A Critical Look at the Automatic SAS® Forecasting System

By: Mohammad Afser Uddin

Advisor: Dr. Stephen M. Scariano

Sam Houston State University

Abstract

The goal of this effort is to study the SAS® automatic model selection and forecasting procedure for studying time series data. This procedure generates various time series model and suggests the best choice model based on either minimum mean squared error (MSE), minimum root mean square error (RMSE), minimum mean absolute error (MAE), and minimum mean absolute percent error (MAPE), along with several other accuracy measures. This study focuses on using the classic time series data sets given in Appendix B of *Introduction to Time Series Analysis and Forecasting* (2nd ed., p 581-626) by Douglas C. Montgomery, Cheryl L. Jennings & Murat Kulahci.

Introduction

All time series forecasting problems involve the use of time series data; however, the SAS(AFS) system can produce automatic forecasts and accuracy statistics for any input time series, and without human interaction. Additionally, this system can suggest the "best model" to use in practice, based on various measures of forecast accuracy, such as MSE, MAE and MAPE. To study this "best model", Table B.1, Table B.5 and Table B.9 data sets, which are given in Appendix B of *Introduction to Time Series Analysis and Forecasting*, are focuses of this presentation.

Table B.1 Analysis

Table B.1 comprises data on the market yield on US Treasury Securities at 10-year constant maturity from April 1953 through February 2007, where the percentage rate variable is collected monthly. The data set excludes last 12 monthly values as a holdout set in order to analyze the accuracy of SAS(AFS), which suggested a damped trend exponential smoothing model as the "best model" for this training, or initialization, data set. SAS(AFS) also produced the predicted values for the next 12 months. I have calculated different types of forecast accuracy measures based on the holdout data and predicted (or test) data. The model equation for this time series under damped trend exponential smoothing is $y_t = \mu_t + \beta_t\, t + \varepsilon_t$ where μ_t denotes for the level at time t, β_t represents the trend at time t, and ε_t denotes the random error associated with time t. Here, the level smoothing constant is $\alpha = 0.999$, the trend smoothing constant is $\gamma = 0.999$ and damping smoothing constant is $\phi=0.314$.

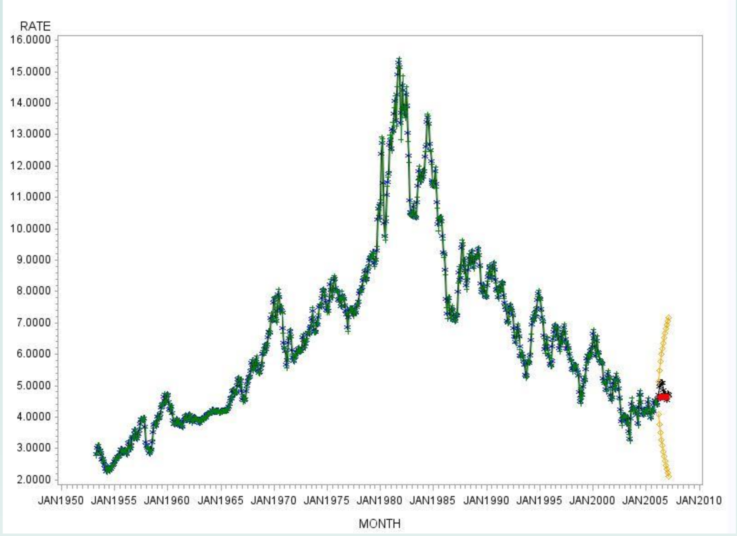

Figure 1a. Time series graph for predicted values, along with 12-month forecasted value and prediction limits. Green is for the predicted values, red is for forecasted values, blue is for the training data and black is for the test data.

