Ex-Ante Forecast Model Performance with Rolling Simulations

Michael Leonard, Ashwini Dixit, Udo Sglavo, SAS Institute Inc.

ABSTRACT

Given a time series data set, you can use automatic time series modeling software to select an appropriate time series model. You can use various statistics to judge how well each candidate model fits the data (in-sample). Likewise, you can use various statistics to select an appropriate model from a list of candidate models (in-sample or out-of-sample or both). Finally, you can use rolling simulations to evaluate ex-ante forecast performance over several forecast origins. This paper demonstrates how you can use SAS® Forecast Server Procedures and SAS® Forecast Studio software to perform the statistical analyses that are related to rolling simulations.

INTRODUCTION

Before you use a time series model to forecast a time series, it is important that you evaluate the model's performance by validating the model's ability to forecast the most recently acquired data. Such performance measures are called *ex-ante* (before the fact) model performance measures. Ex-ante model performance measures are similar to ex-post (after the fact) forecast performance measures; but ex-post forecast performance measures evaluate the performance of forecasts regardless of their source (model-based, judgment-based, possible adjustments, and so on). *Rolling simulations* enable you to perform ex-ante model performance by repeating the analyses over several forecast origins and lead times.

The concept of rolling simulations is complicated, but SAS Forecast Server makes it easy to understand. This paper describes how you can use rolling simulations in SAS Forecast Server Procedures and SAS Forecast Studio to measure ex-ante model performance. This paper focuses on ex-ante model performance of reasonably long time series. This paper is less applicable to short times series and new product forecasting.

BACKGROUND

You can find introductory discussions about time series and automatic forecasting in Makridakis, Wheelwright, and Hyndman (1997); Brockwell and Davis (1996); and Chatfield (2000). You can find a more detailed discussion of time series analysis and forecasting in Box, Jenkins, and Reinsel (1994); Hamilton (1994); Fuller (1995); and Harvey (1994). You can find a more detailed discussion of large-scale automatic forecasting in Leonard (2002) and a more detailed discussion of large-scale automatic forecasting systems with input and calendar events in Leonard (2004).

TIME SERIES MATHEMATICAL NOTATION

This section explains the mathematical notation that is used to describe rolling simulations. If you already understand the mathematical notation that is used in most time series analysis, you might choose to skip this section and just look at the SAS Forecast Server sections that follow it.

TIME INDEXING

- Rolling simulations use the following time indices:Let t = 1, …, T represent a (discrete) time index, where T represents the length of the historical time series.
- Let l = 1, …, L, represent the index of future time periods (which are predicted by forecasting), where L represents the forecast horizon (also called the lead).
- Let h = 1, …, H represent the holdout region index, where h represents the length of out-of-sample selection region (also called the holdout region). You use this index when you want to use out-of-sample evaluation techniques to select time series models. When H = 0, in-sample fit selection is used.
- Let b = 1, …, B represent the performance region index, where B represents the length of the out-of-sample performance region (also called the holdback region). When B = 0, in-sample performance analysis is used.

Various combinations of $T > 0$, $B \geq 0$, $H \geq 0$, and $L \geq 0$ produce results. When $B > 0$, B is usually equal to L, but this is not necessarily so.

Figure 1 illustrates how the time series indices can divide the time line.

Time Regions (H > 0, B > 0)

Historical Time Series (1,...,T)		
In-Sample Region (1,...,T − B)		Out-Sample Region (T − B + 1,..., T)
Initial Fit Region (1,...,T − B − H)	Selection Region (T − B − H + 1,...,T − B)	Performance Region (T − B + 1,..., T)
Fit Region (1,...,T − B)		Forecast Horizon (T − B + 1,..., T − B + L)

