Chapter 26
Choosing the Best Forecasting Model

Chapter Table of Contents

1308

# Chapter 26
Choosing the Best Forecasting Model

The Time Series Forecasting System provides a variety of tools for identifying potential forecasting models and for choosing the best fitting model. It allows you to decide how much control you want to have over the process, from a hands-on approach to one that is completely automated. This chapter begins with an exploration of the tools available through the Series Viewer and Model Viewer. It presents an example of identifying models graphically and exercising your knowledge of model properties. The remainder of the chapter shows you how to compare models by using a variety of statistics and by controlling the fit and evaluation time ranges. It concludes by showing you how to refit existing models and how to compare models using hold-out samples.

Time Series Viewer Features

The `Time Series Viewer` is a graphical tool for viewing and analyzing time series. It can be used separately from the `Time Series Forecasting System` using the TSVIEW command or by selecting `Time Series Viewer` from the `Analysis` pull-down menu under `Solutions`.

In this chapter you will use the Time Series Viewer to examine plots of your series before fitting models. Begin this example by invoking the forecasting system and selecting the `View Series Graphically` button, as shown in Display 26.1, or the `View Series` toolbar icon.
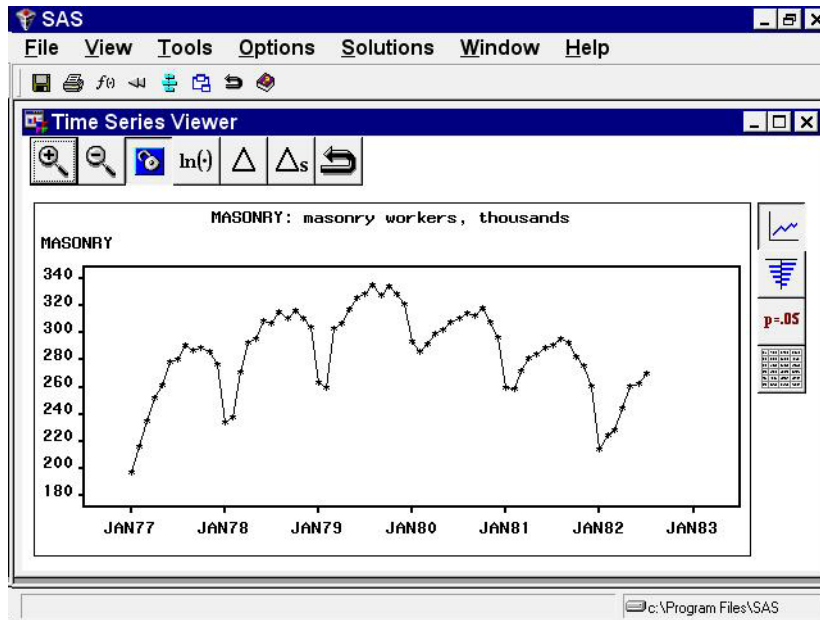
**Display 26.1.** Invoking the Time Series Viewer
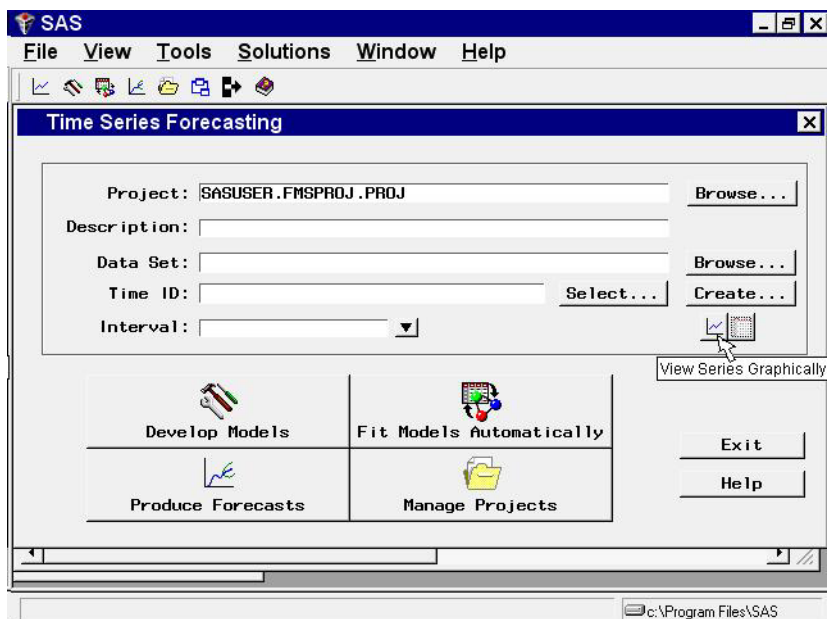
From the Series Selection window, select SASHELP as the library, WORKERS as the data set, and MASONRY as the time series, and then click the `Graph` button. The Time Series Viewer displays a plot of the series, as shown in Display 26.2.

