

The Time Series Forecasting System

Charles Hallahan, Economic Research Service/USDA, Washington, DC
Renee Samy, SAS Institute Inc., Cary, NC

INTRODUCTION

The Time Series Forecasting System (TSFS) is a component of SAS/ETS[®] that provides a menu-based front-end for forecasting activities. The tasks of creating a date variable, graphing a data series and quickly seeing the results of differencing and/or applying a log transformation, testing for unit roots, examining autocorrelation and partial autocorrelation plots, performing seasonality tests, and, finally, estimating models and producing forecasts are just a mouse click away. This Hands-on Workshop is an introduction to the TSFS. Part 1 shows how to access the system and generate forecasting models for several variables through automatic selection from a default list of models provided with the TSFS. Part 2 shows how to override the defaults of the TSFS and take control of specifying a model. Part 3 shows how to add explanatory variables to the forecasting model and control the period of evaluation and fit of a model. This paper is written using the SAS[®] System Version 6.12.

PART 1: USING THE TSFS IN AUTOMATIC MODE

The first step is to make sure a libname statement has been executed to point to the directory containing the SAS data set to be analyzed. In the workshop, you will use SAS data sets found in the SASHELP directory. The TSFS can be accessed either through menu choices (see Figure 1) as

Globals → Analyze → Forecasting

or by entering the command **forecast** on the command line or by assigning the command **forecast** to a function key.

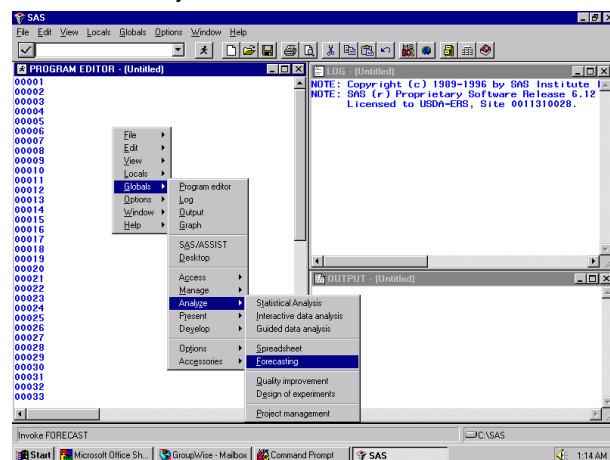


Figure 1 Accessing the TSFS

You will be using two variables from the CITIMON SAS data set found in the SASHELP library. When the initial TSFS screen appears (see Figure 2), name the project SASUSER.TSFS.PROJ with a description of "SUGI'24 TSFS Hands-on Workshop: Part 1" (no quotes). Click on the down arrow beside Data Set: and select SASHELP.CITIMON. Note that a date variable was found in the data set with the name DATE and with a periodicity of MONTH.

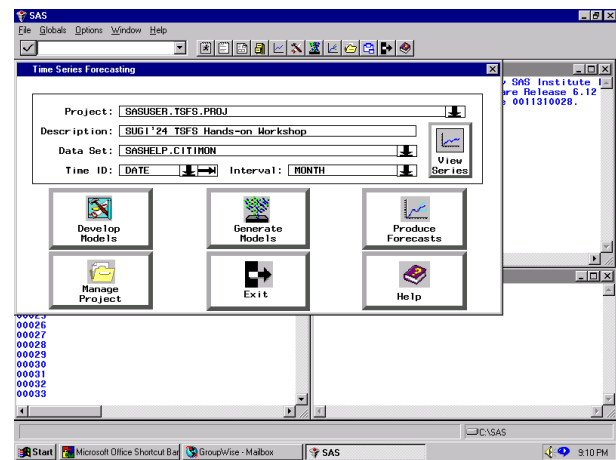


Figure 2 Starting a Project

You will be working with two variables in this data set: EEGP = (retail gasoline price) and RCARD = (retail sales of new domestic cars). To view a series, click on the View Series button and select RCARD (see Figure 3).

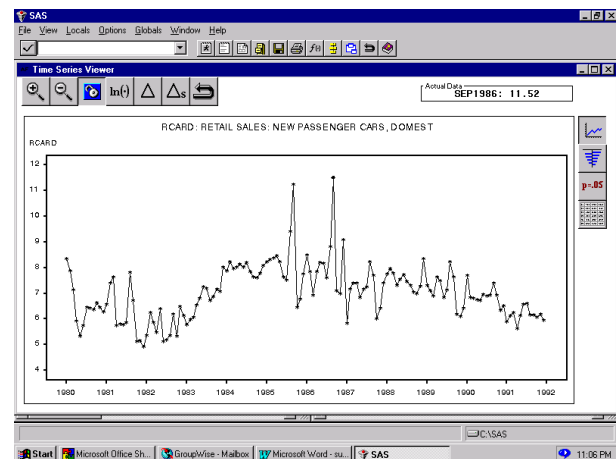


Figure 3 Viewing a Series

Clicking on a point on the graph produces the actual data value in a box on the upper right. The buttons along the upper left are for zooming and transformations. For example, clicking on the Δ key creates the first difference and immediately graphs it. The keys down the upper right produce the sample autocorrelation and partial autocorrelation functions and also carry out white noise and unit root tests. You will be looking at these buttons later when you fit specific models.