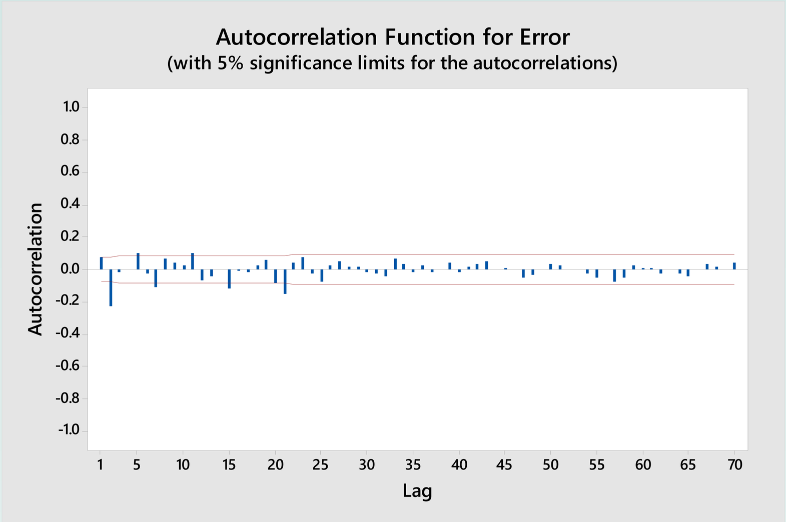

Figure 1b. Autocorrelation Function (ACF) for residuals

Data set	MSE	$\widehat{\sigma_{e(1)}}$	MAD	MPE	MAPE
Full data	0.0693	0.2633	0.1804	0.0204	2.7092
Training data	0.0703	0.2652	0.1818	0.0196	2.7202
Holdout data	0.0744	0.2728	0.2162	3.9258	4.3515

Table 1A. Accuracy Measures for table B.1 time series data

In Figure 1a, it seems the model fits the data appropriately, however, the SAS(AFS) could not forecast values with acceptably small prediction error. There are a few spikes in the ACF function, which means forecast errors may have additional structure not captured by the model. The MSE, RMSE and MAD are similar for both the training and holdout data sets; that is, the variability in forecast errors are approximately similar.

Table B.5 Analysis

Table B.5 comprises data on the US Beverage Manufacturer Product Shipments, Unadjusted from January 1992 through December 2006, where the dollar variable (in Millions) is collected monthly. The data set excludes the last 12 monthly values as a holdout set in order to analyze the accuracy of SAS(AFS), which suggested a Holt Winters–Additive model as the "best model" for this training, or initialization, data set. SAS(AFS) also produced the predicted values for the next 12 months. I have calculated different types of forecast accuracy measures based on the holdout data and predicted (or test) data. The model equation for this time series under the Holt Winters-Additive method is $y_t = \mu_t + \beta_t\, t + s_p(t) + \varepsilon_t$ in which μ_t stands for the level at time t, β_t represents the trend at time t, $s_p(t)$ represents the appropriate seasonal factor at time t and ε_t denotes the random error at time t. Here, the level smoothing constant is $\alpha = 0.407$, the trend smoothing constant is $\gamma = 0.00100$ and seasonal smoothing constant is $\delta=0.00100$.

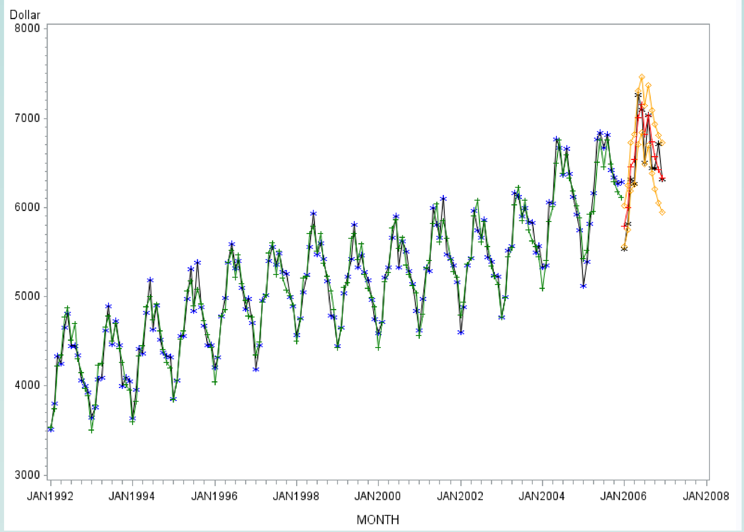

Figure 2a. Time series graph for predicted values, along with 12-month forecasted value and prediction limits. Green is for the predicted values, red is for the forecasted values, blue is for the training data and black is for the test data.