Figure 1: Time Indices Divide the Time Line. Figure 2 illustrates how the time series indices can divide an example time series data set.
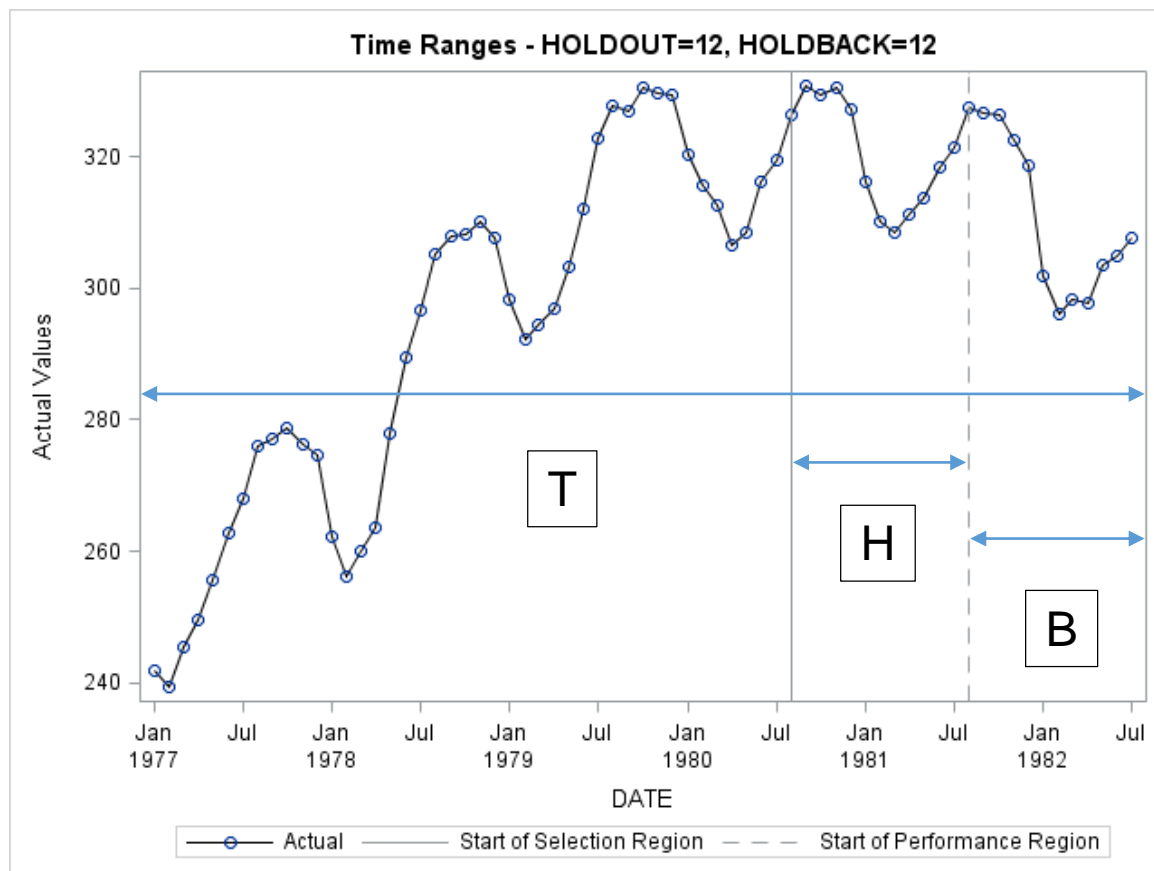

Figure 2: Time Indices Divide a Time Series Data Set

TIME SERIES DATA

Let y_t represent a continuous time series value at time index t. Let $Y_T = \{y_t\}_{t=1}^{T}$ represent the *dependent series* (the historical time series vector) that you want to model and forecast.

The historical time series data can also contain *independent series*, such as inputs and calendar events that help model and forecast the dependent series. Let $\vec{X}_T = \{\vec{x}_t\}_{t=1}^{T+L}$ represent the historical and future predictor series vectors.

Although this paper focuses only on extrapolation techniques, the concepts generalize to more complicated time series models.