Clicking on the return arrow along the top brings you back to the main menu. Your first task will be to let the software find the best fitting model for the two variables by selecting from a preset list of models and using RMSE, the root mean square error, as the default goodness-of-fit measure.

From the Time Series Forecasting Menu (see Figure 2), click on the **Generate Models** button. From the **Generate Forecasting Models** screen (see Figure 4) click on the down arrow on the line **Series to Process**: and select the two variables EEGP and RCARD from the eighteen variables available in the data set. Hold down the CTRL key while clicking on the two selected variables.

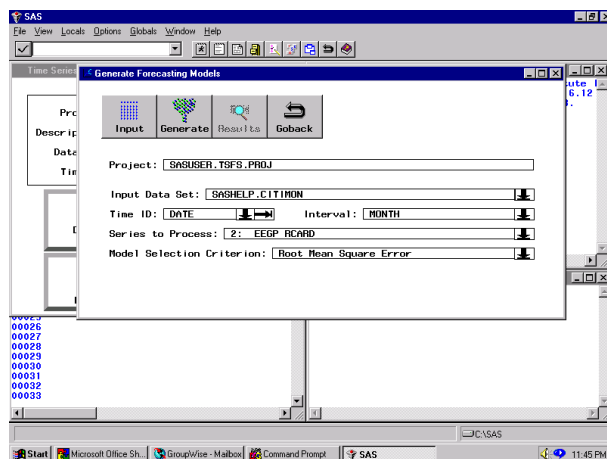


Figure 4 The Generate Models Screen

Before clicking on the **Generate** button to have the best model selected on the basis of the default Model Selection Criterion of the Root Mean Square Error, look at some of the other default options. The default is to keep just the Best Model. Clicking on the **Options** pull-down menu along the top, click on the option **Automatic Fit...** and change the default to keep the Best 5 Models based on RMSE. Note that the option to have the TSFS automatically perform diagnostics on the variables is checked. These diagnostics will be discussed later. See Figure 5 for this step.

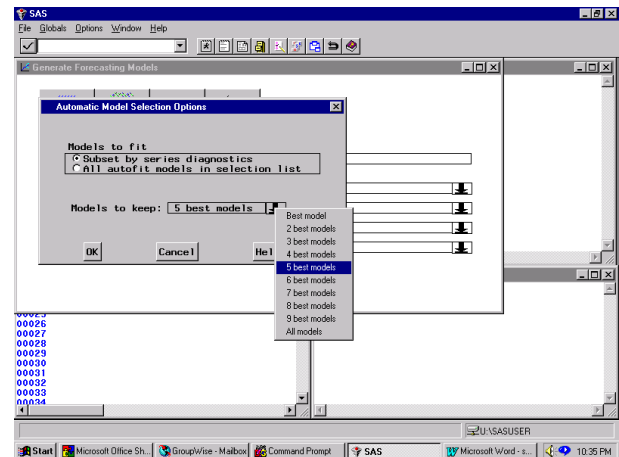


Figure 5 The Automatic Model Selection Options

Another option you will look at now is to see exactly which models are in the selection list as candidates for best fitting model. Again click on the Options pull-down menu and select **Model Selection List...**. The result is Figure 6.

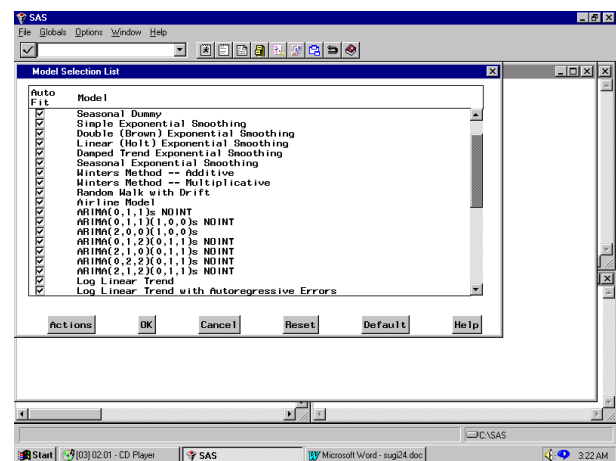


Figure 6 Model Selection List

The models in this list are characterized by three factors:

- trend - either none, deterministic or stochastic (that is, first or second difference)
- log or not
- seasonality or not.

For example, Winter's method is a smoothing model for data exhibiting a trend and seasonality; no log is taken. The TSFS documentation clearly explains all the available models. Favorite models not included in this list can be added, as you will see in Part 2. Since you have selected the option to automatically have TSFS perform the preliminary diagnostics, the list of models actually fit to a data series is much reduced.

Clicking on the **Generate** button leads to the message box stating “Models will be fit automatically for 2 series.” Based on the results of the diagnostic tests, a status box identifies each candidate model as it is being fit. If the diagnostic tests are inconclusive, then, for example, models both with and without logging are fit. For this example 10 models are fit to EEGP and 20 models to RCARD. Upon completion of the model fitting, the 5 best models (as previously designated) for each series are summarized in the **Generate Models Results** window (see Figure 7).