Display 26.2. Series Plot

Select the Zoom In icon, the first one on the window's horizontal toolbar. Notice that the mouse cursor changes shape and that "Note: Click on a corner of the region, then drag to the other corner" appears on the message line. Outline an area, as shown in Display 26.3, by clicking the mouse at the upper-left corner, holding the button down, dragging to the lower right corner, and releasing the button.

Display 26.3. Selecting an Area for Zoom

1310
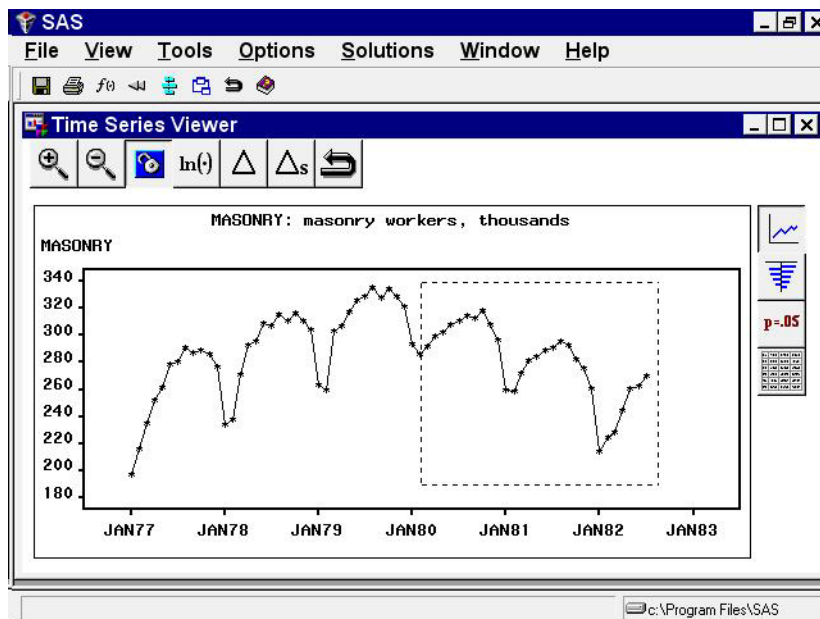
The zoomed plot should appear as shown in Display 26.4.

Display 26.4. Zoomed Plot
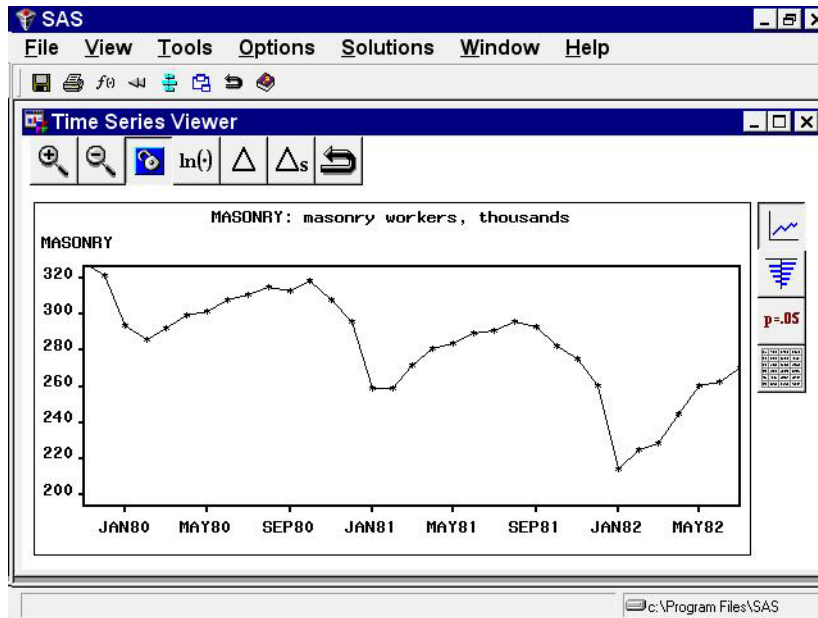
You can repeat the process to zoom in still further. To return to the previous view, select the Zoom Out icon, the second icon on the window's horizontal toolbar.

The third icon on the horizontal toolbar is used to link or unlink the viewer window. By default, the viewer is linked, meaning that it is automatically updated to reflect selection of a different time series. To see this, return to the Series Selection window by clicking on it or using the Window pull-down menu or `Next Viewer` toolbar icon. Select the Electric series in the `Time Series Variables` list box. Notice that the Time Series Viewer window is updated to show a plot of the ELECTRIC series. Select the `Link/unlink` icon if you prefer to unlink the viewer so that it is not automatically updated in this way. Successive selections toggle between the linked and unlinked state. A note on the message line informs you of the state of the Time Series Viewer window.

When a Time Series Viewer window is linked, selecting `View Series` again will make the linked Viewer window active. When no Time Series Viewer window is linked, selecting `View Series` brings up an additional Time Series Viewer window. You can bring up as many Time Series Viewer windows as you want.