MODEL INDICES

You often analyze time series data by comparing different time series models. Automatic time series modeling selects among several competing models or uses combinations of models. Let $F_m(\ )$ represent the *m*th time series model, where *m* = 1, …, *M* and *M* represents the number of candidate models under consideration.

THEORETICAL MODEL FOR ANY TIME SERIES

You can use a theoretical time series model to model the historical time series values of any time series. Given any historical time series data, $Y_T = \{y_t\}_{t=1}^{T}$ and $\vec{X}_T = \{\vec{x}_t\}_{t=1}^{T+L}$, you can construct a time series model to generate predictions of future time periods. Let $y_{t+l|t}^{(m)} = F_m\left(Y_t, \vec{X}_{t+l} : \theta_m\right)$ represent the theoretical time series model, where $y_{t+l|t}^{(m)}$ represents the prediction for y_{t+l} that uses only the historical time series data that end at time index t, for the *l*th future value for the *m*th time series model, $F_m(\ )$, and where θ_m represents the parameter vector to be estimated. The theoretical model can also be a combination of two or more models.

FITTED MODEL FOR A SPECIFIC TIME SERIES

You can use a fitted time series model to predict the future time series values of a specific time series. Given specific historical time series data, $Y_T = \{y_t\}_{t=1}^{T}$ and $\vec{X}_T = \{\vec{x}_t\}_{t=1}^{T+L}$, and a theoretical time series model, $y_{t+l|t}^{(m)} = F_m\left(Y_t, \vec{X}_{t+l} : \theta_m\right)$, you can estimate a fitted model to predict future time periods. Let $\hat{y}_{t+l|t}^{(m)} = F_m\left(Y_t, \vec{X}_{t+l} : \hat{\theta}_m\right)$ represent the fitted model, where $\hat{y}_{t+l|t}^{(m)}$ represents the prediction for y_{t+l} and $\hat{\theta}_m$ represents the estimated parameter vector.

MODEL PREDICTIONS

You can use a fitted time series model to generate model predictions. Let $\hat{y}_{t|t-l}^{(m)} = F_m\left(Y_t, \vec{X}_{t+l} : \hat{\theta}_m\right)$ represent the model predictions for the *m*th time series model for the *t*th time period that uses the historical time series data up to the *(t – l)* time period. Let $\hat{Y}_T^{(m)} = \left\{\hat{y}_{t|t-l}^{(m)}\right\}_{t=1}^{T}$ represent the prediction vector.

For *l* = 1, let $\hat{y}_{t|t-1}^{(m)}$ represent the *one-step-ahead prediction* for the *t*th time period and for the *m*th time series model. The one-step-ahead predictions are used for *in-sample* evaluation. They are commonly used for model fit evaluation, in-sample model selection, and in-sample performance analysis.

For *l* ≥ 1, let $\hat{y}_{t|t-l}^{(m)}$ represent the *multi-step-ahead prediction* for the *t*th time period and for the *m*th time series model. The multi-step-ahead predictions are used for *out-of-sample* evaluation. They are commonly used for holdout selection, performance analysis, and forecasting.

PREDICTION ERRORS

There are many ways to compare predictions that are associated with time series data. These techniques usually

compare the historical time series values, y_t, with their time series model predictions, $\hat{y}_{t|t-l}^{(m)} = F_m\left(Y_t, \vec{X}_{t+l} : \hat{\theta}_m\right)$.

Let $\hat{e}_{t|t-l}^{(m)} = y_t - \hat{y}_{t|t-l}^{(m)}$ represent the *prediction error* for the *m*th time series model for the *t*th time period that uses the historical time series data up to the *(t – l)*th time period.

For *l* = 1, let $\hat{e}_{t|t-1}^{(m)} = y_t - \hat{y}_{t|t-1}^{(m)}$ represent the *one-step-ahead prediction error* for the *t*th time period and for the *m*th time series model. The one-step-ahead prediction errors are used for *in-sample* evaluation. They are commonly used for evaluating model fit, in-sample model selection, and in-sample performance analysis.

