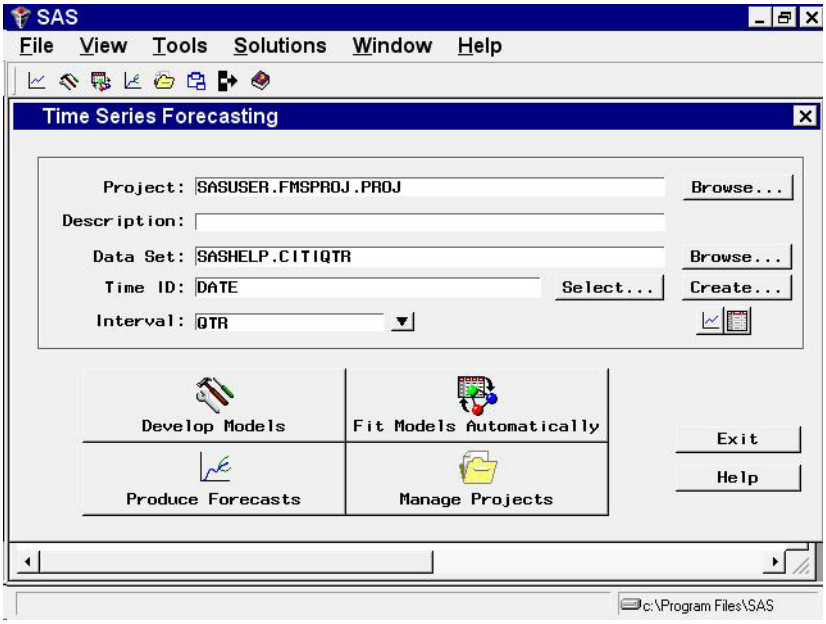
Display 23.7. CITIQTR Data Set Selected



Note that the `Time ID` field is now set to DATE and the `Interval` field is set to QTR. These fields are explained in the following section.

Now select the OK button to complete selection of the CITIQTR data set. This closes the Data Set Selection window and returns to the Time Series Forecasting window, as shown in Display 23.8.

Display 23.8. Time Series Forecasting Window



Time Series Data Sets, ID variables, and Time Intervals

Before you continue with the example, it is worthwhile to consider how the system determined the values for the Time ID and Interval fields in the Data Set Selection window.

The Forecasting System requires that the input data set contain time series observations, with one observation for each time period. The observations must be sorted in increasing time order, and there must be no gaps in the sequence of observations. The time period of each observation must be identified by an ID variable, which is shown in the Time ID field.

If the data set contains a variable named DATE, TIME, or DATETIME, the system assumes that that variable is the SAS date or datetime valued ID variable, and the Time ID field is filled in automatically. The time ID variable for the SASHELP.CITIQTR data set is named DATE, and therefore the system set the Time ID field to DATE.

If the time ID variable for a data set is not named DATE, TIME, or DATETIME, you must specify the time ID variable name. You can specify the time ID variable either by typing the ID variable name in the Time ID field or by clicking the Select button.

If your data set does not contain a time ID variable with SAS date values, you can add a time ID variable using one of the windows described in Chapter 24, “Creating Time ID Variables.”

Once the time ID variable is known, the Forecasting System examines the ID values to determine the time interval between observations. The data set SASHELP.CITIQTR contains quarterly observations. Therefore, the system determined that the data have a quarterly interval, and set the `Interval` field to QTR.

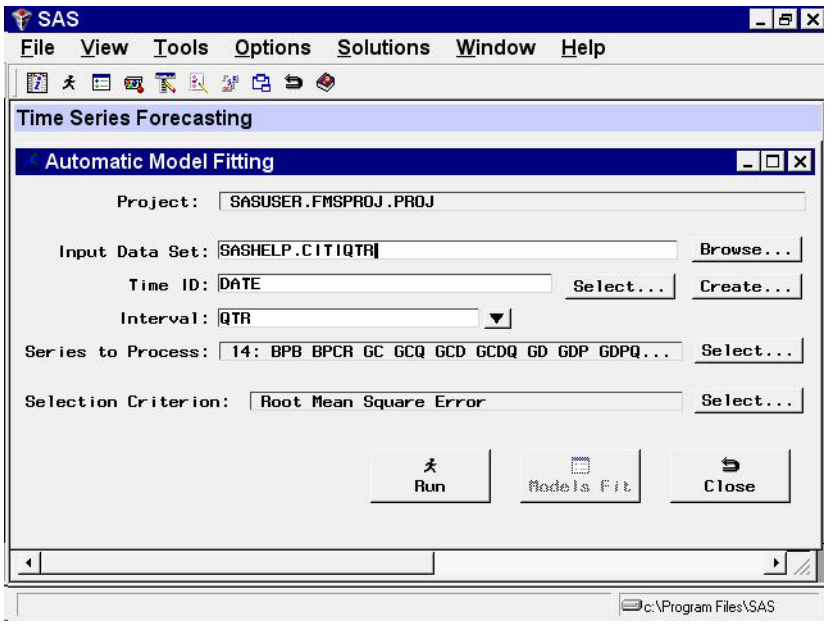
If the system cannot determine the data frequency from the values of the time ID variable, you must specify the time interval between observations. You can specify the time interval using the `Interval` combo box. In addition to the interval names provided in the pop-up list, you can type in more complex interval names to specify an interval that is a multiple of other intervals or that has date values in the middle of the interval (such as monthly data with time ID values falling on the 10th day of the month).

See Chapter 2, "Working with Time Series Data," and Chapter 3, "Date Intervals, Formats, and Functions," for more information on time intervals, SAS date values, and ID variables for time series data sets.

Automatic Model Fitting Window

Before you can produce forecasts, you must fit forecasting models to the time series. Select the `Fit Models Automatically` button. This brings up the `Automatic Model Fitting` window, as shown in Display 23.9.

Display 23.9. Automatic Model Fitting Window



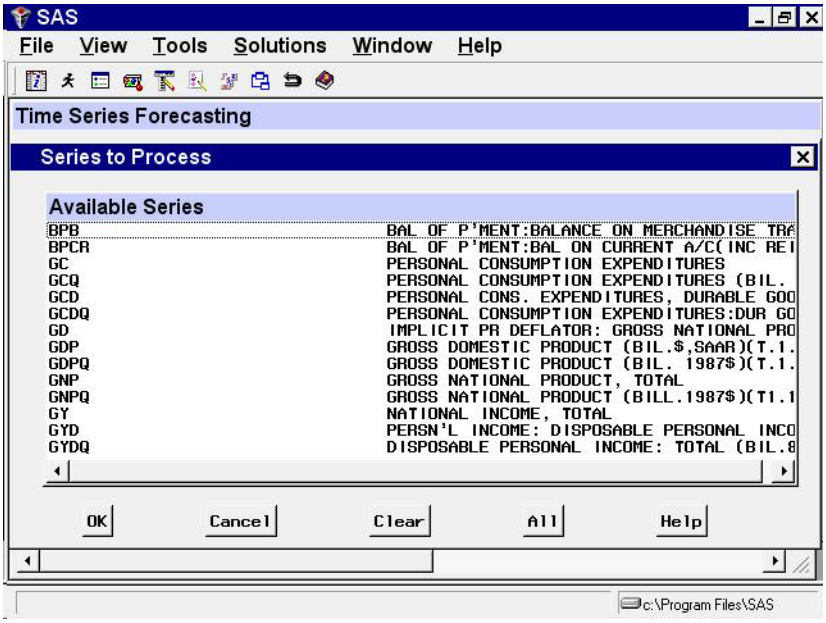
The first part of the Automatic Model Fitting window confirms the project filename and the input data set name.

The `Series to Process` field shows the number and lists the names of the variables in the input data set to which the Automatic Model Fitting process will be applied. By default, all numeric variables (except the time ID variable) are processed.

However, you can specify that models be generated for only a select subset of these variables.

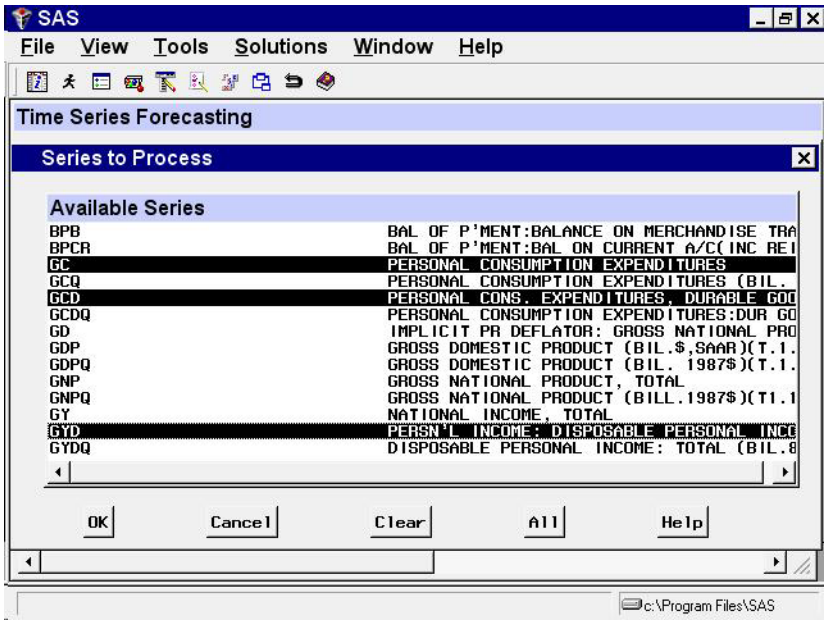
Click the Select button to the right of the `Series to Process` field. This brings up the `Series to Process` window, as shown in Display 23.10.

Display 23.10. Series to Process Window



Use the mouse and the CTRL key to select the personal consumption expenditures series (GC), the personal consumption expenditures for durable goods series (GCD), and the disposable personal income series (GYD), as shown in Display 23.11. (Remember to hold down the CTRL key as you make the selections; otherwise, selecting a second series will deselect the first.)

Display 23.11. Selecting Series for Automatic Model Fitting

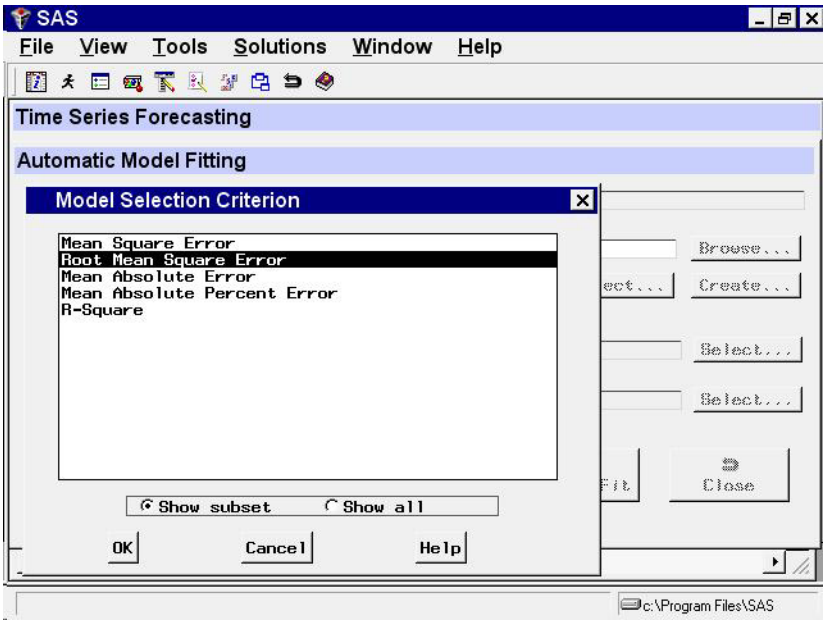


Now select the `OK` button. This returns you to the Automatic Model Fitting window. The `Series to Process` field now shows the selected variables.

The `Selection Criterion` field shows the goodness-of-fit measure that the Forecasting System will use to select the best fitting model for each series. By default, the selection criterion is the root mean square error. To illustrate how you can control the selection criterion, this example will use the mean absolute percent error to select the best fitting models.