Having seen the plot in Display 26.2, you might suspect that the series is nonstationary and seasonal. You can gain further insight into this by examining the sample autocorrelation function (ACF), partial autocorrelation function (PACF), and inverse autocorrelation function (IACF) plots. To switch the display to the autocorrelation plots, select the second icon from the top on the vertical toolbar at the right side of the Time Series Viewer. The plot appears as shown in Display 26.5.

Display 26.5. Sample Autocorrelation Plots
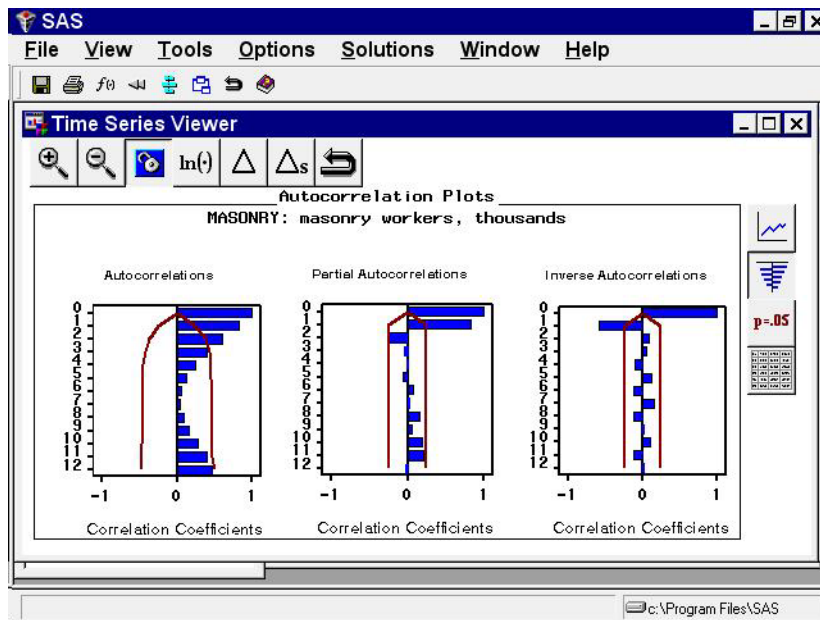
Each bar represents the value of the correlation coefficient at the given lag. The overlaid lines represent confidence limits computed at plus and minus two standard errors. You can switch the graphs to show significance probabilities by selecting `Correlation Probabilities` under the `Options` pull-down menu, or by selecting the `Toggle ACF Probabilities` toolbar icon.

The slow decline of the ACF suggests that first differencing may be warranted. To see the effect of first differencing, select the simple difference icon, the fifth icon from the left on the window's horizontal toolbar. The plot now appears as shown in Display 26.6.

Display 26.6. ACF Plots with First Difference Applied

1312
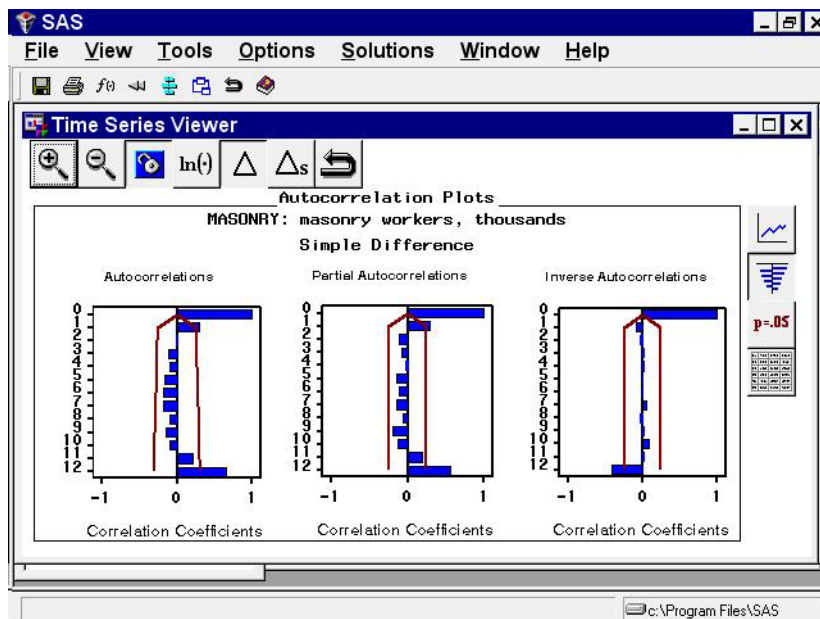
Since the ACF still displays slow decline at seasonal lags, seasonal differencing is appropriate (in addition to the first differencing already applied). Select the Seasonal Difference icon, the sixth icon from the left on the horizontal toolbar. The plot now appears as shown in Display 26.7.