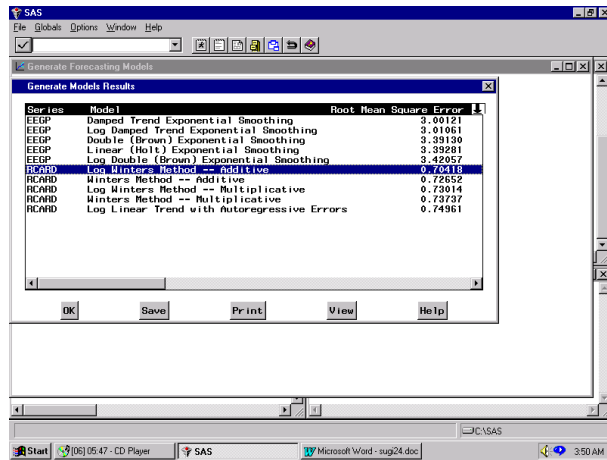


Figure 7 Generate Models Results Window

Since the object of all this work is to produce forecasts, at this point you could accept the best fitting model as found by TSFS and find forecasts for the next 12 months (the default).

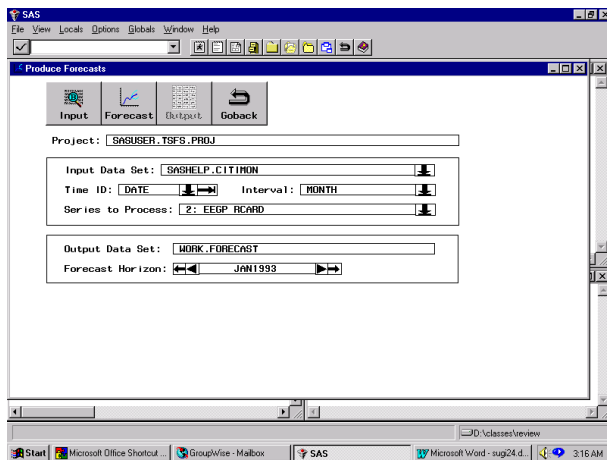


Figure 8 Produce Forecasts

Press the **Goback** key until the main menu reappears (see Figure 2), and click on the **Produce Forecasts**

button to get the **Produce Forecasts** screen (see Figure 8).

To stay with the defaults, click on the Forecast button and a note appears stating that “Forecasts for 2 series were written to output data set WORK.FORECAST.” There are three output formats available. Since context sensitive help is only an F1 function key click away, click F1 to see that the three output types are simple, interleaved, and concatenated. The default is simple and is explained on the Help screen as

Simple

specifies the simple format for the output data set. The data set contains the time ID variable and the forecast variables and 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.

This concludes Part 1 on using the TSFS in automatic mode. Return to the main screen (see Figure 2) by clicking on the **Goback** key. In Part 2, you will take a more proactive role.

PART 2: USING THE TSFS TO DEVELOP MODELS

You will use the SAS data set, SASHELP.CITIMON, as in Part 1. Click on the **Develop Models** button from the main screen to display the **Series Selection**

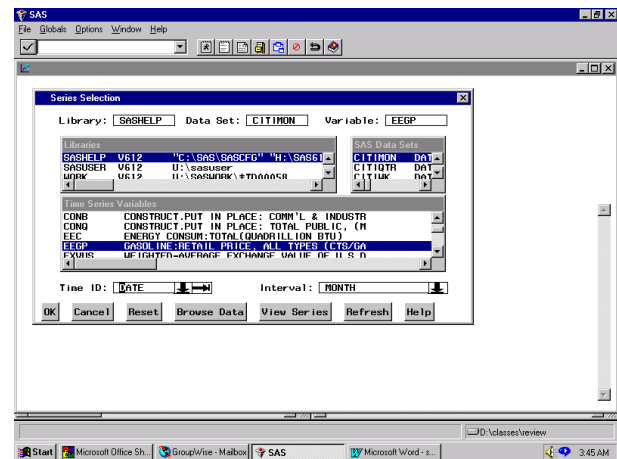


Figure 9 Series Selection

screen (see Figure 9).

Start with the variable EEGP. The first step is to look at a graph of the data, so click the **View Series** button (see Figure 10). The series does not appear to have a

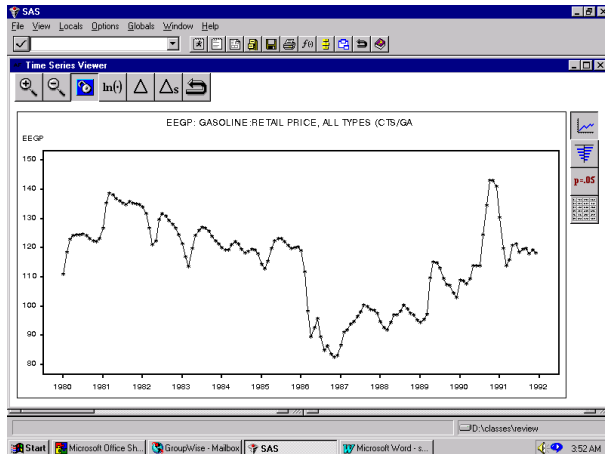


Figure 10 Plot of EEGP

trend. If you click on the high and low points on the graph, low prices seem to occur generally during the winter months and most peaks are in the summer months. There could be seasonality in the data.