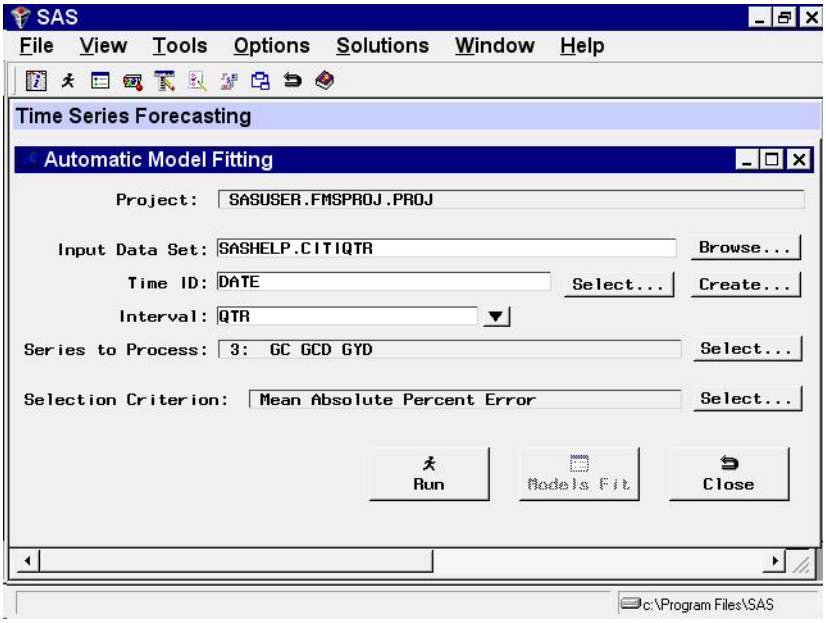
Click the Select button to the right of the Selection Criterion field. This brings up a list of statistics of fit, as shown in Display 23.12.

Display 23.12. Choosing the Model Selection Criterion



Select *Mean Absolute Percent Error* and then select the OK button. The Automatic Model Fitting window now appears as shown in Display 23.13.

Display 23.13. Automatic Model Fitting Window

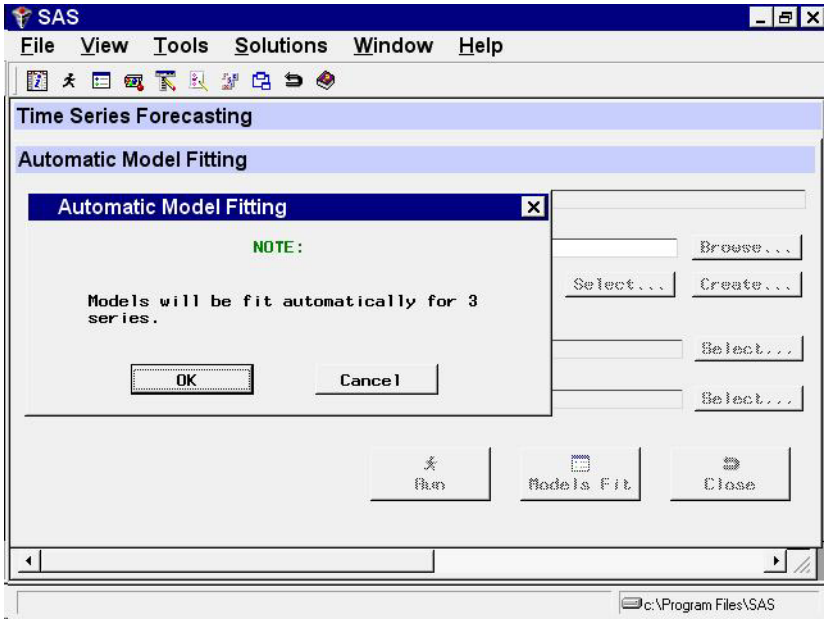


Now that all the options are set appropriately, select the Run button.

The Forecasting System now displays a notice, shown in Display 23.14, confirming that models will be fit for 3 series using the automatic forecasting model search feature. This prompt is displayed because it is possible to fit models for a large number of series at once, which may take a lot of time, and so the system gives you a chance to cancel if you accidentally ask to fit models for more series than you intended.

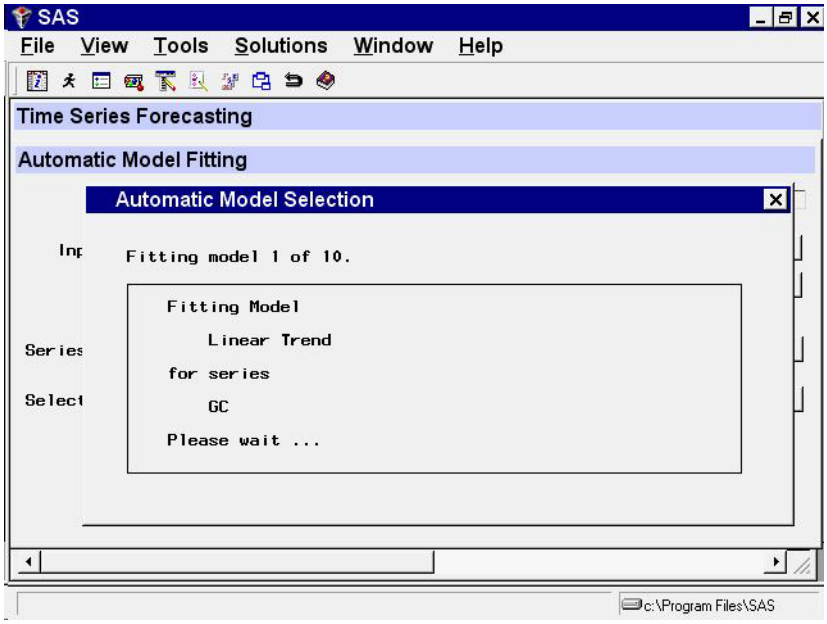
Select the OK button.

Display 23.14. Automatic Model Fitting Note



The system now fits several forecasting models to each of the three series you selected. While the models are being fit, the Forecasting System displays notices indicating what it is doing so that you can observe its progress, as shown in Display 23.15.

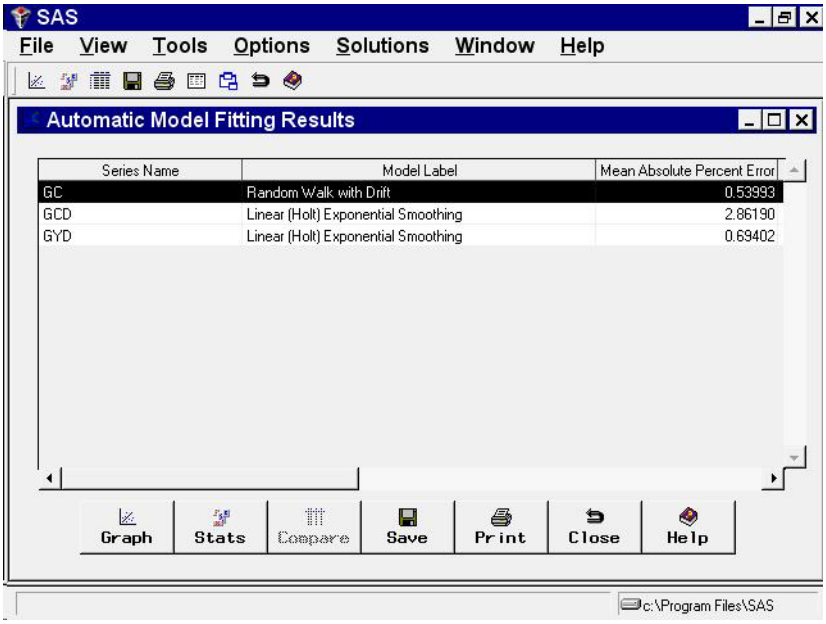
Display 23.15. "Working" Notice



For each series, the system saves the model that produces the smallest mean absolute percent error. You can have the system save all the models fit by selecting `Automatic Fit` from the `Options` pull-down menu.

After the Automatic Model Fitting process has completed, the results are displayed in the `Automatic Model Fitting Results` window, as shown in Display 23.16.

Display 23.16. Automatic Model Fitting Results



This resizable window shows the list of series names and descriptive labels for the forecasting models chosen for them, as well as the values of the model selection criterion and other statistics of fit. Select the `Close` button.

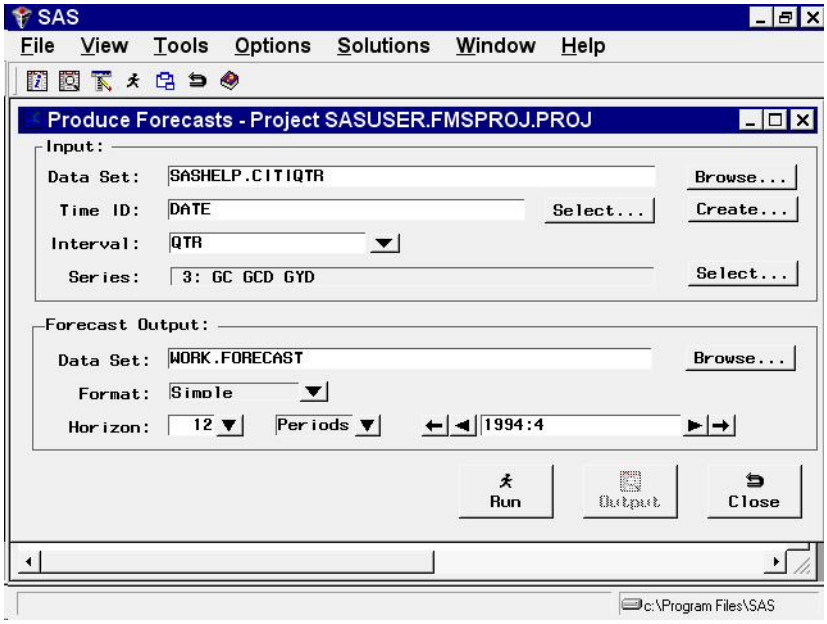
This returns you to the Automatic Model Fitting window. You can now fit models for other series in this data set or change to a different data set and fit models for series in the new data set.

Select the `Close` button to return to the `Time Series Forecasting` window.

Produce Forecasts Window

Now that you have forecasting models for these three series, you are ready to produce forecasts. Select the `Produce Forecasts` button. This brings up the `Produce Forecasts` window, as shown in Display 23.17.

Display 23.17. Produce Forecasts Window

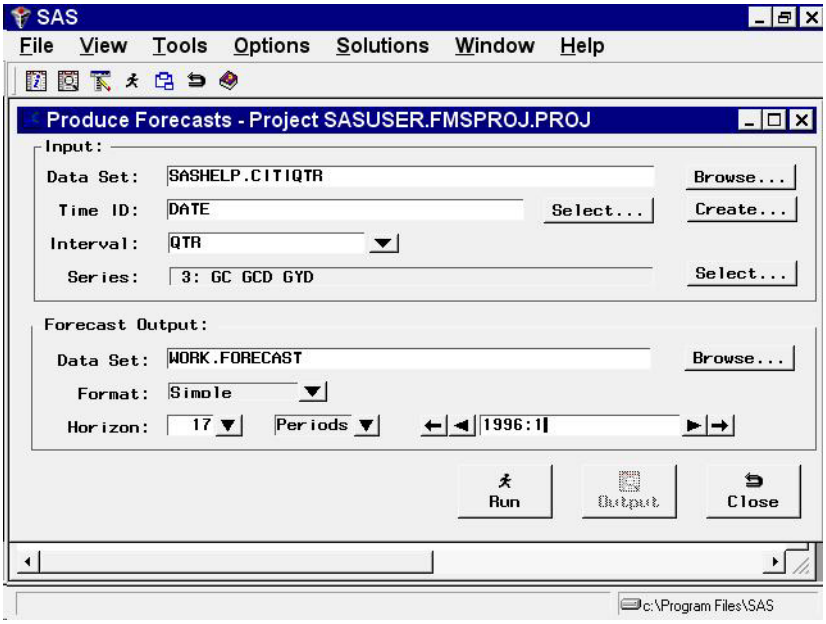


The Produce Forecasts window shows the input data set information and indicates the variables in the input data set for which forecasting models exist. Forecasts will be produced for these series. If you want to produce forecasts for only some of these series, use the control arrow at the right of the `Series` field to select the series to forecast. The `Data Set` field in the `Forecast Output` box contains the name of the SAS data set in which the system will store the forecasts. The default output data set is WORK.FORECAST.

You can set the forecast horizon using the controls on the line labeled `Horizon`. The default horizon is 12 periods. You can change it using number of periods, number of years, or the date of the last forecast period. Position the cursor in the date field and change the forecast ending date to 1 January 1996 by typing `jan1996` and pressing the ENTER key.