Display 26.7. ACF Plot with Simple and Seasonal Differencing
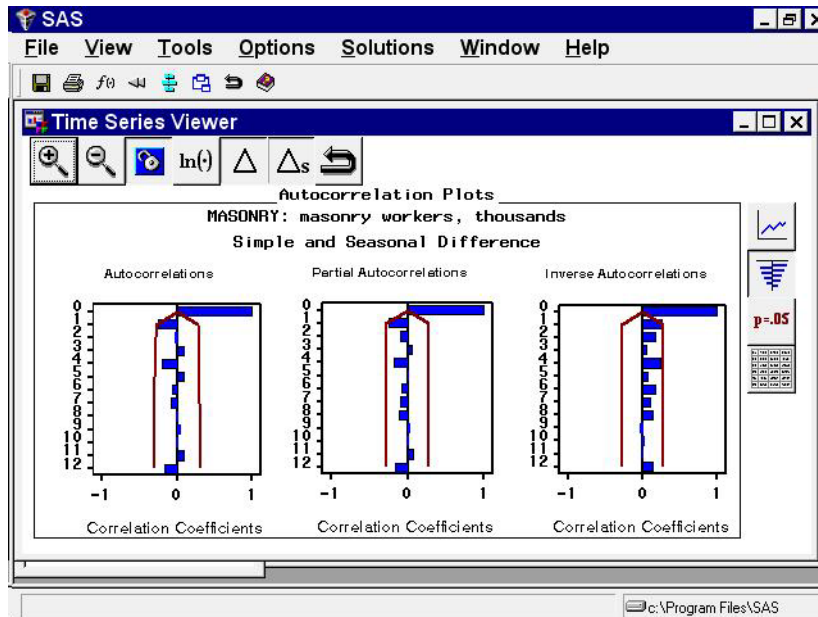

Model Viewer Prediction Error Analysis

Leave the Time Series Viewer open for the remainder of this exercise. Drag it out of the way or push it to the background so that you can return to the Time Series Forecasting window. Select Develop Models, then click an empty part of the table to bring up the pop-up menu, and select Fit ARIMA Model. Define the ARIMA(0,1,0)(0,1,0)s model by selecting 1 for differencing under ARIMA Options, 1 for differencing under Seasonal ARIMA Options, and No for Intercept, as shown in Display 26.8.

Display 26.8.   Specifying the ARIMA(0,1,0)(0,1,0)s Model
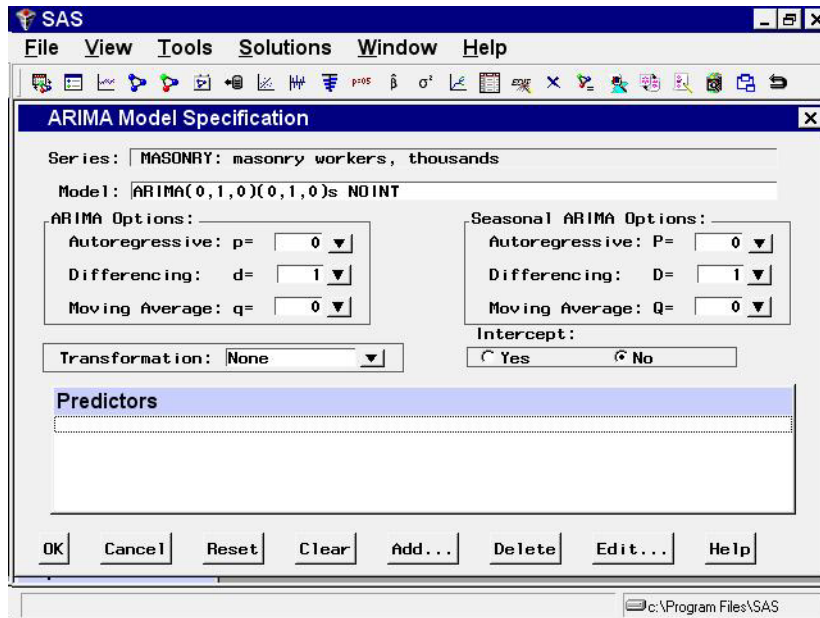
When you select the OK button, the model is fit and you are returned to the Develop Models window. Click on an empty part of the table and choose Fit Models from List from the pop-up menu. Select Airline Model from the window. (Airline Model is a common name for the ARIMA(0,1,1)(0,1,1)s model, which is often used for seasonal data with a linear trend.) Select the OK button. Once the model has been fit, the table shows the two models and their root mean square errors. Notice that the Airline Model provides only a slight improvement over the differencing model, ARIMA(0,1,0)(0,1,0)s. Select the first row to highlight the differencing model, as shown in Display 26.9.