At this point, you could run the three diagnostic tests by selecting

Locals → Diagnose Series

from the menu pull-down along the top. The result is

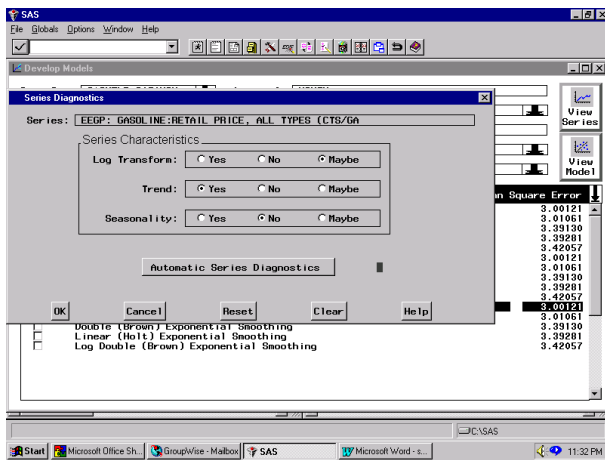


Figure 11 Series Diagnostics

Figure 11.

Clicking on the **Automatic Series Diagnostics** box yields

Log Transform: Maybe
Trend: Yes
Seasonality: No

Other transformations available when developing a model are the logistic, square root, and Box Cox transformations. Trends can be modeled by first or

second differences or by deterministic curves, such as a linear or quadratic trend. Seasonality can be modeled deterministically with seasonal dummies or via seasonal ARIMA models. Smoothing models can also incorporate trends and seasonality. The documentation spells out the details of all the possible models. The documentation also states that “The series diagnostic tests are intended as a heuristic tool only, and no statistical validity is claimed for them.” As of Release 7, the diagnostic tests are as follows:

- **Log transform test.** A high-order autoregressive model is fit to the original series, and its log and goodness-of-fit measures for prediction errors are compared.
- **Trend test.** An augmented Dickey-Fuller test and random walk with drift tests are performed.
- **Seasonality test.** A seasonal dummy model with AR(1) errors is fit and an F-test performed on the seasonal dummies.

Before taking first differences of EEGP, you will close the **Series Diagnostics** window and click the **View Series** button. Clicking on the Tests button on the right (designated as “p=.05”), displays graphical results for the Trend tests discussed above. You can see numerical p-values by clicking on one of the blue bars for any of the tests. For example, clicking on the bar for lag 1 of the Unit Root Tests yields Figure 12.

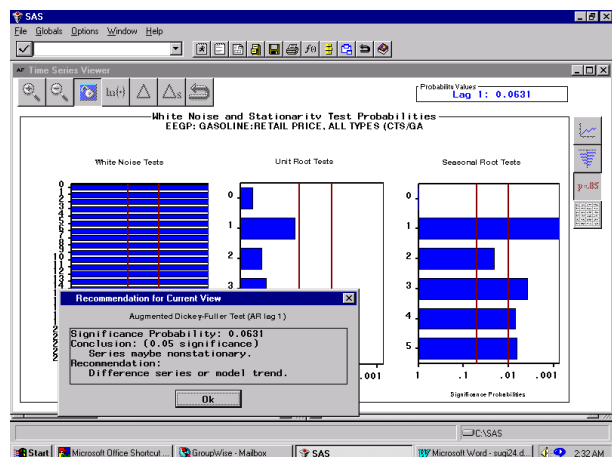


Figure 12 p-value for Unit Root Test

The p-value of 0.0631 leads to the recommendation that the series be differenced.

Another indicator of the need for first differencing comes from examination of the autocorrelation function. This plot is obtained by clicking the second icon from the top along the right (see Figure 13).

The slow decay of the autocorrelations is the classic sign of nonstationarity of the series.

To look at a plot of the first differences of EEGP, you can go back to the **Time Series Viewer** (see Figure

10) by clicking on the top icon along the right and then clicking on the first difference icon, Δ .

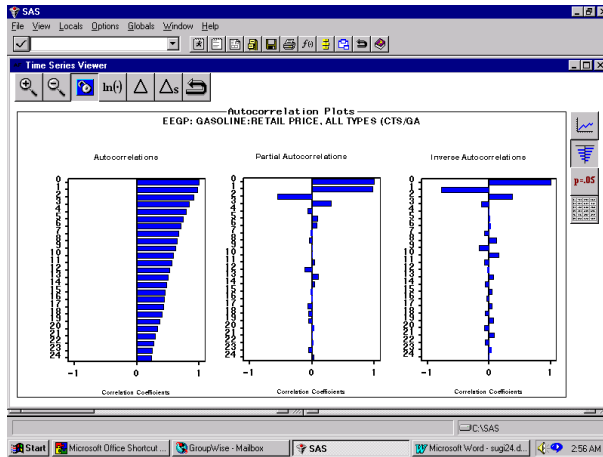


Figure 13 Autocorrelation Plots

The plot of the first differences is shown in Figure 14.

