Comparison of Forecasting Approaches Using SAS Proc Reg and Proc Arima with SAS Macros vs. SAS Time Series Forecasting System[1]

Martin Selzer, Ph.D.
mselzer@att.net
Chatham Decision Sciences
Chatham, NJ

1. PURPOSE

This paper describes a SAS macro based program using Proc Reg and Proc Arima developed for forecasting. The program is compared to a separate forecasting process using the SAS Time Series Forecasting System (SASTSFS). Forecasts are prepared with the macro based program and compared to those produced by other programs and SASTSFS. Results show that the new program gives results quite near those generated by SASTSFS. While no effort was made to compare the accuracy of the two approaches against actual results, the mathematical formulation of the statistical models used in the macro based program suggest the program would provide more accurate goals. The macro based approach also has fewer manual interventions and produces forecasts more quickly and less prone to error.

2. PROGRAMS COMPARED

2.1. Macro Based Program

The Macro Based program developed uses the same basic models as the current forecasting process. These models are listed below:

- Linear Trend without Seasonal Terms
- Linear Trend with Seasonal Terms
- Simple Exponential Smoothing
- Double (Brown's) Exponential Smoothing

The difference between the macro based program and the process using SASTSFS is that the macro based program uses two SAS procedures in programmer written code to generate forecasts. These procedures are Proc Reg and Proc Arima. Proc Reg is used to produce the Linear Trend models. The Linear Trend with Seasonal Terms model uses Proc Reg and dummy variable regression with monthly zero-one dummy variables to produce forecasts. The Exponential Smoothing models use Proc Arima. Simple Exponential Smoothing is produced with a (0,1,1) Arima process. Double Exponential Smoothing uses a (0,2,2) process.

2.2. SASTSFS Process

The SAS Time Series Forecasting System process describe in this paper uses programs in two separate steps. The Step 1 program prepares the data for SASTSFS and the Step 2 program processes the output of this system.

---

[1] The SAS program with macros described in this paper can be downloaded at **urdobro.com.**

2.3. SASTSFS Process - Continued

The two steps involve multiple tasks in which the user inputs the data, chooses start dates, selects models, runs models fitted to historical data, and produces forecasts.[2]

While the SASTSFS Process used here at first appears complex, after repeated applications the user becomes familiar with the system. It does require a careful tracking of start dates for the data inputs and that all steps in the process are followed correctly. One handy feature of the system is that it allows graphing of historical observations and fitted values from models.

3. DATA and COMPARISON APPROACH

The data comes from an actual large manufacturer of five products sold through sales people in a number of distribution territories.

The two approaches were run and the outputs compared to evaluate if they gave similar results. This was done for various products with different number of historical periods. Each product evaluated had a certain number of distribution territories for which three month forecasts were prepared. A set of products was selected to evaluate different numbers of historical periods for modeling. Forecasts also were prepared using two approaches to make forecasts consistent with overall product goals provided by higher level managers. Territories' forecasts either were **aligned** or **adjusted** to over all product forecasts. Aligning forecasts modifies forecasts to be consistent with overall growth rates specified by managers. Adjusting forecasts makes the sum of all forecasts equal to amounts specified by managers. Tables 1 and 2 below show the forecast methods and products evaluated in the analysis.

Table 1
TEST RESULTS MACRO BASED PROGRAM VS SAS TIME SERIES FORECASTING SYSTEM

Products, Periods, and Adjustment Methods

24 Months		LT 24 and GT 12 Months		<= 12 Months	
Align	**Adjust**	**Align**	**Adjust**	**Align**	**Adjust**
Trend					
Prod1	Prod2	Prod3	Prod4	Prod5	Prod6
Trend with Seasonal Terms					
Prod1	Prod2	Prod3	Prod4	Prod5	Prod6
Single Exponential Smoothing or Arima(0,1,1)					
Prod1	Prod2	Prod3	Prod4	NA	NA
Double Exponential Smoothing or Arima(0,2,2)					
Prod1	Prod2	Prod3	Prod4	NA	NA

---

[2] The results reported in this paper also adjusted for variance in the data for a product across territories. This was done when deriving forecast using the Macro Based Program and SASTFS. The program on line does not include this feature.

Table 2
DATASETS USED IN TESTING OF MACRO
BASED PROGRAM AGAINST SAS TIME
SERIES FORECASTING SYSTEM

Prod	# of Distribution Terrs	# of Historical Time Periods
Prod1	79	24
Prod2	121	24
Prod3	79	18
Prod5	53	8
Prod6	496	7
Prod4	496	17

The measure used to compare forecasts for each territory from the two approaches is the absolute value of the sum of difference between the three forecasts from each approach divided by the sum of the forecasts using the SASFTS. This measure is shown mathematically below:

$$M = abs\left(\frac{\sum_{i=1}^{3}\left(F_{i,Macro} - F_{i,SASFTFS}\right)}{\sum_{i=1}^{3}F_{i,SASFTFS}} \right)$$

The minimum, maximum, average, and standard deviation of this measure across the territories for the products evaluated are reported. The analysis was done for the various products with different number of historical data periods, territories, and consistency methods with product goals.