The window now appears as shown in Display 23.18.

Display 23.18. Produce Forecasts Window



Now select the Run button to produce the forecasts. The system indicates that the forecasts have been stored in the output data set. Select OK to dismiss the notice.

The Forecast Data Set

The Forecasting System can save the forecasts to a SAS data set in three different formats. Depending on your needs, you may find one of these output formats more convenient. The output data set format is controlled by the Format combo box. You can select the following output formats. The simple format is the default.

Simple — The data set contains time ID variables and the forecast variables, and it contains one observation per time period. Observations for earlier time periods contain actual values copied from the input data set; later observations contain the forecasts.

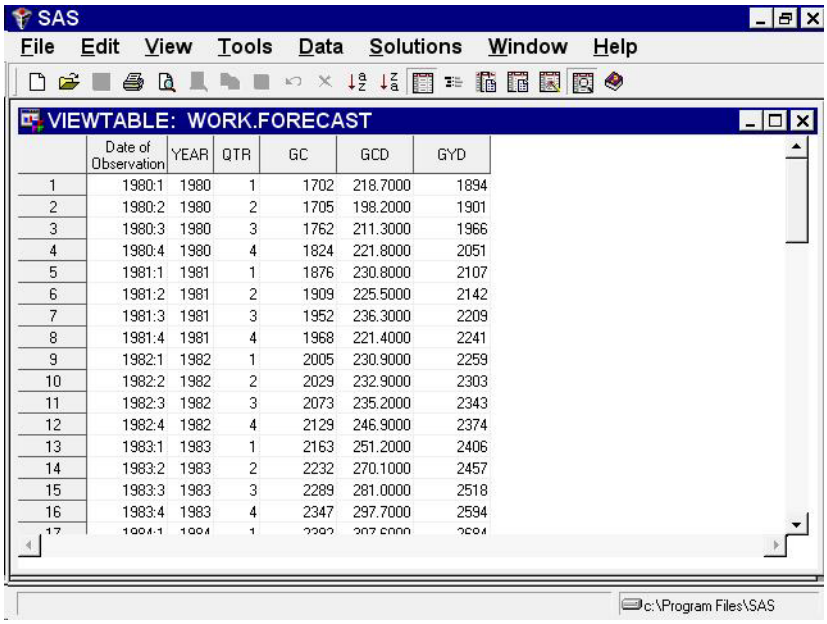
Interleaved — The data set contains time ID variables, the variable TYPE, and the forecast variables. There are several observations per time period, with the meaning of each observation identified by the TYPE variable.

Concatenated — The data set contains the variable SERIES, time ID variables, and the variables ACTUAL, PREDICT, ERROR, UPPER, LOWER, and STD. There is one observation per time period per forecast series. The variable SERIES contains the name of the forecast series, and the data set is sorted by SERIES and DATE.

Simple Format Forecast Data Set

To see the simple format forecast data set that the system created, select the Output button. This brings up a VIEWTABLE window to display the data set, as shown in Display 23.19.

Display 23.19. Forecast Data Set – Simple Format



	Date of Observation	YEAR	QTR	GC	GCD	GYD
1	1980:1	1980	1	1702	218.7000	1894
2	1980:2	1980	2	1705	198.2000	1901
3	1980:3	1980	3	1762	211.3000	1966
4	1980:4	1980	4	1824	221.8000	2051
5	1981:1	1981	1	1876	230.8000	2107
6	1981:2	1981	2	1909	225.5000	2142
7	1981:3	1981	3	1952	236.3000	2209
8	1981:4	1981	4	1968	221.4000	2241
9	1982:1	1982	1	2005	230.9000	2259
10	1982:2	1982	2	2029	232.9000	2303
11	1982:3	1982	3	2073	235.2000	2343
12	1982:4	1982	4	2129	246.9000	2374
13	1983:1	1983	1	2163	251.2000	2406
14	1983:2	1983	2	2232	270.1000	2457
15	1983:3	1983	3	2289	281.0000	2518
16	1983:4	1983	4	2347	297.7000	2594

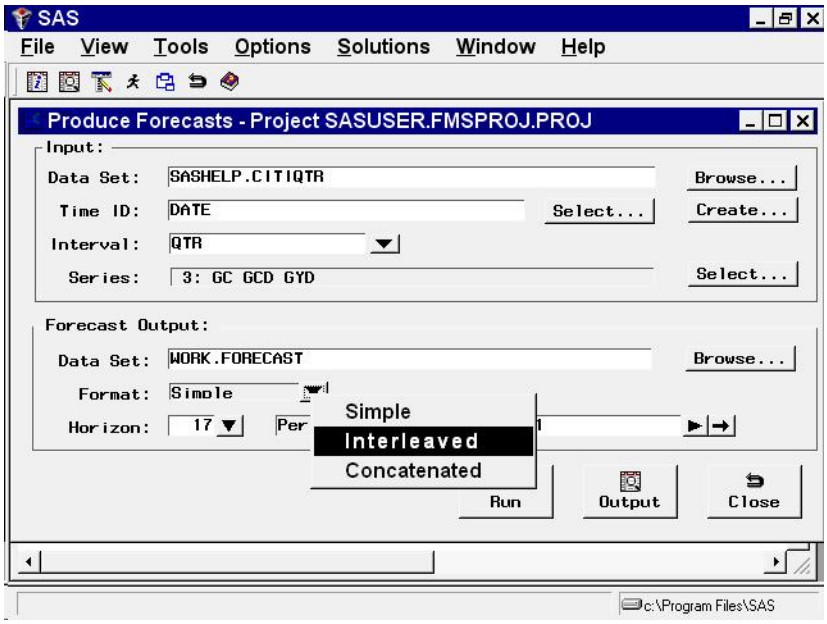
Display 23.19 shows the default simple format. This form of the forecast data set contains time ID variables and the variables that you forecast. The forecast variables contain actual values or predicted values, depending on whether the date of the observation is within the range of data supplied in the input data set.

Select `File` and `Close` to close the VIEWTABLE window.

Interleaved Format Forecast Data Set

From the Produce Forecasts window, use the combo box to select the `Interleaved` format, as shown in Display 23.20.

Display 23.20. Forecast Data Set Options

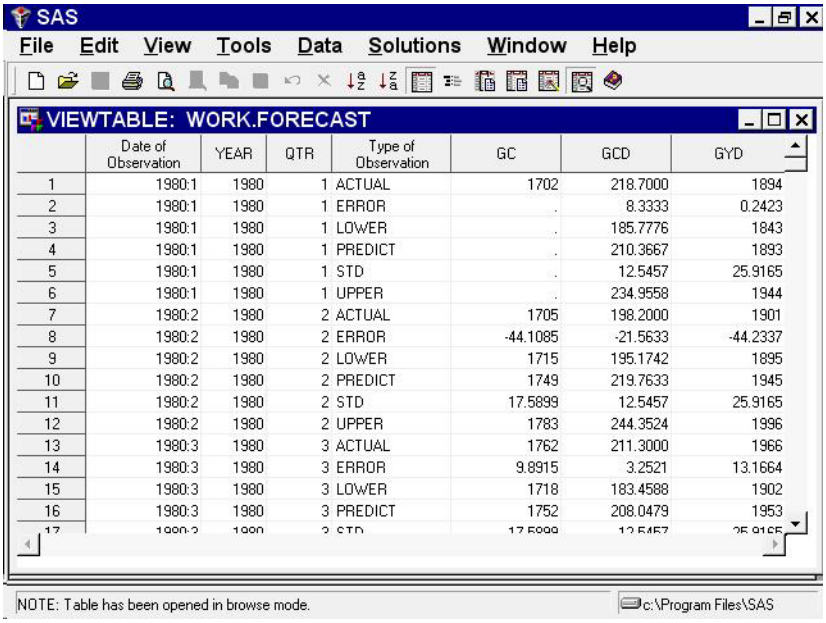


Now select the Run button again. The system presents a warning notice reminding you that the data set WORK.FORECAST already exists and asking if you want to replace it. Select Replace.

The forecasts are stored in the data set WORK.FORECAST again, this time in the *Interleaved* format. Dismiss the notice that the forecast was stored.

Now select the Output button again. This brings up a VIEWTABLE window to display the data set, as shown in Display 23.21.

Display 23.21. Forecast Data Set – Interleaved Format



	Date of Observation	YEAR	QTR	Type of Observation	GC	GCD	GYD
1	1980:1	1980	1	ACTUAL	1702	218.7000	1894
2	1980:1	1980	1	ERROR	.	8.3333	0.2423
3	1980:1	1980	1	LOWER	.	185.7776	1843
4	1980:1	1980	1	PREDICT	.	210.3667	1893
5	1980:1	1980	1	STD	.	12.5457	25.9165
6	1980:1	1980	1	UPPER	.	234.9558	1944
7	1980:2	1980	2	ACTUAL	1705	198.2000	1901
8	1980:2	1980	2	ERROR	-44.1085	-21.5633	-44.2337
9	1980:2	1980	2	LOWER	1715	195.1742	1895
10	1980:2	1980	2	PREDICT	1749	219.7633	1945
11	1980:2	1980	2	STD	17.5899	12.5457	25.9165
12	1980:2	1980	2	UPPER	1783	244.3524	1996
13	1980:3	1980	3	ACTUAL	1762	211.3000	1966
14	1980:3	1980	3	ERROR	9.8915	3.2521	13.1664
15	1980:3	1980	3	LOWER	1718	183.4588	1902
16	1980:3	1980	3	PREDICT	1752	208.0479	1953

In the interleaved format, there are several output observations for each input observation, identified by the TYPE variable. The values of the forecast variables for observations with different TYPE values are as follows.

ACTUAL — actual values copied from the input data set.

ERROR — the difference between the actual and predicted values.

LOWER — the lower confidence limits.

PREDICT — the predicted values from the forecasting model. These are within-sample, one-step-ahead predictions for observations within the historical period, or multistep predictions for observations within the forecast period.

STD — the estimated standard deviations of the prediction errors.

UPPER — the upper confidence limits.

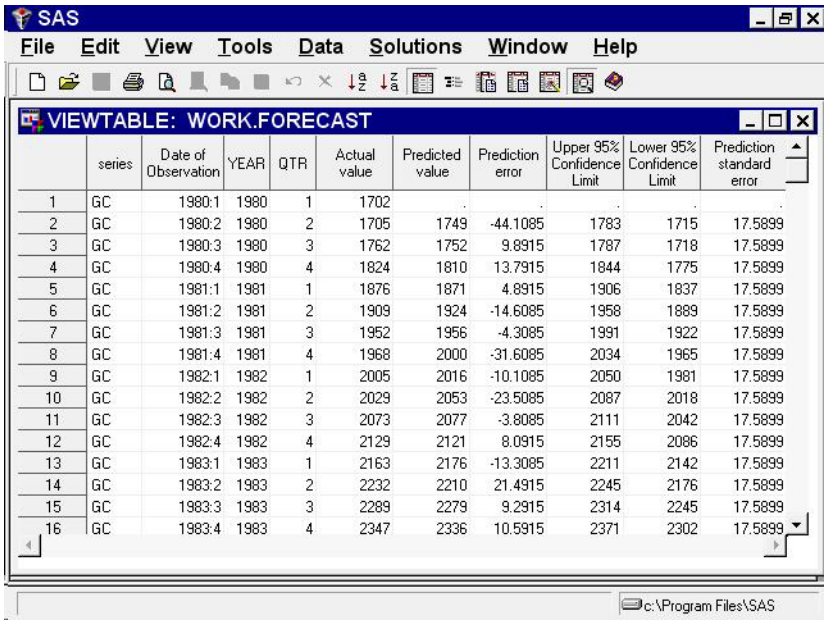
Select File and Close to close the VIEWTABLE window.

Concatenated Format Forecast Data Set

Use the combo box to select the Concatenated format. Re-create the forecast data set again, and then select the Output button.

The VIEWTABLE window showing the concatenated format of the forecast data set appears, as shown in Display 23.22.