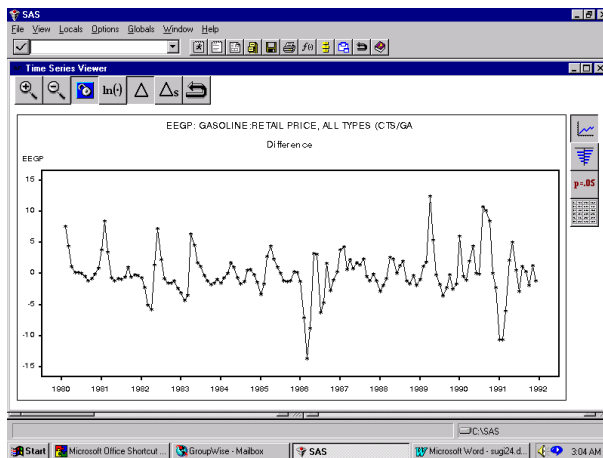


Figure 14 Plot of First Differences

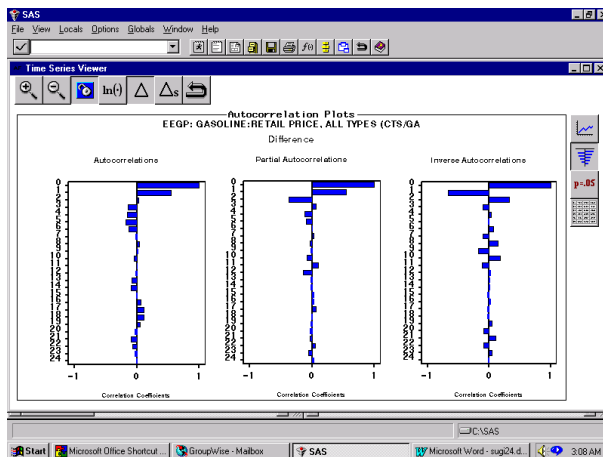


Figure 15 Autocorrelations of First Differences

The autocorrelations for the first differences are shown in Figure 15.

The plot of first differences shows no obvious evidence of a trend remaining in the series, and the autocorrelation plot is consistent with a stationary series. Examination of the autocorrelation and partial autocorrelation plots suggests that an autoregressive model in first differences might be appropriate, that is, an IAR(2,1) model. Clicking on the **Goback** icon on the right along the top returns you to the **Develop Models** screen (see Figure 16).

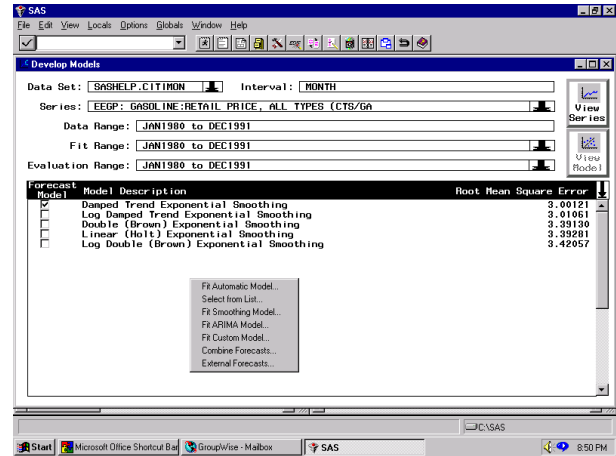


Figure 16 Develop Models

Note that the best model according to minimum RMSE found from the automatic model selection in Part 1 is the Damped Trend Exponential Smoothing model with a RMSE of 3.0012. Make sure that none of the five models on this screen are highlighted and right-click on an empty part of the screen. Select **Fit ARIMA Model** from the pop-up menu. The **ARIMA Model Specification** screen appears. Click on the down arrow beside **Autoregressive: p=** to select 2 and beside **Differencing: d=** to select 1 (see Figure 17).

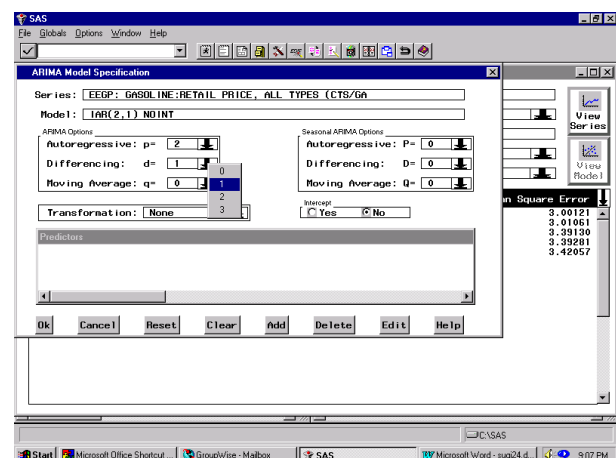


Figure 17 ARIMA Model Specification

Clicking "Ok" estimates the IAR(2,1) Model and returns you to the **Develop Models** screen where you see that the IAR(2,1) model has a RMSE of 2.82 and, hence, is the best model found so far according to the RMSE criterion. The IAR(2,1) model was not found before because it did not appear on the **Model Selection List**. You will see how to add a model to the list later. For now, highlight the IAR(2,1) model, right-click and select the first choice on the popup menu, **View Model....** This brings up the **Model Viewer** screen (see Figure 18).