**Display 26.9.**   Selecting a Model
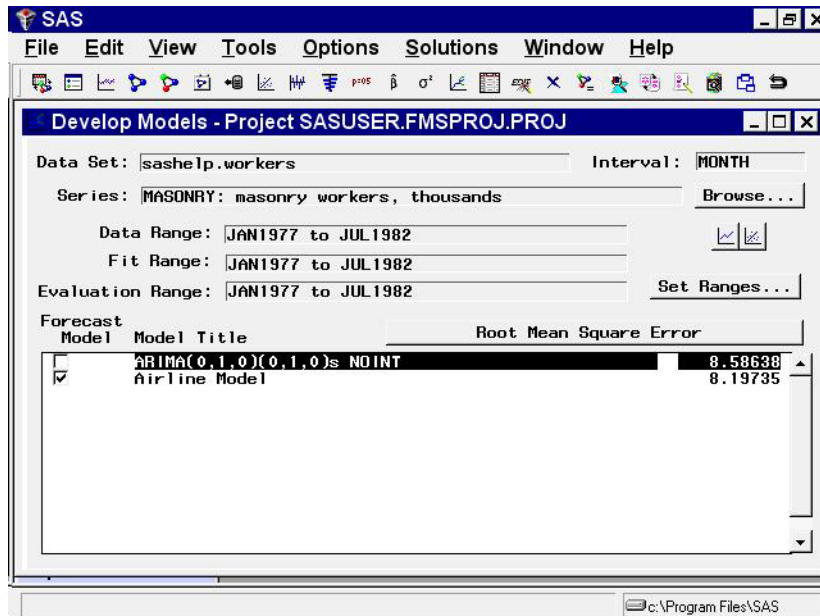
Now select the `View Selected Model Graphically` button, below the `Browse` button at the right side of the Develop Models window. The `Model Viewer` window appears, showing the actual data and model predictions for the MASONRY series. (Note that predicted values are missing for the first 13 observations due to simple and seasonal differencing.)

To examine the ACF plot for the model prediction errors, select the third icon from the top on the vertical toolbar. For this model, the prediction error ACF is the same as the ACF of the original data with first differencing and seasonal differencing applied. This differencing is apparent if you bring the Time Series Viewer back into view for comparison.

Return to the Develop Models Window by clicking on it or using the window pull-down menu or the Next Viewer toolbar icon. Select the second row of the table in the Develop Models window to highlight the Airline Model. The Model Viewer is automatically updated to show the prediction error ACF of the newly selected model, as shown in Display 26.10.

Display 26.10. Prediction Error ACF Plot for the Airline Model
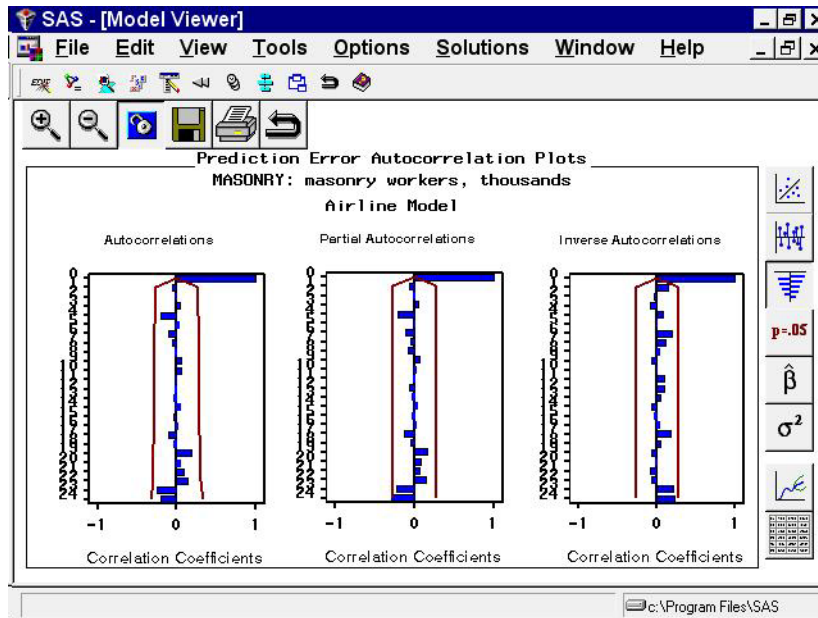
Another helpful tool available within the Model Viewer is the parameter estimates table. Select the fifth icon from the top of the vertical toolbar. The table gives the parameter estimates for the two moving average terms in the Airline Model, as well as the model residual variance, as shown in Display 26.11.

**Display 26.11.** Parameter Estimates for the Airline Model

You can adjust the column widths in the table by dragging the vertical borders of the column titles with the mouse. Notice that neither of the parameter estimates is significantly different from zero at the .05 level of significance, since `Prob>|t|` is greater than .05. This suggests that the Airline Model should be discarded in favor of the more parsimonious differencing model, which has no parameters to estimate.

The Model Selection Criterion

Return to the Develop Models window (Display 26.9) and notice the Root Mean Square Error button at the right side of the table banner. This is the model selection criterion–the statistic used by the system to select the best fitting model. So far in this example you have fit two models and have left the default criterion, root mean square error (RMSE), in effect. Because the Airline Model has the smaller value of this criterion, and because smaller values of the RMSE indicate better fit, the system has chosen this model as the forecasting model, indicated by the check box in the `Forecast Model` column.