Display 23.22. Forecast Data Set – Concatenated Format



	series	Date of Observation	YEAR	QTR	Actual value	Predicted value	Prediction error	Upper 95% Confidence Limit	Lower 95% Confidence Limit	Prediction standard error
1	GC	1980:1	1980	1	1702	.	.	.	.	.
2	GC	1980:2	1980	2	1705	1749	-44.1085	1783	1715	17.5899
3	GC	1980:3	1980	3	1762	1752	9.8915	1787	1718	17.5899
4	GC	1980:4	1980	4	1824	1810	13.7915	1844	1775	17.5899
5	GC	1981:1	1981	1	1876	1871	4.8915	1906	1837	17.5899
6	GC	1981:2	1981	2	1909	1924	-14.6085	1958	1889	17.5899
7	GC	1981:3	1981	3	1952	1956	-4.3085	1991	1922	17.5899
8	GC	1981:4	1981	4	1968	2000	-31.6085	2034	1965	17.5899
9	GC	1982:1	1982	1	2005	2016	-10.1085	2050	1981	17.5899
10	GC	1982:2	1982	2	2029	2053	-23.5085	2087	2018	17.5899
11	GC	1982:3	1982	3	2073	2077	-3.8085	2111	2042	17.5899
12	GC	1982:4	1982	4	2129	2121	8.0915	2155	2086	17.5899
13	GC	1983:1	1983	1	2163	2176	-13.3085	2211	2142	17.5899
14	GC	1983:2	1983	2	2232	2210	21.4915	2245	2176	17.5899
15	GC	1983:3	1983	3	2289	2279	9.2915	2314	2245	17.5899
16	GC	1983:4	1983	4	2347	2336	10.5915	2371	2302	17.5899

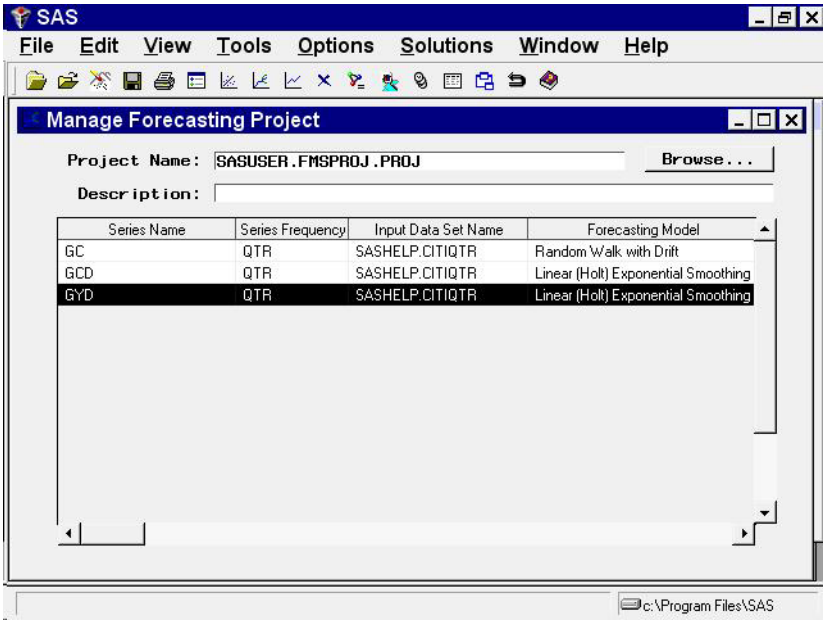
This completes the example of how to use the Produce Forecasts window. Select `File` and `Close` to close the VIEWTABLE window. Select the `Close` button to return to the `Time Series Forecasting` window.

Forecasting Projects

The system collects all the forecasting models you create, together with the options you set, into a package called a *forecasting project*. You can save this information in a SAS catalog entry and restore your work in later forecasting sessions. You can store any number of forecasting projects under different catalog entry names.

To see how this works, select the `Manage Projects` button. This brings up the `Manage Forecasting Project` window, as shown in Display 23.23.

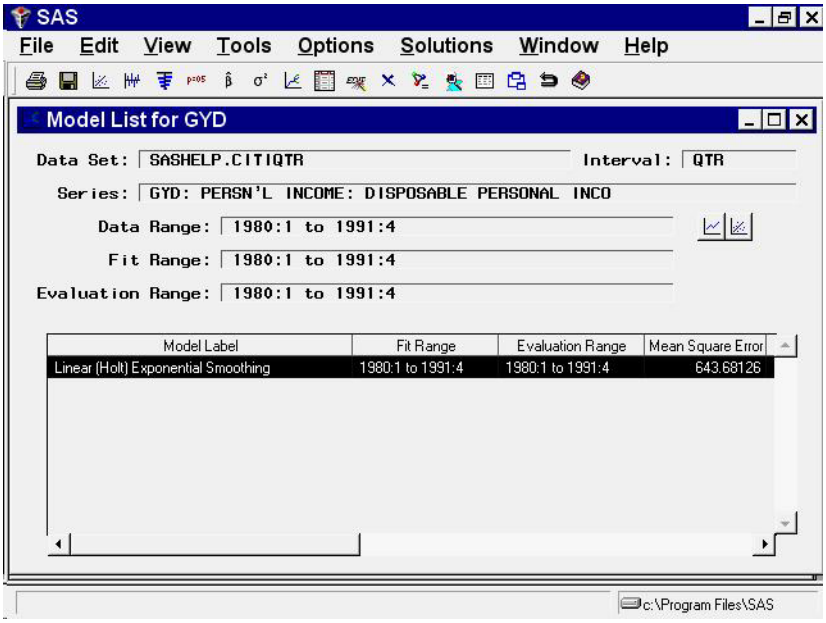
Display 23.23. Manage Forecasting Project Window



The table in this window lists the series for which forecasting models have been fit, and it shows for each series the forecasting model used to produce the forecasts. This window provides several features that allow you to manage the information in your forecasting project.

You can select a row of the table to drill down to the list of models fit to the series. Select the GYD row of the table, either by double-clicking with the mouse or by clicking once to highlight the table row and then selecting `List Models` from the toolbar or from the `Tools` pull-down menu. This brings up the `Model List` window for this series, as shown in Display 23.24.

Display 23.24. Model List Window



Because the Automatic Model Fitting Process process kept only the best fitting model, only one model appears in the model list. You can fit and retain any number of models for each series, and all the models fit and kept will appear in the series' model list.

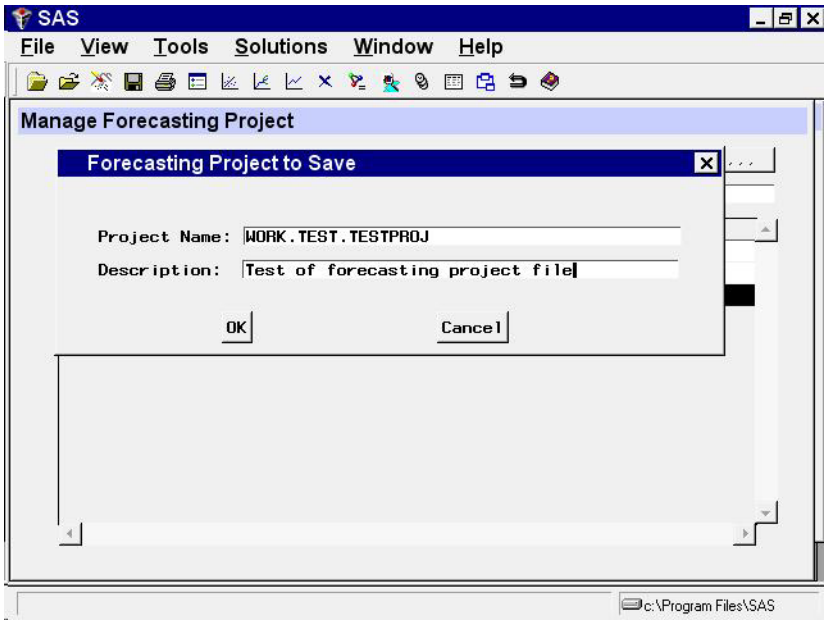
Select `Close` from the toolbar or from the File pull-down menu to return to the Manage Forecasting Project window.

Saving and Restoring Project Information

To illustrate how you can save your work between sessions, in this section you will exit and then re-enter the Forecasting System.

From the Manage Forecasting Project window, select `File` and `Save as`. This brings up the `Forecasting Project to Save` window. In the `Project Name` field type the name WORK.TEST.TESTPROJ. In the `Description` field, type "Test of forecasting project file". The window should now appear as shown in Display 23.25.

Display 23.25. Project to Save Name and Description



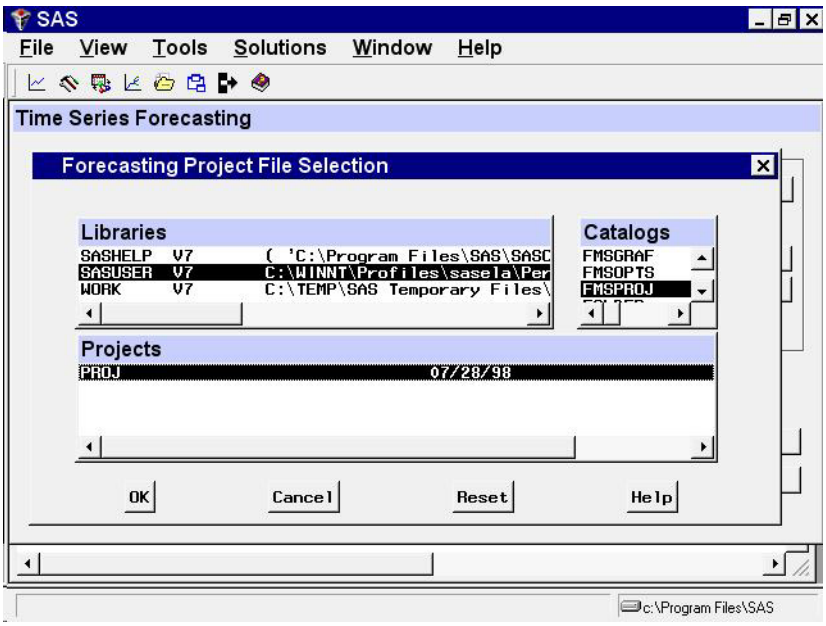
Select the `OK` button. This returns you to the Project Management window, and displays a message indicating that the project was saved.

Select `Close` from the tool-bar or from the File pull-down menu to return to the `Time Series Forecasting` window. Now select the `Exit` button. The system asks if you are sure you want to exit the system; select `Yes`. The forecasting application now terminates.

Bring up the forecasting application again. A new project name is displayed by default.

Now restore the forecasting project you saved previously. Select the Browse button to the right of the Project field. This brings up the `Forecasting Project File Selection` window, as shown in Display 23.26.

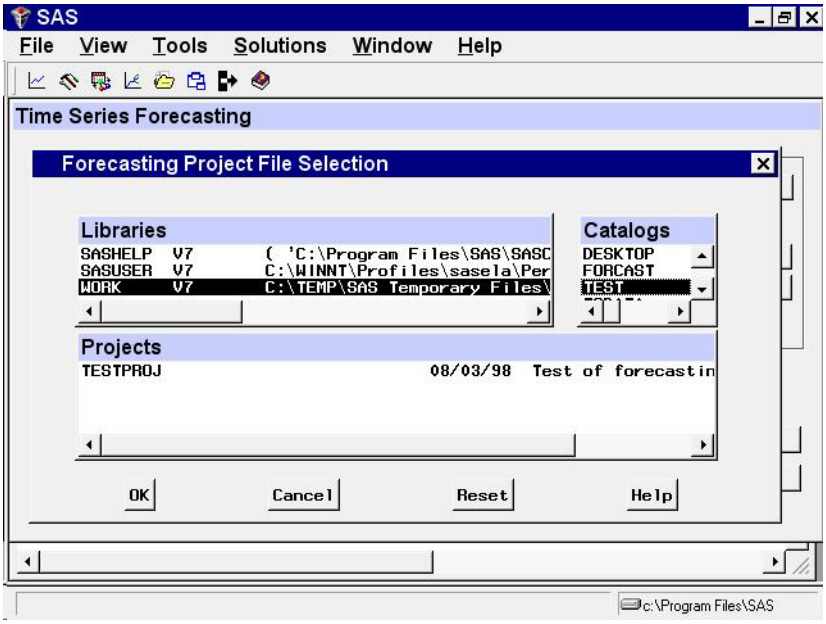
Display 23.26. Forecasting Project File Selection Window



Select the WORK library from the Libraries list. The Catalogs list now shows all the SAS catalogs in the WORK library.