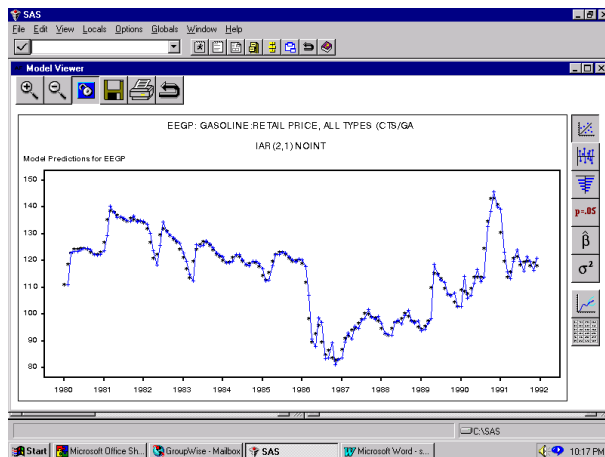


Figure 18 Model Viewer

The buttons along the right aid in assessing the adequacy of the model. They are, in order from the top

- Predictions (the default in Figure 18)
- Prediction Errors
- Residual ACF
- Tests
- Parameters
- Statistics
- Plot Forecasts
- Browse Forecasts

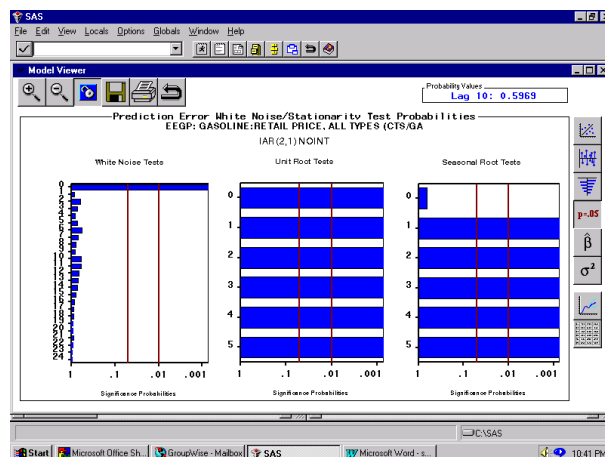


Figure 19 Residual Tests

Clicking on the 4th button for Tests gives Figure 19 and shows no evidence of structure left in the residuals (the significant result for a seasonal unit root test with 0 lags seems to be spurious).

Clicking on the 5th button for Parameters shows that the two AR parameter estimates are highly significant. At this point, you will stop searching for further models and return to the issue of adding a model to the Model Selection List.

From the menu bar along the top, click on

Options → Model Selection List...

Click on the **Actions** button on the lower left and click on **Add**. Three choices are presented:

Smoothing Model...
Custom Model...
ARIMA Model...

Select **ARIMA Model...** to display the **ARIMA Model Specification** screen (as in Figure 17). Make the same choices as before to define the IAR(2,1) model and click on **OK**. The **Model Selection List** screen now includes IAR(2,1) NOINT.

Click on **OK** to return to the **Develop Models** screen. To see the forecasts for the IAR(2,1) model, click on the line for that model to highlight it (it should be the last model in the list). Clicking on the **View Model** button on the upper right (see Figure 16) brings up the **Model Viewer** screen (see Figure 18). Click on the bottom button on the right, **Plot Forecasts**. To zoom in on the end of the series, click on the **Zoom In** button at the left along the top. The mouse pointer changes its appearance; holding down the left mouse button, trace out a box including the last few years of the graph (see Figure 20).

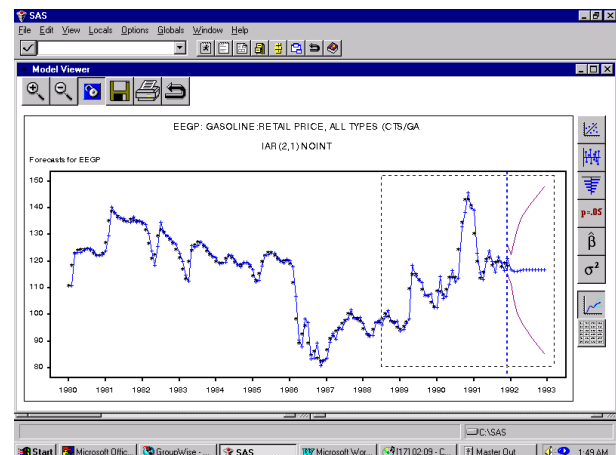


Figure 20 Zooming In on Plot Forecasts

Releasing the mouse button results in the zoomed plot, Figure 21.

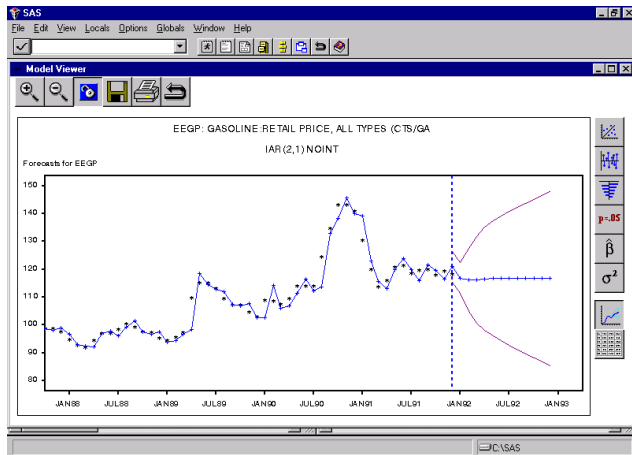


Figure 21 Zoomed Plot Forecasts

A spreadsheet view of the actual and predicted values, errors, and confidence limits is available by clicking on the **Browse Forecasts** button on the bottom along the right.

