Forecasting Operational Indices Using SAS/ETS@ Software

Laura McCarty, Federal Express Corporation, Memphis, TN

Background

Forecasts of operational indices are useful tools not only when planning for similar conditions, but also when introducing new technologies. In the latter case, they can be used as bench-marks by which to measure increases or decreases in productivity that are attributable to the new process as opposed to time. Therefore, before introducing new technologies that are intended to improve productivity, Federal Express produces prediction values of what is expected under current operating policies.

Recently, a need for such forecasts arose. Forecasts of eighteen operational indices for seven different areas of the country were needed. This means that a total of one hundred and twenty-six forecasts of time series data were required. For anyone who has ever done forecasting, this is an overwhelming number of forecasts to generate.

In order to produce reliable forecasts of the time series data quickly and efficiently, a software package that fits the data to multiple models is necessary. The application also has to produce quantitative statistics, such as R-Square, RMSE, and MAPE for each model. Also, plots of the original data, including the data values, the autocorrelation function, and the partial autocorrelations, and of the predicted values, including residuals and residual autocorrelations, are necessary. Finally, the software has to provide predicted values and graphs of them. SAS/ETS@ version 6.9 software meets all of there requirements. Therefore, it was chosen for the above project.

Model Generation

For the analysis, four years of monthly data was used so each of the eighteen indices has forty-eight data points. The data, which resided in a mainframe database, was downloaded to a Unix file. Then, it was converted into a SAS data set that the forecasting application can read.

SAS/ETS software requires an interval variable that must be in a SAS date format. In this case, the interval variable is month, which is in the mmddyy6.

format. It also requires a series variable. This is the variable that is being forecasted. This project has eighteen series variables which were each forecasted independently. They will be referred to as index 1, index 2, ..., index 18. However, the process is the same for all of them.

After identifying an interval and a series variable, the series variable was plotted against the interval variable. SAS/ETS software provides six different ways to see this data. They are called View Series Options and are as follows:

1. View Data Values
2. Plot Data Values
3. Display Descriptive Statistics
4. Plot Autocorrelation Function
5. Plot Inverse Autocorrelations
6. Plot Partial Autocorrelations

Of the six options, three were implemented. They were Plot Data Values, Plot Autocorrelation Function, and Plot Inverse Autocorrelations. Also, the Difference Series and the Log Transform Series can be displayed, but these options were not used.

After viewing the series to check for seasonality, trends, and randomness, a criterion was selected from the Automatic Model Selection Criterion menu, which contains eighteen choices. For the first iteration, Root Mean Square Error was used as the criterion.
Next, the Add Model function was employed. When using Add Model to fit a model to the series data, there are three choices about how to do this. The are as follows:

1. Available Model List
2. New Specification
3. Automatic Model Selection.

The first option allows the user to select one of the thirty-seven available models which is then fit to the data. The second lets the user specify whether or not to use Time Trend, Log Transform, Arima, and many others. The final choice, which was the one used, fits all of the models selected by the user to the data and returns the one that provides the best fit based on the criterion selected previously.

On the first iteration, all models were included as candidates and evaluated. After fitting them, SAS/ETS software returned the model that best fit the data based on the criterion selected. Then, the Add Model function was rerun, this time without the model selected above. The result is the model that fits the data second best.

After obtaining two models for the first criterion, another criterion was selected.

The second criterion was Mean Absolute Percent Error. After choosing the criteria, the Add Model steps were repeated. On the first attempt at fitting a model with the second criteria, all models were again fitted to the data, including those chosen for the first criteria. The reason why will be explained later in the Model Comparison section.

After running the Add Model function using Automatic model selection twice for the second criterion, two more models were added to the model list. However, one or more of the new models might be the same as the two models selected using criterion one. Frequently, this is the case.

Next, another Automatic Model Selection Criterion was chosen, and the Add Model functions were run again. The third criterion used for this project is R-Square. This sequence was repeated once more. This time the Automatic Model Selection Criterion was Theil's Inequality Coefficient.

Finally, after adding models based on four different criterion, eight models resulted. However, some of them may have been the same. In fact, only two are necessarily different. For each of the indices forecasted in this analysis, the number of unique models ranged from two to six.

Model Comparison

After obtaining models as described above, they were compared to each other to get the model that best fit the data. Several methods of comparison from the View Model Options menu were used. The available choices are as follows:

1. Model Specifications
2. Model Parameters
3. Statistics of Fit
4. Plot of Residuals
5. Plot of Residual Autocorrelations
6. Plot of Residual Inverse Autocorrelations
7. Plot of Residual Partial Autocorrelations
8. Plot of Actual Versus Predicted Values

The first way that the models were compared was by using the Selection Criterion. On the View Model Options menu, this is the third choice, Statistics of Fit. The RSME's of each model were compared, and the model with the RMSE closest to zero was considered best. Then the MAPE's were compared. Again, the model with the MAPE closest to zero was considered best. The model with the R - Square closest to one was chosen, and the model with the smallest Theil's inequality coefficient was selected.

If one model had the closest to optimal values for all four criterion, it was chosen as the best model. However, as was often the case, different models were better depending on which criterion was used. In this case, all of the models that were best for a given criteria were included in further analysis. Also, models that performed reasonably well for all of the criteria, but not best for any one criteria, were included.

This is why two models were generated for each criteria. By having two, more options were available to choose from. In cases where there was not a definite best model, having second best models for each criteria was useful. In some cases, a model was best for one criteria and second best for one or more of the others. In this case, the second model might turn out to be the model that best represents the data.

After comparing models based on the selection criteria, they were also compared using graphs. The graphical tools, which were also chosen from the View Model Options menu, included Plot of Residuals, Plot of Residual Autocorrelations, and Plot of Residual Partial Autocorrelations.

Models with residuals that were either not random or not within control limits were immediately rejected.

Even if one model was definitely the best based on selection criteria, it was still judged graphically to make sure that the model not only mapped the existing data but also followed a likely trend into the future. This was done using Plot of Actual Versus Predicted Values.

Sometimes, models are able to match past data quite well, but do not serve as good predictors of the future. This happens when a trend changes late in the past data, and the model puts too much emphasis on the change.

Forecasts

After selecting the best one or two models based on the comparison techniques described above, forecasts were generated using those algorithms. The forecasts were displayed using options from the View Forecast Options menu which includes the following:

1. View Forecast Values
2. Plot Forecast
3. Display Descriptive Statistics

First, Forecast Values were displayed. The listing includes the forecast date, the series variable, the forecast, the upper ninety-five percent confidence interval, and the lower ninety-five percent confidence interval. It can be used to not only see the forecasted values but also to determine the size of the confidence interval.

Then, the Plot Forecast option was employed. The graph shows the historic data and the predicted values along with the ninety-five percent confidence intervals. Basically, it conveys the same information as the View forecast values option, but in graphical form.

Conclusion

The results obtained from the View Forecast Values were used to fulfill the project requirements. By using SAS/ETS software to obtain them, reliable and thorough results were obtained quickly and efficiently.