Select the TEST catalog. The Projects list now shows the list of forecasting projects in the catalog TEST. So far, you have created only one project file, TESTPROJ, so TESTPROJ is the only entry in the Projects list, as shown in Display 23.27.

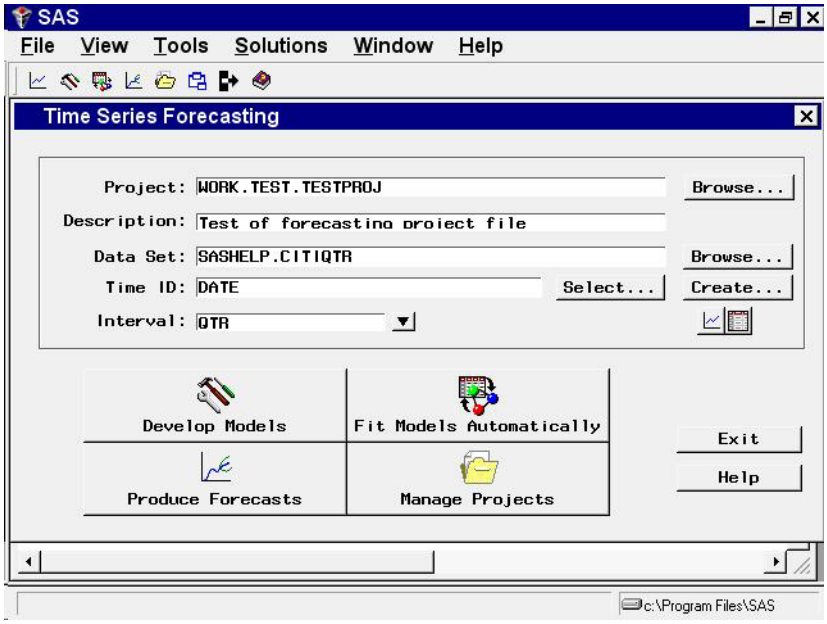
Display 23.27. Forecasting Projects List



Select TESTPROJ from the Projects list and then select the OK button. This returns you to the Time Series Forecasting window.

The system loads the project information you saved in TESTPROJ and displays a message indicating this. The Project field is now set to WORK.TEST.TESTPROJ, and the description is the description you previously gave to TESTPROJ, as shown in Display 23.28.

Display 23.28. Time Series Forecasting Window after Loading Project



If you now select the Manage Projects button, you will see the list of series and forecasting models you created in the previous forecasting session.

Develop Models Window

In the first forecasting example, you used the Automatic Model Fitting window to fit and select the forecasting model for each series automatically. In addition to this automatic forecasting process, you can also work with time series one at a time to fit forecasting models and apply your own judgement to choose the best forecasting model for each series.

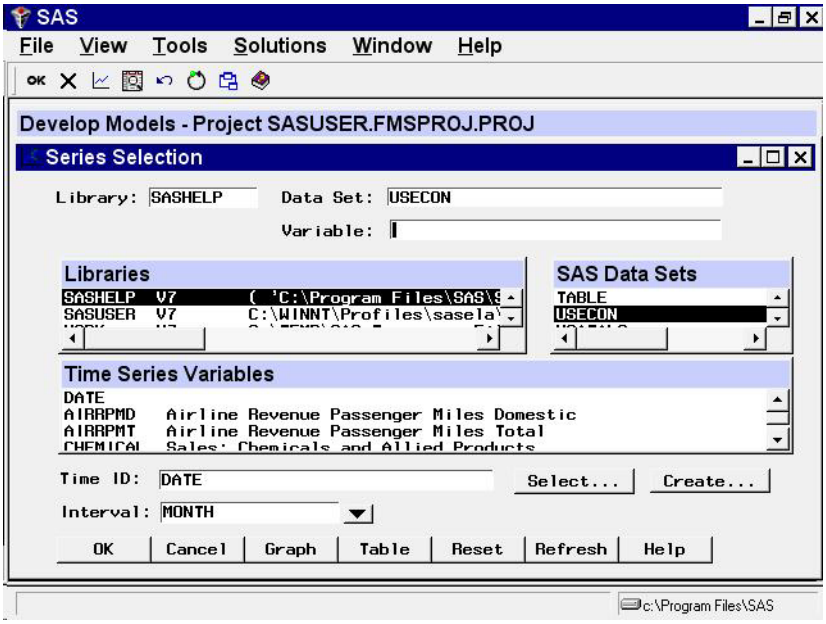
Using the Automatic Model Fitting feature, the system acts like a "black box." This section goes inside the black box to look at the kinds of forecasting methods that the system provides and introduces some of the tools the system offers to help you find the best forecasting model.

Introduction

From the Time Series Forecasting window, select the Browse button to the right of the Data Set field to bring up the Data Set Selection window. Select the USECON data set from the SASHELP library. This data set contains monthly data on the U.S. economy.

Select OK to close the selection window. Now select the `Develop Models` button. This brings up the `Series Selection` window, as shown in Display 23.29. You can enlarge this window for easier viewing of lists of data sets and series.

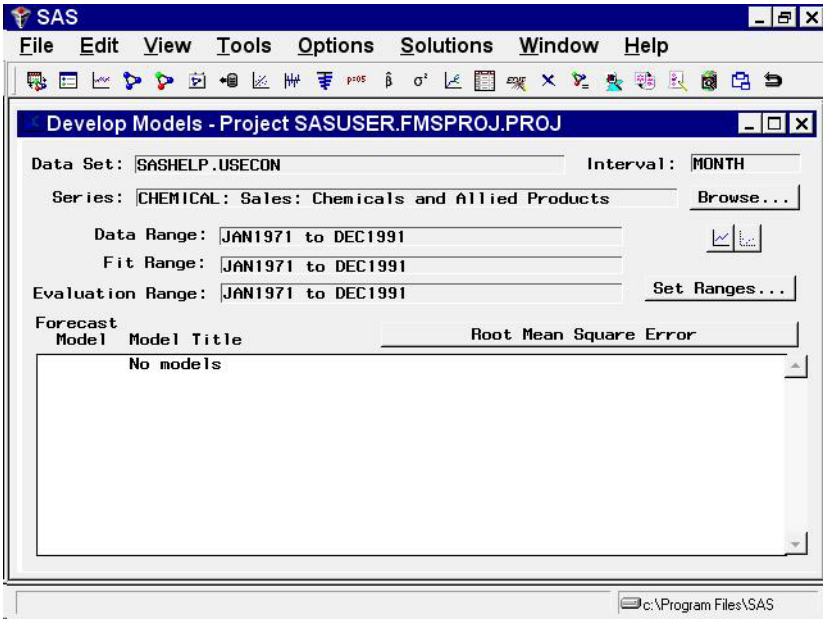
Display 23.29. Series Selection Window



Select the series CHEMICAL, *Sales of Chemicals and Allied Products*, and then select the `OK` button.

This brings up the `Develop Models` window, as shown in Display 23.30.

Display 23.30. Develop Models Window



The Data Set, Interval, and Series fields in the upper part of the Develop Models window indicate the series with which you are currently working. You can change the settings of these fields by selecting the Browse button.

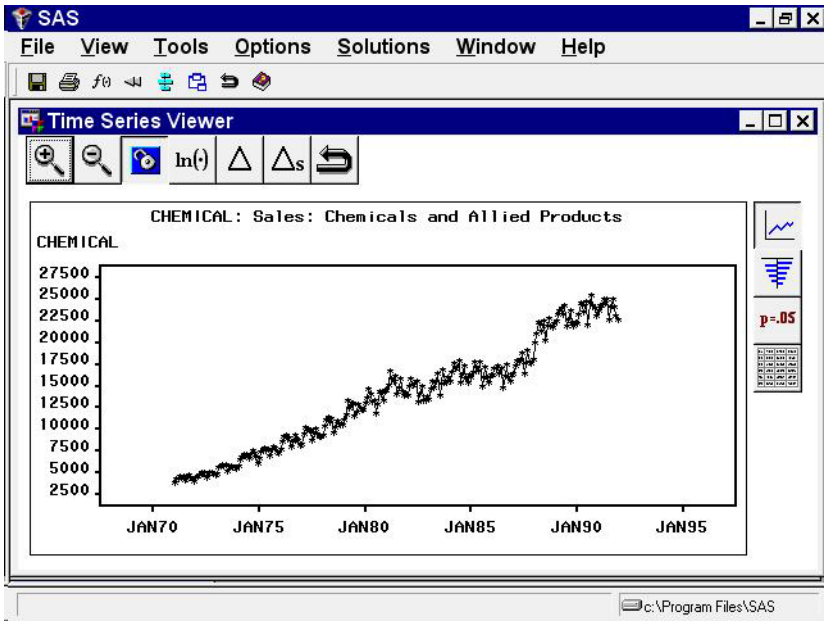
The Data Range, Fit Range, and Evaluation Range fields show the time period over which data are available for the current series, and what parts of that time period will be used to fit forecasting models to the series and to evaluate how well the models fit the data. You can change the settings of these fields by selecting the Set Ranges button.

The bottom part of the Develop Models window consists of a table of forecasting models fit to the series. Initially, the list is empty, as indicated by the message "No models." You can fit any number of forecasting models to each series and designate which one you want to use to produce forecasts.

Graphical tools are available for exploring time series and fitted models. The two icons below the Browse button access the Time Series Viewer and the Model Viewer.

Select the left icon. This brings up the Time Series Viewer window, as shown in Display 23.31.

Display 23.31. Chemical and Allied Product Series



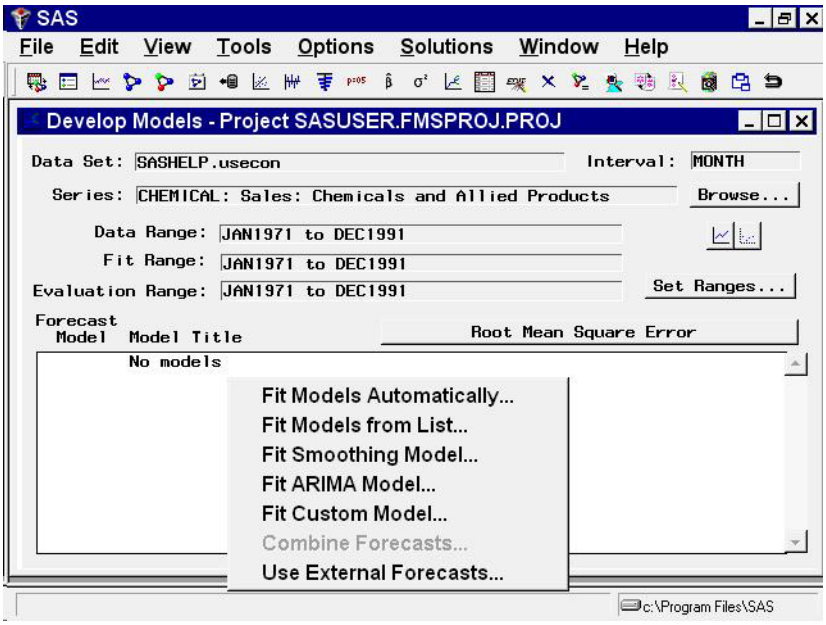
The Time Series Viewer displays a plot of the CHEMICAL series. The Time Series Viewer offers many useful features, which are explored in later sections.

The Time Series Viewer appears in a separate resizable window. You can switch back and forth between the Time Series Viewer window and other windows. For now, return to the Develop Models window. You can close the Time Series Viewer window or leave it open. (To close the Time Series Viewer window, select Close from the toolbar or from the File pull-down menu.)

Fitting Models

To bring up a menu of model fitting choices, select `Edit` from the menu bar and then select `Fit Model`, or select `Fit Models from List` in the tool-bar, or simply select a blank line in the table as shown in Display 23.32.

Display 23.32. Menu of Model Fitting Choices



The Forecasting System provides several ways to specify forecasting models. The seven choices given by the menu shown in Display 23.32 are as follows:

Fit Models Automatically
: performs for the current series the same automatic model selection process that the Automatic Model Fitting window applies to a set of series.

Fit Models from List
: presents a list of commonly used forecasting models for convenient point-and-click selection.

Fit Smoothing Model
: displays the `Smoothing Model Specification` window, which allows you to specify several kinds of exponential smoothing and Winters method forecasting models.