The statistics available as model selection criteria are a subset of the statistics available for informational purposes. To access the entire set, select `Options` from the menu bar, and then select `Statistics of Fit`. The `Statistics of Fit Selection` window appears, as shown in Display 26.12.

Display 26.12. Statistics of Fit
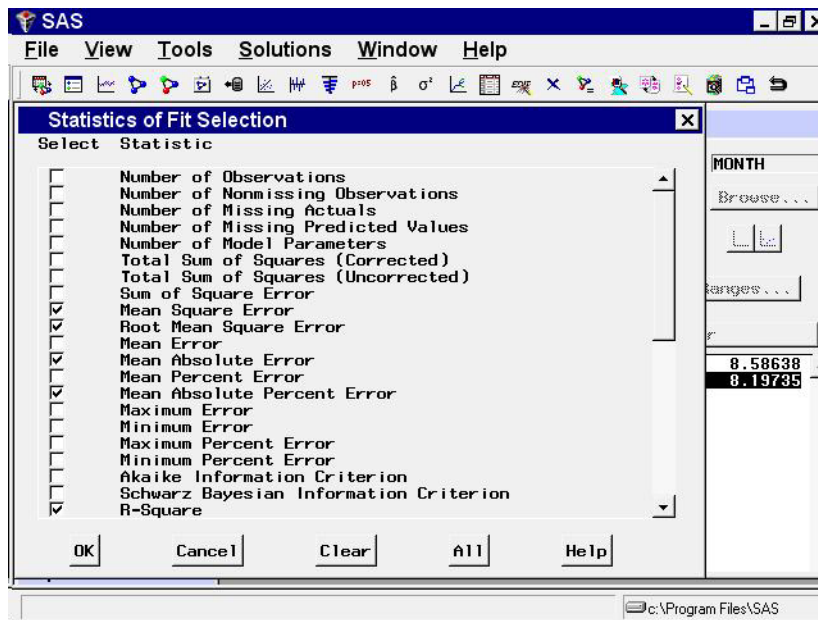
By default, five of the more well known statistics are selected. You can select and deselect statistics by clicking the check boxes in the left column. For this exercise, select `All`, and notice that all the check boxes become checked. Select the `OK` button to close the window. Now if you choose `Statistics of Fit` in the `Model Viewer` window, all of the statistics will be shown for the selected model.

To change the model selection criterion, click the `Root Mean Square Error` button or select `Options` from the menu bar and then select `Model Selection Criterion`. Notice that most of the statistics of fit are shown, but those which are not relevant to model selection, such as number of observations, are not shown. Select `Schwarz Bayesian Information Criterion` and click `OK`. Since this statistic puts a high penalty on models with larger numbers of parameters, the ARIMA(0,1,0)(0,1,0)s model comes out with the better fit.

Notice that changing the selection criterion does not automatically select the model that is best according to that criterion. You can always choose the model you want to use for forecasts by selecting its check box in the `Forecast Model` column.

Now bring up the Model Selection Criterion window again and select `Akaike Information Criterion`. This statistic puts a lesser penalty on number of parameters, and the Airline Model comes out as the better fitting model.

Sorting and Selecting Models

Select `Sort Models` on the `Tools` pull-down menu or from the toolbar. This sorts the current list of fitted models by the current selection criterion. Although some selection criteria assign larger values to better fitting models (for example, R-square) while others assign smaller values to better fitting models, Sort Models always orders

models with the best fitting model–in this case, the Airline Model–at the top of the list.

When you select a model in the table, its name and criterion value become highlighted, and actions that apply to that model become available. If your system supports a right mouse button, you can click it to invoke a pop-up menu, as shown in Display 26.13.

Display 26.13. Right Mouse Button Pop-up Menu
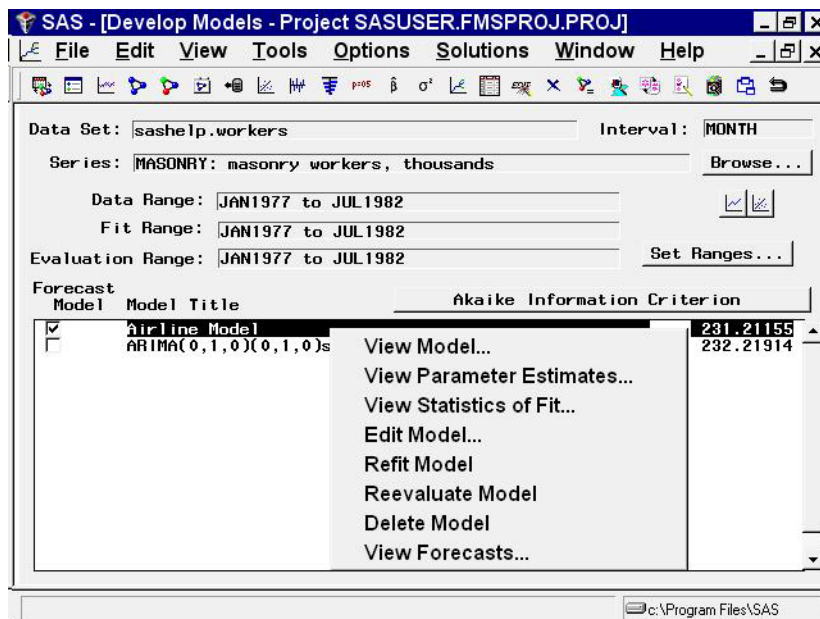
Whether or not you have a right mouse button, the same choices are available under `Edit` and `View` from the menu bar. If the model viewer has been invoked, it is automatically updated to show the selected model, unless you have unlinked the viewer using the `Link/Unlink` toolbar button.