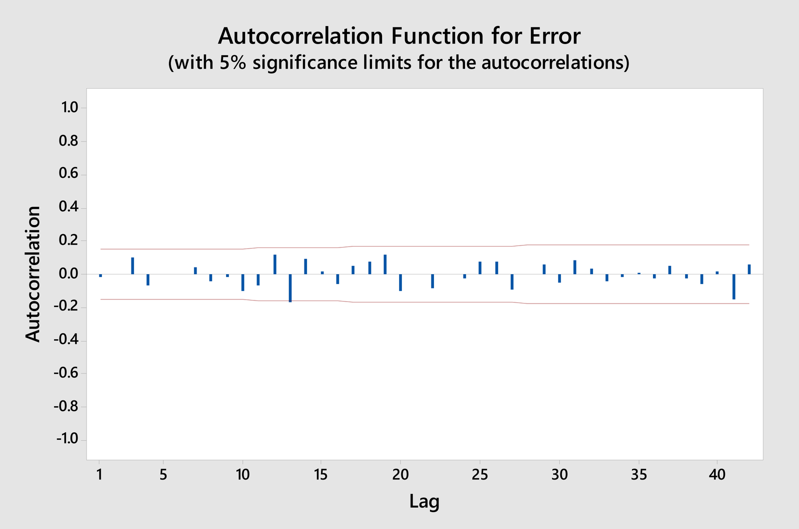

Figure 2b. Autocorrelation Function (ACF) for residuals

Data set	MSE	$\widehat{\sigma_{e(1)}}$	MAD	MPE	MAPE
Full data	15749.3	125.496	93.867	0.03149	1.77268
Training data	13639	116.787	91.348	0.01675	1.77321
Holdout data	45175.5	212.545	183.167	-1.52511	2.86881

Table 2A. Accuracy Measures for table B.5 time series data

Figure 2a gives the impression that the model fits the data appropriately well and it can forecast future values successfully. The ACF function in figure 2b suggests that the forecast errors are structureless. Although the MSE, RMSE and MAD appear large for the holdout data when compared to the training data, note that the MAPE (mean average percent error) is still quite small in all cases.

Table B.9 Analysis

Table B.9 comprises data on the International Sunspot Numbers from 1700 through 2004, where the sunspot number is collected yearly. The data set excludes the last ten year values as a holdout set in order to analyze the accuracy of SAS(AFS), which suggested a Simple Exponential Smoothing model as the "best model" for this training, or initialization, data set. SAS(AFS) also produced the predicted values for the next 10 years. I have calculated different types of forecast accuracy measures based on the holdout data and predicted (or test) data. The model equation for this time series, under simple exponential smoothing, is $y_t = \mu_t + \varepsilon_t$ where μ_t stands for the level at time t, and ε_t denotes the error associated with time t. Here, the level smoothing constant is $\alpha = 0.999$, indicating non-stationarity.