Fit ARIMA Model
: displays the `ARIMA Model Specification` window, which allows you to specify many kinds of autoregressive integrated moving average (ARIMA) models, including seasonal ARIMA models and ARIMA models with regressors, transfer functions, and other predictors.

Fit Custom Model
: displays the `Custom Model Specification` window, which allows you to construct a forecasting model by specify-

ing separate options for transforming the data, modeling the trend, modeling seasonality, modeling autocorrelation of the errors, and modeling the effect of regressors and other independent predictors.

Combine Forecasts
: displays the `Forecast Combination Model Specification` window, which allows you to specify models that produce forecasts by combining, or averaging, the forecasts from other models. (This option is not available unless you have fit at least two models.)

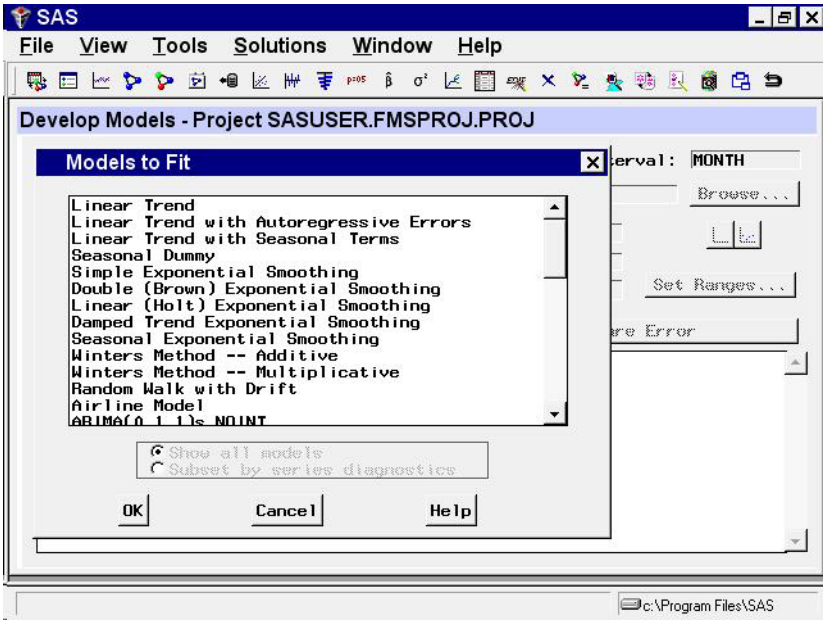
Use External Forecasts
: displays the `External Forecast Model Specification` window, which allows you to use judgmental or externally produced forecasts that have been saved in a separate series in the data set.

All the forecasting models used by the system are ultimately specified through one of the three windows: `Smoothing Method Specification`, `ARIMA Model Specification`, or `Custom Model Specification`. You can specify the same models with either the ARIMA Model Specification window or the Custom Model Specification window, but the Custom Model Specification window may provide a more natural way to specify models for those who are less familiar with the Box-Jenkins style of time series model specification.

The Automatic Model feature, the `Models to Fit` window, and the `Forecast Combination Model Specification` window all deal with lists of forecasting models previously defined through the Smoothing Model, ARIMA Model, or Custom Model specification windows. These windows are discussed in detail in later sections.

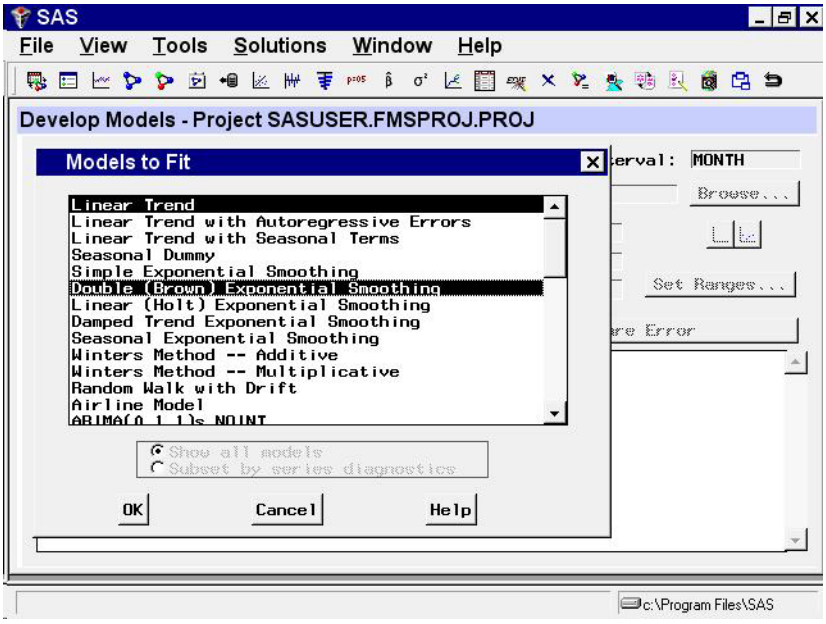
To get started using the Develop Models window, select the `Fit Models from List` item from the menu shown in Display 23.32. This brings up the `Models to Fit` window, as shown in Display 23.33.

Display 23.33. Models to Fit Window



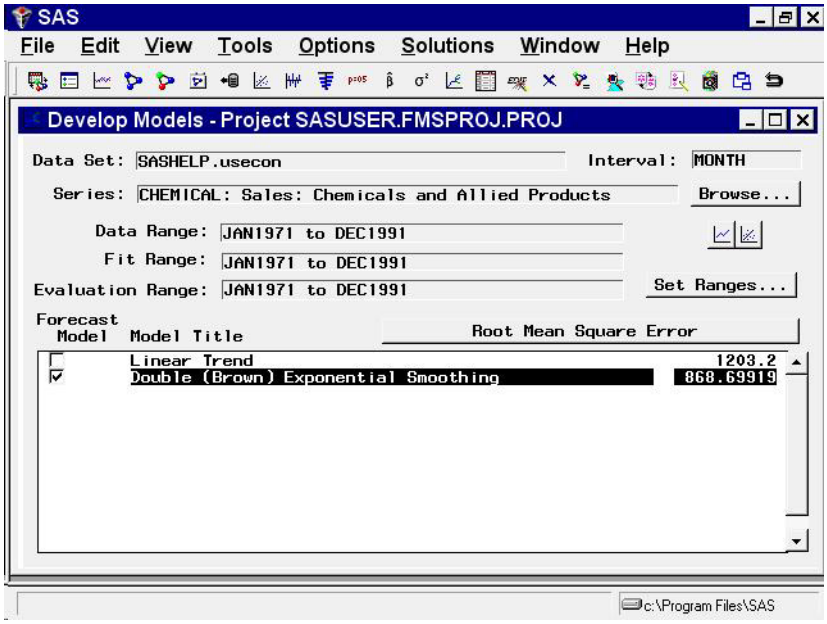
You can select several models to fit at once by holding down the CTRL key as you make the selections. Select `Linear Trend` and `Double (Brown) Exponential Smoothing`, as shown in Display 23.34, and then select the `OK` button.

Display 23.34. Selecting Models to Fit



The system fits the two models you selected. After the models are fit, the labels of the two models and their goodness-of-fit statistic are added to the model table, as shown in Display 23.35.

Display 23.35. Fitted Models List



Model List and Statistics of Fit

In the model list, the *Model Title* column shows the descriptive labels for the two fitted models, in this case Linear Trend and Double Exponential Smoothing.

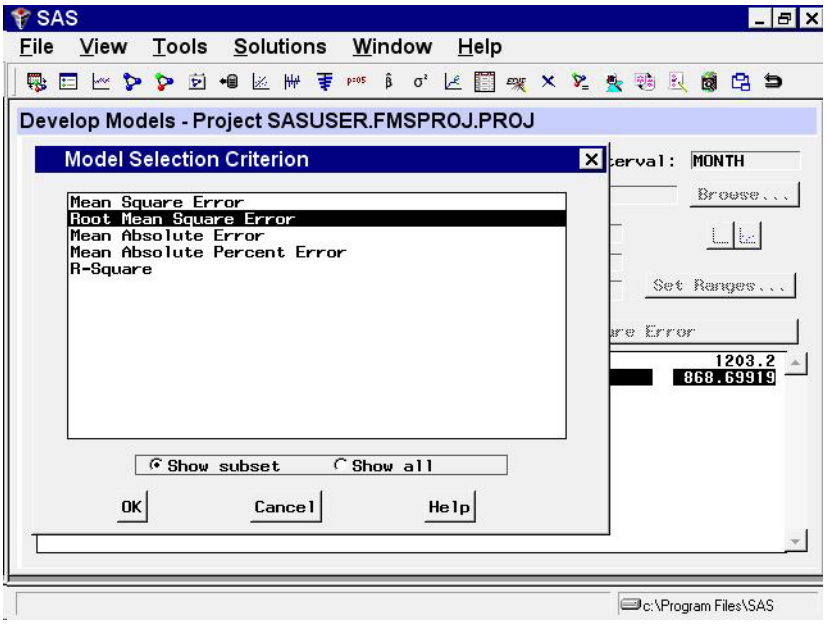
The column labeled *Root Mean Square Error* (or labeled *Mean Absolute Percent Error* if you continued from the example in the previous section) shows the goodness-of-fit criterion used to decide which model fits better. By default, the criterion used is the root mean square error, but you can choose a different measure of fit. The linear trend model has a root mean square error of 1203, while the double exponential smoothing model fits better, with a RMSE of only 869.

The left column labeled *Forecast Model* consists of check boxes that indicate which one of the models in the list has been selected as the model to use to produce the forecasts for the series. When new models are fit and added to the model list, the system sets the Forecast Model flags to designate the one model with the best fit–as measured by the selected goodness-of-fit statistic–as the forecast model. (In the case of ties, the first model with the best fit is selected.)

Because the Double Exponential Smoothing model has the smaller RMSE of the two models in the list, its Forecast Model check box is set. If you would rather produce forecasts using the Linear Trend model, choose it by selecting the corresponding check box in the Forecast Model column.

To use a different goodness-of-fit criterion, select the button with the current criterion name on it (Root Mean Square Error or Mean Absolute Percent Error). This brings up the `Model Selection Criterion` window, as shown in Display 23.36.

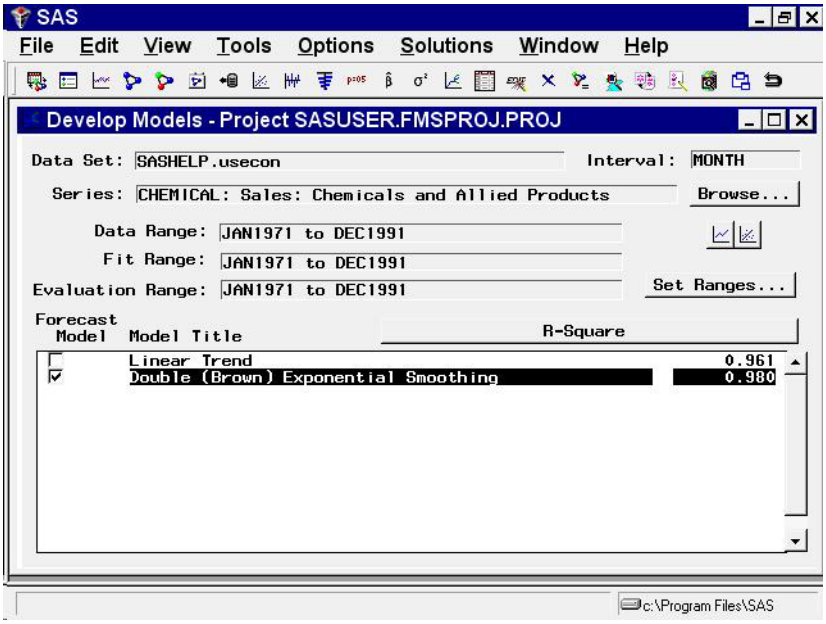
Display 23.36. Model Selection Criterion Window



The system provides many measures of fit that you can use as the model selection criterion. To avoid confusion, only the most popular of the available fit statistics are shown in this window by default. To display the complete list, you can select the `Show all` option. You can control the subset of statistics listed in this window through the `Statistics of Fit` item in the `Options` menu on the Develop Models window.

Initially, Root Mean Square Error is selected. Select R-Square and then select the `OK` button. This changes the fit statistic displayed in the model list, as shown in Display 23.37.