4.   RESULTS

The results of the comparisons are shown in the table below:

Table 3
TEST RESULTS OF MACRO BASED PROGRAM
VS SAS TIME SERIES FORECASTING SYSTEM

Minimum, Maximum, Mean, and Standard Deviation of Sum of Differences in 3 Month Forecasts Divided by SASFTFS Forecast

Measure	24 Months		<24 >12 Months		< 12 Months	
	Align	Adjust	Align	Adjust	Align	Adjust
Trend						
Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Max	0.0047	0.0432	0.0047	0.1016	0.0664	0.0000
Mean	0.0011	0.0016	0.0011	0.0123	0.0129	0.0000
Std Dev	0.0008	0.0039	0.0008	0.0118	0.0122	0.0000
Trend with Seasonal Terms						
Min	0.0000	0.0000	0.0000	0.0000	NA	
Max	0.0035	0.0140	0.0047	0.2058		
Mean	0.0011	0.0033	0.0011 (see Note 1)	0.0442		
Std Dev	0.0008	0.0050	0.0008	0.0351		
Single Exponential Smoothing or Arima(0,1,1)						
Min	0.0006	0.0000	0.0001	0.0000	0.0033	0.0000
Max	0.1482	0.0720	0.1648	1.2422	0.5274	0.5267
Mean	0.0337	0.0187	0.0434	0.1019	0.1315	0.0588
Std Dev	0.0304	0.0135	0.0339	0.0979	0.1019	0.0575
Double Exponential Smoothing or Arima(0,2,2)						
Min	0.0000	0.0000	0.0003	0.0000	see Note 2	NA
Max	0.2266	0.0786	0.2342	1.5967		
Mean	0.0352	0.0195	0.0521	0.1154		
Std Dev	0.0324	0.0154	0.0448	0.1271		

Note 1: For Prod3 the Trend and Trend with Seasonal Terms results were the same. This product had 18 historical data months. Apparently, there is negligible seasonality for this product in the last forecast months where seasonality can be evaluated.

Note 2: For Prod5 there were 8 historical months. In the SAS Forecasting System, double exponential smoothing gave many odd forecasts with many of them negative. Clearly, for data sets with few historical values the model program gives superior results for this model as compared to the SAS Forecasting System.

4. RESULTS - Continued

For the Trend with Seasonal Terms and Trend Model, both models gave similar results. The average comparison measure, M, across territories for these models was almost always less than .01 or 1%. The Exponential Smoothing methods results were more different, most likely because different algorithms were used from the SAS Analysis Systems as compared to Proc Arima. This is understood because for Simple Exponential Smoothing the SAS Analysis system gave the same forecast for each of the three forecast periods. The Proc Arima approach gave different forecasts because it uses maximum likely methods to generate starting smoothing predictions and weights. It also uses an iterative process to generate forecasts.

5. ADVANTAGES AND DISADVANTAGES OF NEW PROGRAMMING METHODOLOGY

5.1. Advantages

The new programming methodology described has advantages over the forecasting methodology using SASTSFS.

- Less processing time to produce final outputs
- More automated processing less prone to error
- Larger files that can be processed in batch mode with less manual intervention
- Better forecasting algorithms for exponential smoothing based on Arima methods which can result in a reduction of revisions[3]
- Consistency of programming across analysts
- Standardization of file inputs across brands and programmers

5.2. Disadvantages

The basic disadvantage of the macro based programming methodology is that it involves a complex SAS program with various macro variables and macro calls. It also does not allow intermediary graphing that SASTSFS provides. With the data described in this analysis, analysts currently use a set of programs that involves using SASTSFS with multiple steps and output of final programs. Although these programs are simpler, they require manual interventions in SAS and MS Excel for formatting and involve more processing time.

---

[3] The SAS Time Series Forecasting System does have the option of running ARIMA (0,1,1) and (0,2,2) models. The current analysis did not test the results of these models with those of the new programming methodology. When forecasters use the SAS Time Series Forecasting System for Exponential Smoothing, they should use the ARIMA approach because this approach gives more accurate forecasts since it uses maximum likelihood estimation of parameters and starting forecast values.

6. SUMMARY

A macro based programming approach to produce goals for brands across territories was described and compared to programs and processes using the SAS Time Series Forecasting System. The macro based program uses SAS macros, Proc Reg, and Proc Arima to produce forecasts. The approach using the SAS Time Series Forecasting System is a two step process that relies on SAS Time Series Forecasting options to produce forecasts.

A comparison was made of the two approaches and results showed the two processes produced very close results for trending with and without seasonal terms. Exponential smoothing algorithms gave more different results. The estimation method used in the Macro Based program suggests that the forecasts from exponential smoothing are more accurate with this program