Select the highlighted model in the table again. Notice that it is no longer highlighted. When no models are highlighted, the right mouse button pop-up menu changes, and items on the menu bar that apply to a selected model become grayed out. For example, you can choose `Edit` from the menu bar, but you can't choose the `Edit Model` or `Delete Model` selections unless you have highlighted a model in the table.

When you select the check box in the `Forecast Model` column of the table, the model in that row becomes the forecasting model. This is the model that will be used the next time forecasts are generated by choosing `View Forecasts`, or by using the `Produce Forecasts` window. Note that this forecasting model flag is automatically set when you use `Fit Automatic Model` or when you fit an individual model that fits better, using the current selection criterion, than the current forecasting model.

Comparing Models

Select `Tools` and `Compare Models` from the menu bar. This displays the `Model Fit Comparison` table, as shown in Display 26.14.

Display 26.14. Model Comparison Window

The two models you have fit are shown as `Model 1` and `Model 2`. When there are more than two models, you can bring any two of them into the table by selecting the up and down arrows. In this way, it is easy to do pairwise comparisons on any number of models, looking at as many statistics of fit as you like. Since you previously chose to display all statistics of fit, all of them are shown in the comparison table. Use the vertical scroll bar to move through the list.

After you have examined the model comparison table, select the `Close` button to return to the `Develop Models` window.

Controlling the Period of Evaluation and Fit

Notice the three time ranges shown on the `Develop Models` window (Display 26.9). The data range shows the beginning and ending dates of the MASONRY time series. The period of fit shows the beginning and ending dates of data used to fit the models. The period of evaluation shows the beginning and ending dates of data used to compute statistics of fit. By default, the fit and evaluate ranges are the same as the data range. To change these ranges, select the `Set Ranges` button, or select `Options` and `Time Ranges` from the menu bar. This brings up the `Time Ranges Specification` window, as shown in Display 26.15.

Display 26.15. Time Ranges Specification Window
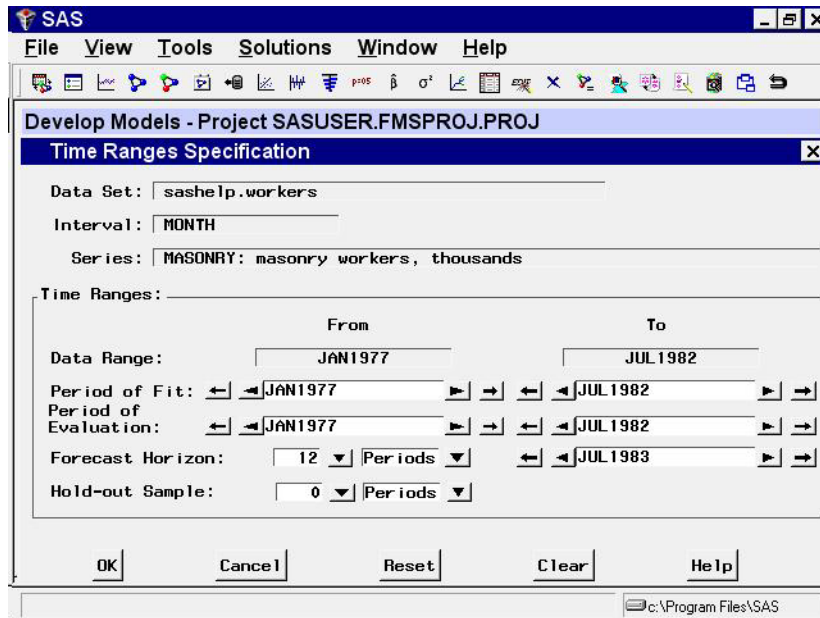
For this example, suppose the early data in the series is unreliable, and you want to use the range June 1978 to the latest available for both model fitting and model evaluation. You can either type in JUN1978 in the From column for Period of Fit and Period of Evaluation, or you can advance these dates by clicking the right pointing arrows. The outer arrow advances the date by a large amount (in this case, by a year), and the inner arrow advances it by a single period (in this case, by a month). Once you have changed the Period of Fit and the Period of Evaluation to JUN1978 in the From column, select the OK button to return to the Develop Models window. Notice that these time ranges are updated at the top of the window, but the models already fit have not been affected. Your changes to the time ranges affect *subsequently fit* models.