At this point you might be interested in combining two of the forecast models, say the Damped Trend Exponential Smoothing and IAR(2,1) models. Return to the **Develop Models** screen, make sure no model is highlighted (click on a model if it is already highlighted to toggle selection off), and right click in a blank area. Select the option **Combine Forecasts...** (the next to last choice) to get the **Forecast Model Specification** screen. Click on the two models mentioned above (see Figure 22).

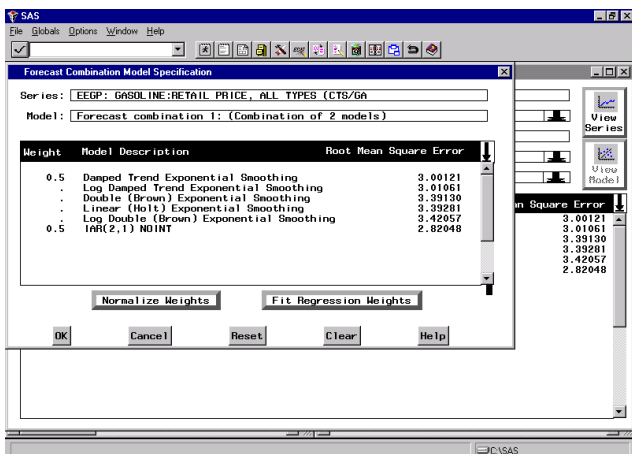


Figure 22 Forecast Combination Model Specification

The default is to **Normalize Weights**, that is, assign equal weights to the forecasts being combined. Click

on **Ok** to complete the forecast combination. Control is passed back to the **Develop Models** screen where you see that the combined forecast has an RMSE of 2.87, not as good as the IAR(2,1) model with an RMSE of 2.82.

You could return to the **Forecast Combination Model Specification** screen and select **Fit Regression Weights** this time (see Figure 22). In this case, the weights for the two models are 0.1094 and 0.8904 with a RMSE of 2.84. The IAR(2,1) still remains the best model (as measured by RMSE) found so far.

PART3: ADDING EXPLANATORY VARIABLES TO THE MODEL

Now you will see if you can improve the forecasting model for RCARD by using EEGP as an explanatory variable. When you fit automatic models to both EEGP and RCARD, you found that the best fitting model was the Log Winters Method – Additive, which had an RMSE = .70. The Winters Method – Additive had only a slightly larger RMSE of 0.72. To use EEGP as an explanatory variable in a model for RCARD, return to the **Develop Models** screen (see Figure 16) and select RCARD as the Series. Click in a blank area of the screen and select **Fit Custom Model...** to get the **Custom Model Specification** screen. In the lower portion of the screen labeled Predictors, click and select **Add Regressors...** and select EEGP. The result is Figure 23.

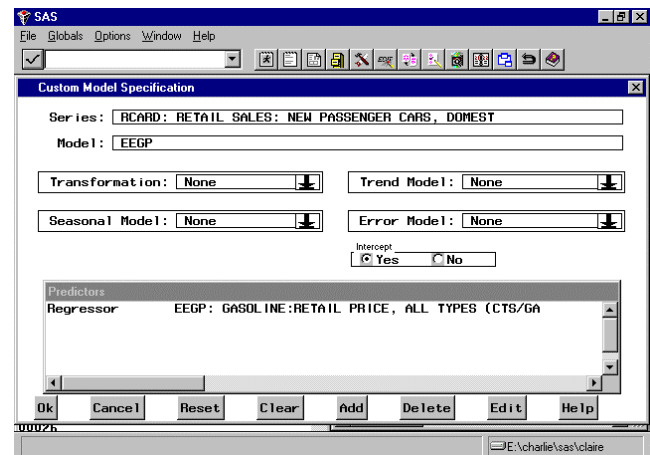


Figure 23 Custom Model Specification

Note that when you select **Add Regressors...** the other choices included Time Trends, Adjustments, Dynamic Regressors, Interventions, and Seasonal Dummies.

Checking OK will fit the model, which you see has a RMSE of 0.97. To try and improve on this model, which is labeled EEGP in the Develop Models screen, highlight this model, right click, and select **Edit Model...** Click on the Add button at the bottom of the screen, and, since RCARD does exhibit seasonality,

you select Seasonal Dummies. Clicking on OK tells us that the model for RCARD with EEGP and seasonal dummies as explanatory variables has an RMSE of 0.91, still not as good as the models found through the Automatic Fit.

The next step would be to examine the residuals of this model by right clicking on the model and selecting **View Model**. Clicking on the second button from the top on the right produces the residual ACF and PACF (see Figure 24).