Display 23.37. Model List with R-Square Statistics

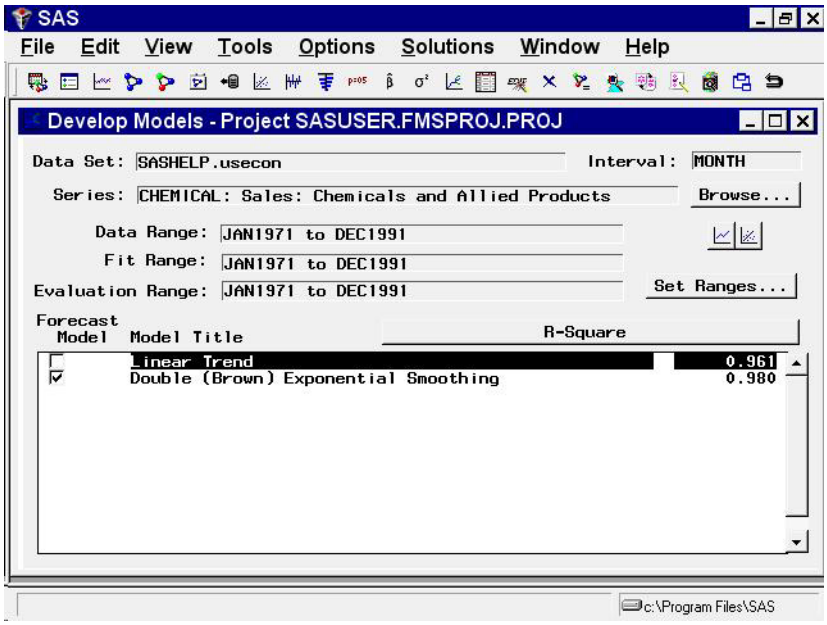


Now that you have fit some models to the series, you can use the Model Viewer button to take a closer look at the predictions of these models.

Model Viewer

In the Develop Models window, select the row in the table containing the Linear Trend model so that this model is highlighted. The model list should now appear as shown in Display 23.38.

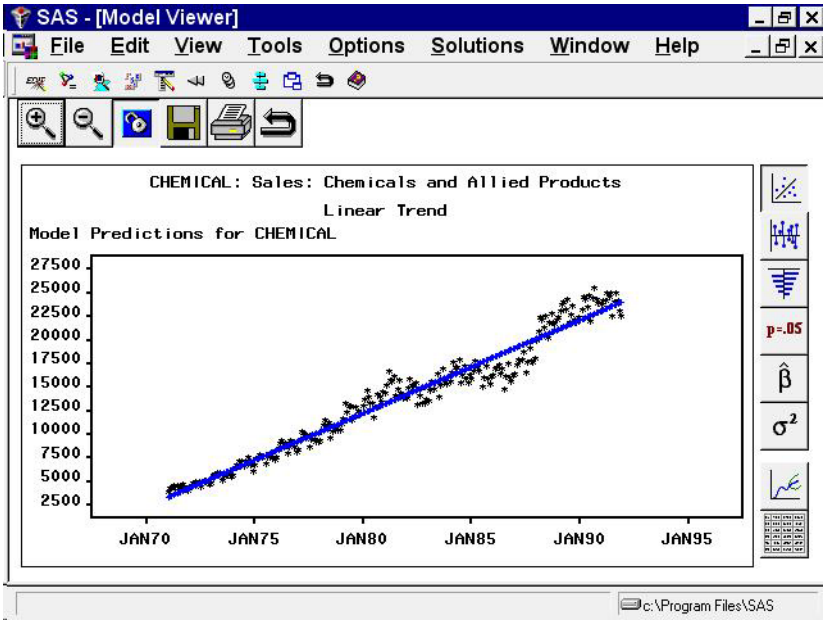
Display 23.38. Selecting a Model to View



Note that the Linear Trend model is now highlighted, but the Forecast Model column still shows the Double Exponential Smoothing model as the model chosen to produce the final forecasts for the series. Selecting a model in the list means that this is the model that menu items such as `View Model`, `Delete`, `Edit`, and `Refit` will act upon. Choosing a model by selecting its check box in the Forecast Model column means that this model will be used by the `Produce Forecasts` process to generate forecasts.

Now bring up the `Model Viewer` by selecting the right icon under the Browse button, or by selecting `Model Predictions` in the tool-bar or from the View pull-down menu. The `Model Viewer` displays the Linear Trend model, as shown in Display 23.39.

Display 23.39. Model Viewer: Actual and Predicted Values Plot

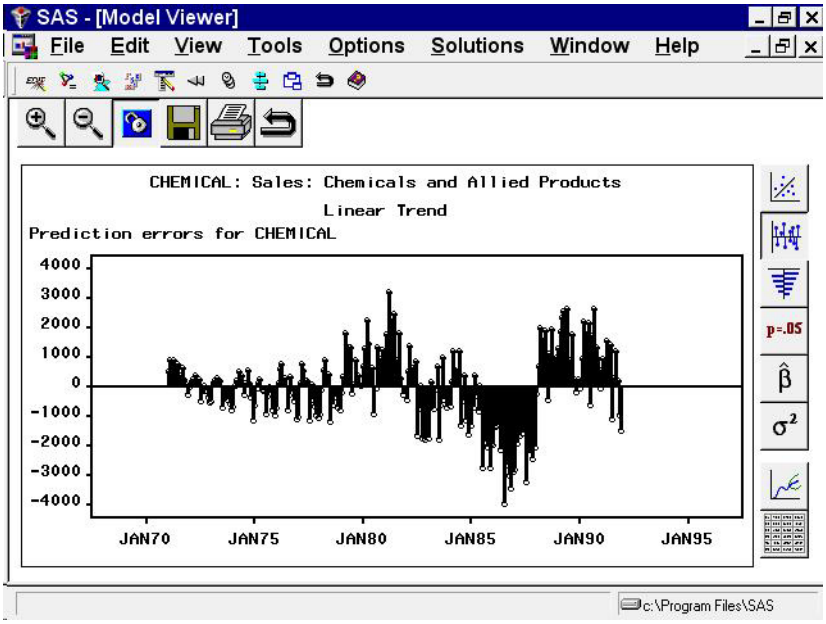


This graph shows the linear trend line representing the model predicted values together with a plot of the actual data values, which fluctuate about the trend line.

Prediction Error Plots

Select the second icon from the top in the vertical tool-bar in the Model Viewer window. This switches the Viewer to display a plot of the model prediction errors (actual data values minus the predicted values), as shown in Display 23.40.

Display 23.40. Model Viewer: Prediction Errors Plot



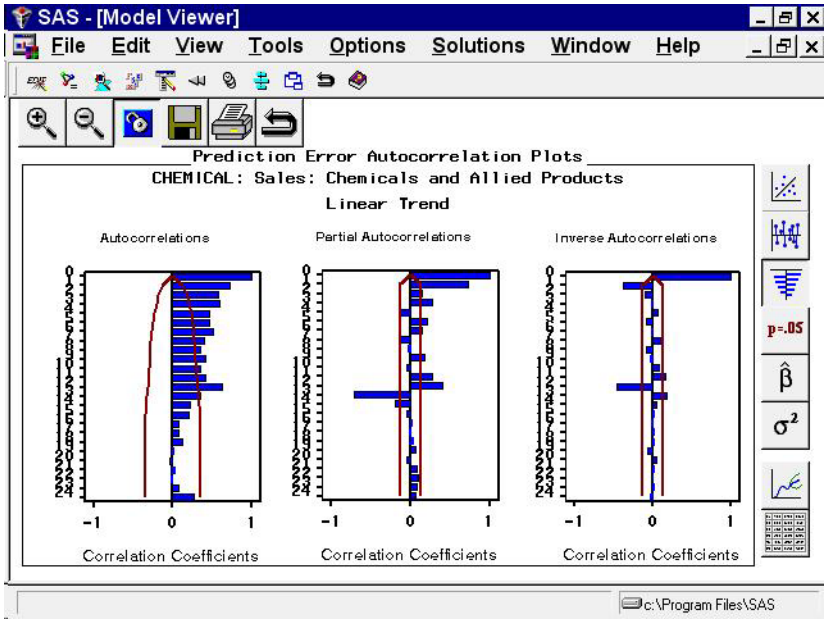
If the model being viewed includes a transformation, prediction errors are defined as the difference between the transformed series actual values and model predictions.

You can choose to graph instead the difference betwen the untransformed series values and untransformed model predictions, which are called *model residuals*. You can also graph normalized prediction errors or normalized model residuals. Use the `Residual Plot Options` submenu under the `Options` pull-down menu.

Autocorrelation Plots

Select the third icon from the top in the vertical tool-bar. This switches the Viewer to display a plot of autocorrelations of the model prediction errors at different lags, as shown in Display 23.41. Autocorrelations, partial autocorrelations, and inverse autocorrelations are displayed, with lines overlaid at plus and minus two standard errors. You can switch the graphs so that the bars represent significance probabilities by selecting the `Correlation Probabilities` item on the tool-bar or from the `View` pull-down menu. For more information on the meaning and use of autocorrelation plots, refer to Chapter 7, "The ARIMA Procedure."

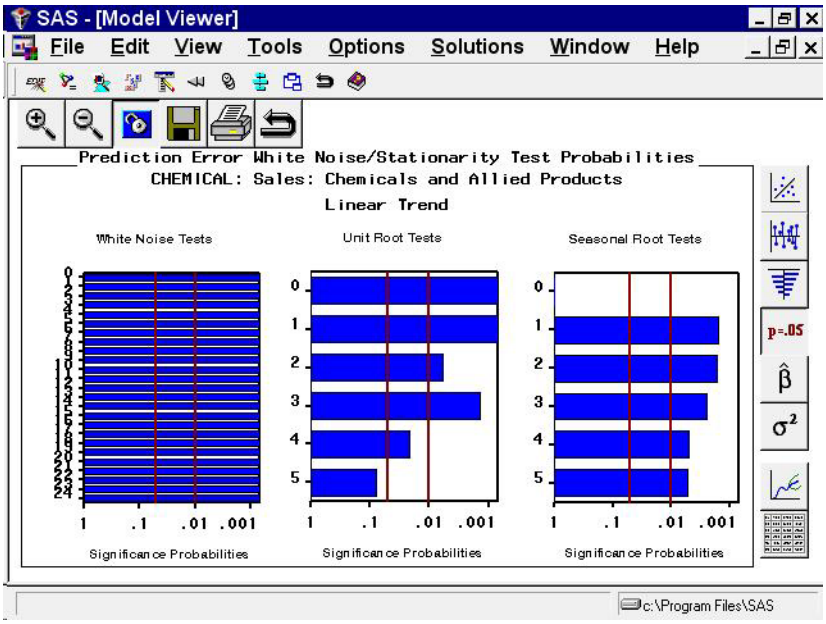
Display 23.41. Model Viewer: Autocorrelations Plot



White Noise and Stationarity Plots

Select the fourth icon from the top in the vertical tool-bar. This switches the Viewer to display a plot of white noise and stationarity tests on the model prediction errors, as shown in Display 23.42.

Display 23.42. Model Viewer: White Noise and Stationarity Plot

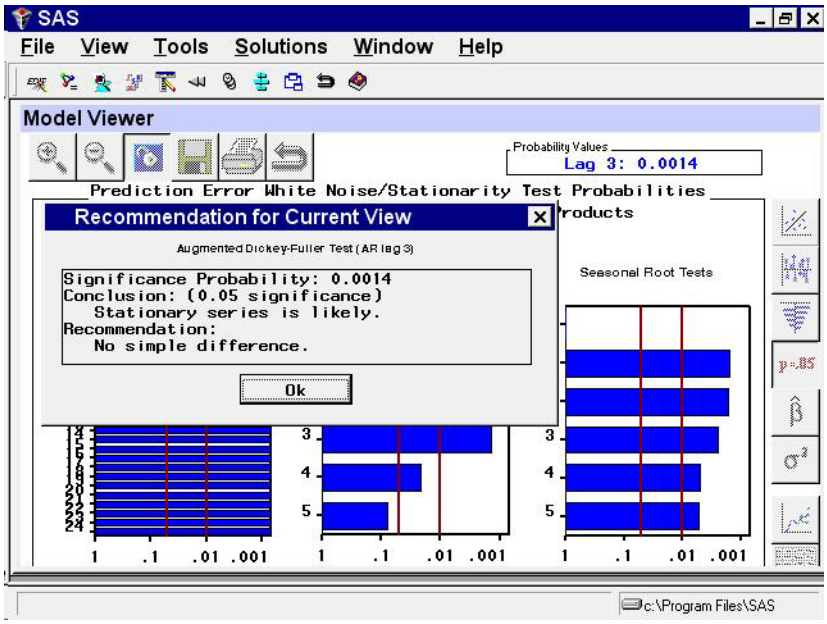


The white noise test bar chart shows significance probabilities of the Ljung-Box Chi Square statistic. Each bar shows the probability computed on autocorrelations up to the given lag. Longer bars favor rejection of the null hypothesis that the prediction errors represent white noise. In this example they are all significant beyond the .001 probability level, so that we reject the null hypothesis. In other words, the high level of significance at all lags makes it clear that the linear trend model is inadequate for this series.