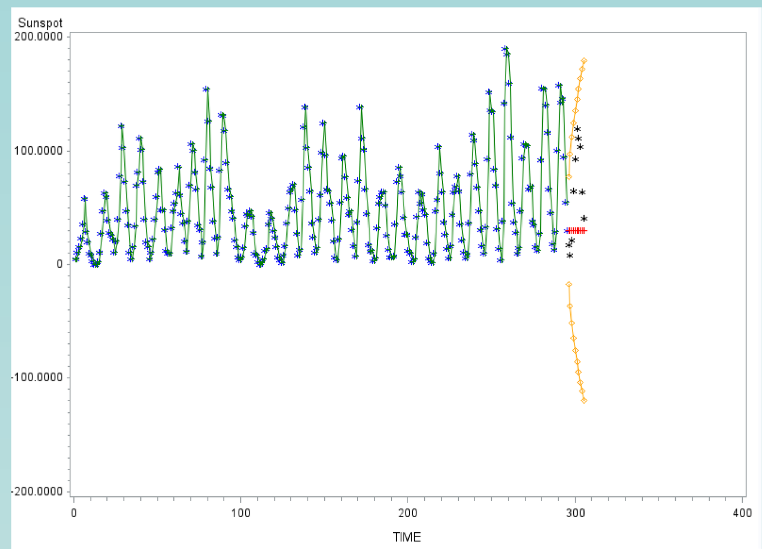

Figure 3a. Time series graph for predicted values, along with 12-month forecasted value and predicted limits. Green is for the predicted values, red is for the forecasted values, blue is for the training data and black is for the testing data.

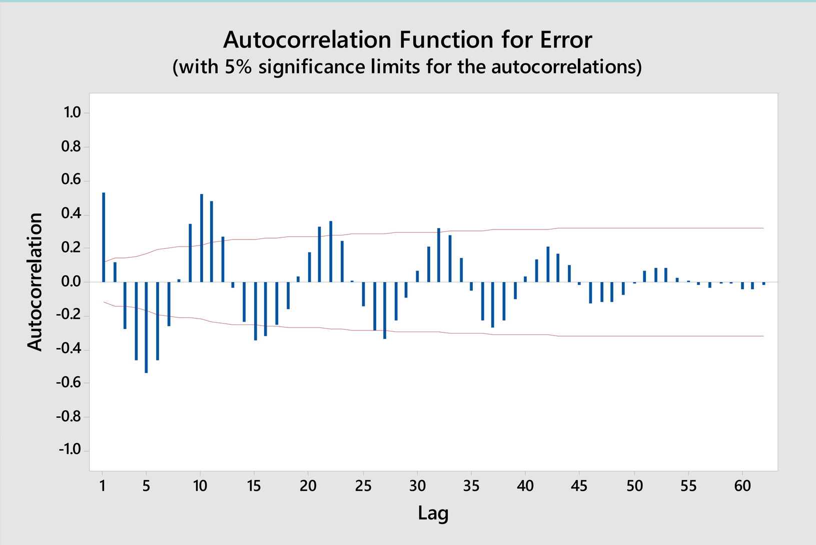

Figure 3b. Autocorrelation Function (ACF) for residuals

Data set	MSE	$\widehat{\sigma_{e(1)}}$	MAD	MPE	MAPE
Full data	579.8208	24.0795	18.2662	-37.0441	72.5157
Training data	578.9276	24.0609	18.1680	-37.8624	73.3212
Holdout data	2723.07	52.1830	42.9001	6.43757	77.3795

Table 3a. Accuracy Measures table B.9 time series data

Figure 3a clearly shows that the model suggested by SAS(AFS) is of little use in forecasting future values appropriately. Additionally, there are many spikes in the ACF function of residuals, which suggests that forecast errors have structure and the original model must be improved. Moreover, MSE, RMSE, MAD and MAPE are also very large. SAS(AFS) does not do well for this data set.

Conclusion

The SAS Automatic Forecasting System (AFS) for identifying the "best model" in time series applications is sometimes quite useful. However, in some instances that system is quite limited and of little value. In the end, it cannot be concluded that SAS(AFS) should be exclusively relied upon; the skill and acumen of the savvy data analyst cannot be replaced in all scenarios. Nonetheless, SAS (AFS) can at least be used as a "first pass" tool for large volumes of individual time series data that requires forecasts within a short period of time. Beyond that, the data analyst must be called upon to review the output and affect the necessary model-building whenever warranted.

References

[1] Montgomery, Douglas C. Author. Wiley Series in Probability and Statistics: Introduction to Time Series Analysis and Forecasting (2nd Edition), Wiley.

[2] SAS ® software

[3] MINITAB ® 17 Statistical Softare

[4] SAS/ETS ® 14.2 User's Guide: Overview of the Time Series Forecasting System