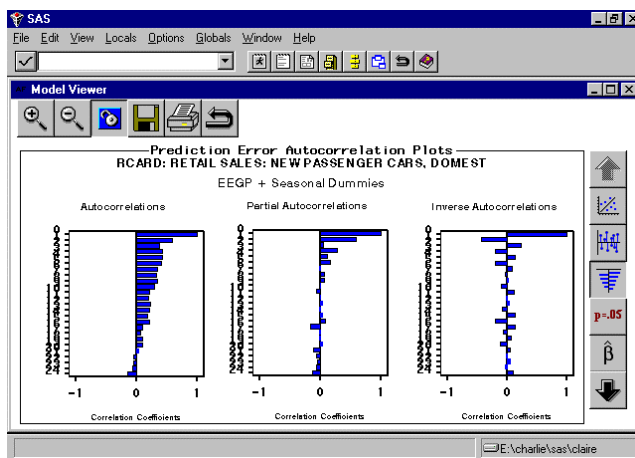


Figure 24 Residual ACF and PACF

Examination of these plots reveals plenty of structure remaining in the residuals. Carry the model development phase one more step by modeling the residual as an ARMA(1,1) process. To do this, click on the **Goback** key along the top of the **Model Viewer** screen and once more right click on the model EEGP + Seasonal Dummies. Select **Edit Model** and, in the **Error Model Options**, select $p=1$ and $q=1$ (see Figure 25).

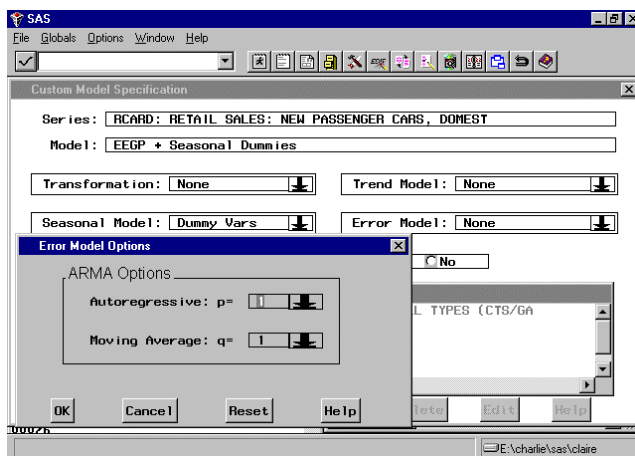


Figure 25 Error Model Options

Click on OK to close the **Error Model Options** window and once more on OK to fit the model. You return to the **Develop Models** screen where you see

that the model now labeled EEGP + Seasonal Dummies + ARMA (1,1) has an RMSE of 0.72, essentially the same as the best model found through the **Automatic Model Generation** option. You could continue to diagnose this model by examining the residuals, but since additional editing of the model should be clear, stop here.

The last point covered in this workshop concerns the time ranges used to fit and assess the models. You can access the **Time Ranges Specification** window several ways, for example, by clicking on the **Options** pull-down and selecting **Time Ranges** (see Figure 26).

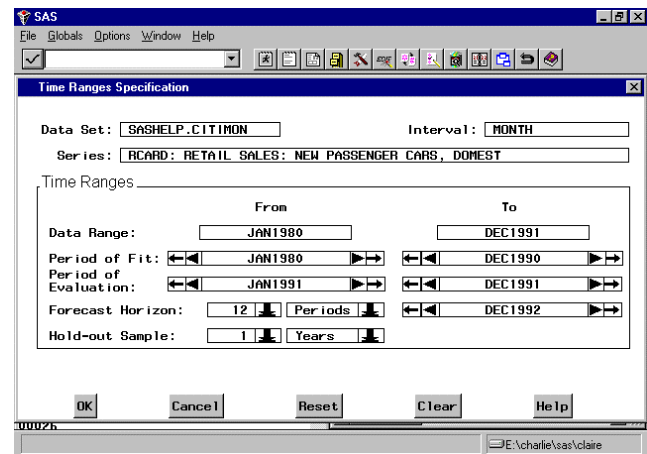


Figure 26 Time Ranges Specification

The **Data Range** is JAN1980 to DEC1991. These are also the default settings for the **Period of Fit** and **Period of Evaluation**. The **Hold-out Sample** is set to 0 periods. Suppose you wanted to fit the model over the period JAN1980 – DEC1990 and evaluate the model over the period JAN1991 – DEC1991. The easiest way to do this is to declare a **Hold-out Sample** of 1 Year. When you click on OK, you return to the **Develop Models** screen. To refit all the models for RCARD based on a hold-out sample of one year, click on the Edit pull-down

Edit → Refit Models → All Models

The RMSE's are now based on the 12 prediction errors for 1991 and they show that the best model is the Log Winters Method – Multiplicative with an RMSE of 0.57. The EEGP + Seasonal Dummies + ARMA(1,1) model has an RMSE of 0.65.

CONCLUSION

The Time Series Forecasting System provides an intuitive point-and-click interface to the forecasting tools of SAS/ETS software. You can extend the list of candidate models provided by the SAS System and through automatic diagnostic tests for trend,

seasonality, and log transformation, a large number of models can be automatically fit to a set of data series. Graphs of data series and error diagnostics are immediately available. Explanatory variables, time trends, seasonal dummies, and interventions are easily included in models. In summary, the TSFS converts the task of forecasting with the SAS System from a batch-oriented process to a straightforward point-and-click operation.

SAS and SAS/ETS are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION

Author Name: Charles Hallahan
Company: Economic Research Service/USDA
Address: Room 3117
1800 M St, NW
City, State ZIP: Washington, DC 20036-5831
Work Phone: 202-694-5051
Fax: 202-694-5718
e-mail: hallahan@econ.ag.gov