The second bar chart shows significance probabilities of the Augmented Dickey-Fuller test for unit roots. For example, the bar at lag three indicates a probability of .0014, so that we reject the null hypothesis that the series is nonstationary. The third bar chart is similar to the second except that it represents the seasonal lags. Since this series has a yearly seasonal cycle, the bars represent yearly intervals.

You can select any of the bars to display an interpretation. Select the fourth bar of the middle chart. This displays the `Recommendation for Current View`, as shown in Display 23.43. This window gives an interpretation of the test represented by the bar that was selected; it is significant, therefore a stationary series is likely. It also gives a recommendation: You do not need to perform a simple difference to make the series stationary.

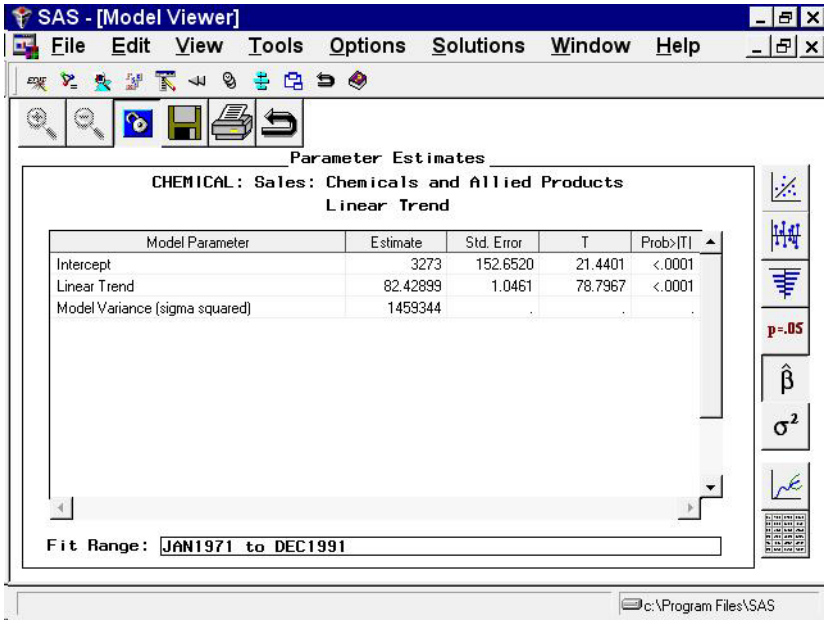
Display 23.43. Model Viewer: Recommendation for Current View



Parameter Estimates Table

Select the fifth icon from the top in the vertical tool-bar to the right of the graph. This switches the Viewer to display a table of parameter estimates for the fitted model, as shown in Display 23.44.

Display 23.44. Model Viewer: Parameter Estimates Table

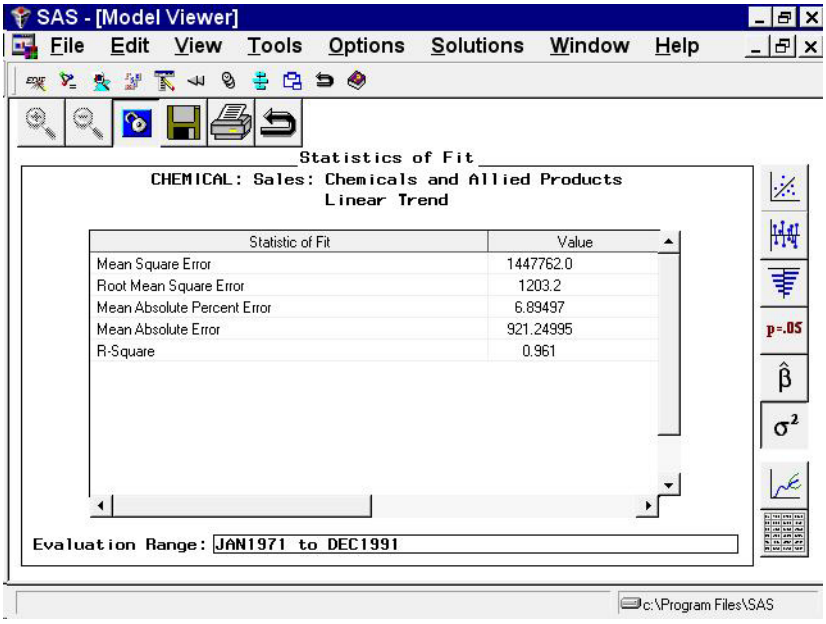


For the linear trend model, the parameters are the intercept and slope coefficients. The table shows the values of the fitted coefficients together with standard errors and *t*-tests for the statistical significance of the estimates. The model residual variance is also shown.

Statistics of Fit Table

Select the sixth icon from the top in the vertical tool-bar to the right of the table. This switches the Viewer to display a table of statistics of fit computed from the model prediction errors, as shown in Display 23.45. The list of statistics displayed is controlled by selecting `Statistics of Fit` from the `Options` pull-down menu.

Display 23.45. Model Viewer: Statistics of Fit Table



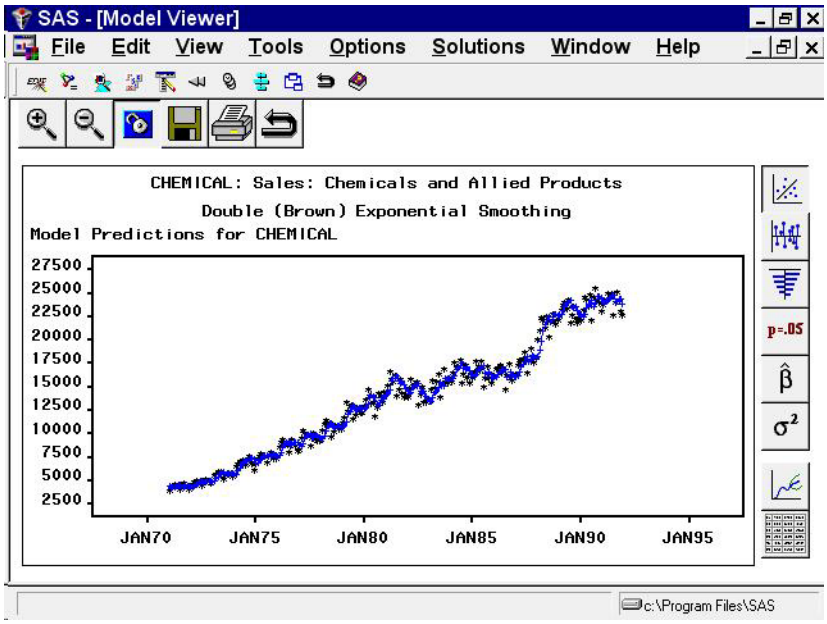
Changing to a Different Model

Select the first icon in the vertical tool-bar to the right of the table to return the display to the predicted and actual values plots (Display 23.39).

Now return to the Develop Models window, but do not close the Model Viewer window. You can use the `Next Viewer` icon in the tool-bar or your system's window manager controls to switch windows. You can resize the windows to make them both visible.

Select the Double Exponential Smoothing model so that this line of the model list is highlighted. The Model Viewer window is now updated to display a plot of the predicted values for the Double Exponential Smoothing model, as shown in Display 23.46. The Model Viewer is automatically updated to display the currently selected model, unless you specify `Unlink` (the third icon in the window's horizontal tool-bar).

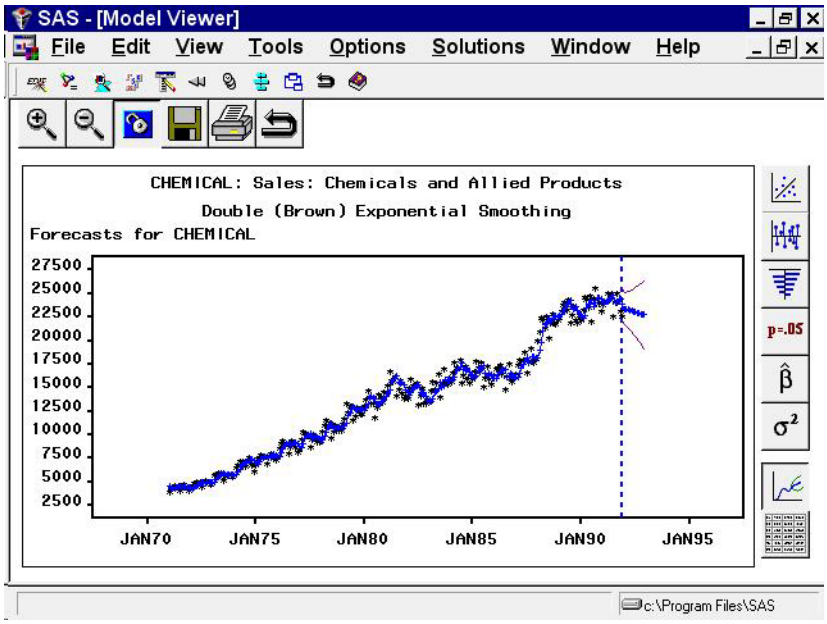
Display 23.46. Model Viewer Plot for Exponential Smoothing Model



Forecasts and Confidence Limits Plots

Select the seventh icon from the top in the vertical tool-bar to the right of the graph. This switches the Viewer to display a plot of forecast values and confidence limits, together with actual values and one-step-ahead within-sample predictions, as shown in Display 23.47.

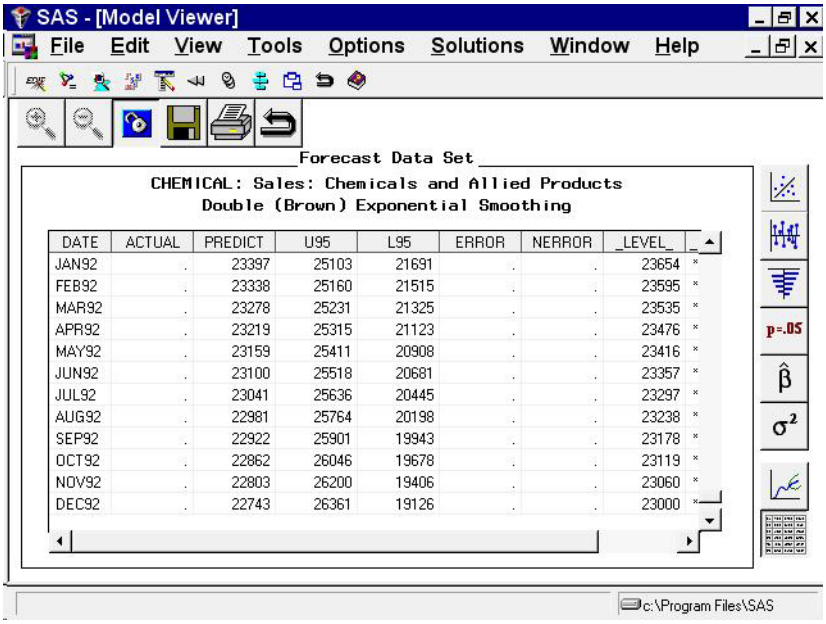
Display 23.47. Model Viewer: Forecasts and Confidence Limits



Data Table

Select the last icon at the bottom of the vertical tool-bar to the right of the graph. This switches the Viewer to display the forecast data set as a table, as shown in Display 23.48.

Display 23.48. Model Viewer: Forecast Data Table



To view the full data set, use the vertical and horizontal scroll bars on the data table or enlarge the window.

Closing the Model Viewer

Other features of the Model Viewer and Develop Models window are discussed later in this book. For now, close the Model Viewer window and return to the `Time Series Forecasting` window.

To close the Model Viewer window, select `Close` from the window's horizontal tool-bar or from the `File` pull-down menu.