Refitting and Reevaluating Models

If you fit the ARIMA(0,1,0)(0,1,0)s and Airline models again in the same way as before, they will be added to the model list, with the same names but with different values of the model selection criterion. Parameter estimates will be different, due to the new fit range, and statistics of fit will be different, due to the new evaluation range.

For this exercise, instead of specifying the models again, refit the existing models by selecting Edit from the menu bar and then selecting Refit Models and All Models. After the models have been refit, you should see the same two models listed in the table but with slightly different values for the selection criterion. The ARIMA (0,1,0)(0,1,0)s and Airline models have now been fit to the MASONRY series using data from June 1978 to July 1982, since this is the period of fit you specified. The statistics of fit have been computed for the period of evaluation, which

1320

was the same as the period of fit. If you had specified a period of evaluation different from the period of fit, the statistics would have been computed accordingly.

In practice, another common reason for refitting models is the availability of new data. For example, when data for a new month become available for a monthly series, you might add them to the input data set, then invoke the forecasting system, open the project containing models fit previously, and refit the models prior to generating new forecasts. Unless you specify the period of fit and period of evaluation in the `Time Ranges Specification` window, they default to the full data range of the series found in the input data set at the time of refitting.

If you prefer to apply previously fit models to revised data without refitting, use `Reevaluate Models` instead of `Refit Models`. This recomputes the statistics of fit using the current evaluation range, but does not re-estimate the model parameters.

Using Hold-out Samples

One important application of model fitting where the period of fit is different from the period of evaluation is the use of hold-out samples. With this technique of model evaluation, the period of fit ends at a time point before the end of the data series, and the remainder of the data are held out as a nonoverlapping period of evaluation. With respect to the period of fit, the hold-out sample is a period in the future, used to compare the forecasting accuracy of models fit to past data.

For this exercise, use a hold-out sample of 12 months. Bring up the `Time Ranges Specification` window again by selecting the `Set Ranges` button. Set `Hold-out Sample` to 12 using the combo box, as shown in Display 26.16. You can also type in a value. To specify a hold-out sample period in different units, you can use the `Periods` combo box. In this case, it will allow you to select years as the unit, instead of periods.

Display 26.16.   Specifying the Hold-out Sample Size
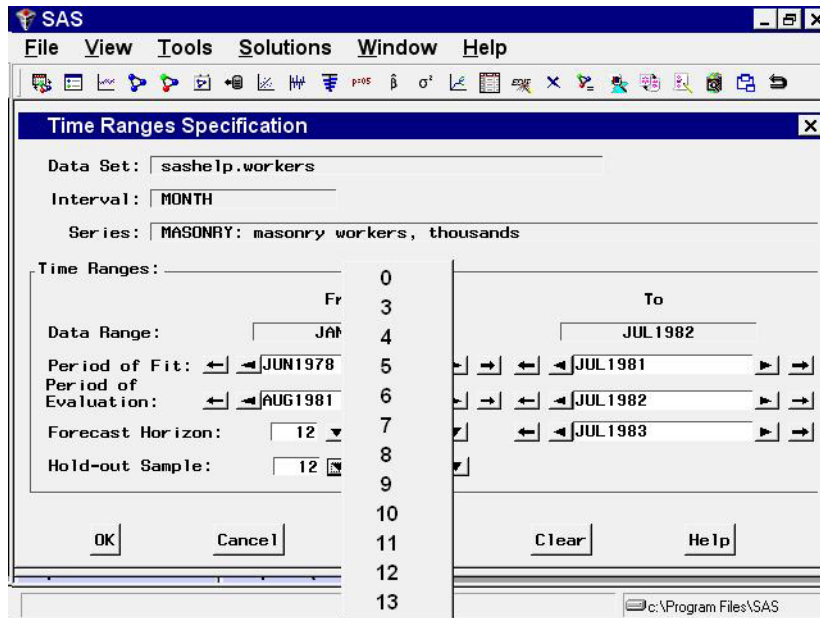
Notice that setting the holdout sample to 12 automatically sets the fit range to JUN1978–JUL1981 and the evaluation range to AUG1981–JUL1982. If you had set the period of fit and period of evaluation to these ranges, the hold-out sample would have been automatically set to 12 periods.

Select the OK button to return to the Develop Models window. Now refit the models again. Select Tools and Compare Models to compare the models now that they have been fit to the period June 1978 through July 1981 and evaluated for the hold-out sample period August 1981 through July 1982. As shown in Display 26.17, the ARIMA $(0,1,0)(0,1,0)$s model now seems to provide a better fit to the data than does the Airline model. It should be noted that the results can be quite different if you choose a different size hold-out sample.

Display 26.17.   Using 12 Month Hold-out Sample

SAS/ETS User's Guide, Version 8