For *l* > 1, let $\hat{e}_{t|t-l}^{(m)} = y_t - \hat{y}_{t|t-l}^{(m)}$ represent the *multi-step-ahead prediction error* for the *t*th time period and for the *m*th time series model. The multi-step-ahead prediction errors are used for *out-of-sample* evaluation. They are commonly used for holdout selection and performance analysis.

STATISTICS OF FIT

A statistic of fit measures how well the predictions match the actual time series data. (A statistic of fit is sometimes called a goodness of fit.) There are many statistics of fit: RMSE, MAPE, MASE, AIC, and so on. Given a time series vector, $Y_T = \left\{y_t\right\}_{t=1}^{T}$, and its associated prediction vector, $\hat{Y}_T^{(m)} = \left\{\hat{y}_{t|t-l}^{(m)}\right\}_{t=1}^{T}$, let $SOF\left(Y_T, \hat{Y}_T^{(m)}\right)$ represent the relevant statistic of fit. The statistic of fit can be based on the one-step ahead or multi-step-ahead predictions.

For $Y_T = \left\{y_t\right\}_{t=1}^{T}$ and $\hat{Y}_T^{(m)} = \left\{\hat{y}_{t|t-l}^{(m)}\right\}_{t=1}^{T}$, let $SOF\left(Y_T, \hat{Y}_T^{(m)}\right)$ represent the *fit statistics* for the *m*th time series model. The one-step-ahead prediction errors are used for *in-sample* evaluation. They are commonly used for model fit evaluation, in-sample model selection, and in-sample performance analysis.

For $Y_H = \left\{y_t\right\}_{t=T-H+1}^{T}$ and $\hat{Y}_H^{(m)} = \left\{\hat{y}_{t|T-H}^{(m)}\right\}_{t=T-H+1}^{T}$, let $SOF\left(Y_H, \hat{Y}_H^{(m)}\right)$ represent the *performance statistics* for the *m*th time series model. The multi-step-ahead prediction errors are used for *out-of-sample* evaluation. They are commonly used for holdout selection and performance analysis.

MODEL SELECTION LIST

No one modeling technique works for all time series data. A model selection list contains a list of candidate models and describes how to select among the models. Let $\left\{F_m(\ )\right\}_{m=1}^{M}$ represent the model selection list.

The model selection list is subsetted by using the selection diagnostics (intermittency, trend, seasonality, and so on) and a model selection criterion (such as RMSE, MAPE, and AIC). Let $F_*(\ ) \in \left\{F_m(\ )\right\}_{m=1}^{M}$ represent the selected model.

After a model is selected, the selection region is dropped. The model is refitted to all the data except the performance region. See Figure 1 and Figure 2 for a graphical description.

MODEL GENERATION

The list of candidate models can be specified by the analyst or automatically generated by the series diagnostics (or both). The diagnostics consist of various analyses that are related to intermittency, trend, seasonality, autocorrelation, and, when predictor variables are specified, cross-correlation analysis.

The *series diagnostics* are used to generate models, whereas the *selection diagnostics* are used to select models.

ROLLING SIMULATIONS

Rolling simulations enable you to evaluate ex-ante forecast performance across forecast origins.

For each holdback index, *b* = 1, …, *B*, and each lead index, *l* = 1, …, *b*, let $\hat{e}_{t-b+l|t-b}^{(m)} = y_{t-b+l} - \hat{y}_{t-b+l|t-b}^{(m)}$ represent the *multi-step-ahead prediction error* for the *(t – b+l)* time period and for the *m*th time series model. The holdback index, *b*, rolls the forecast origin forward in time. For each holdback index, *b* = 1, …, *B*, multi-step-ahead prediction errors are computed for each lead index, *l* = 1, …, *b*.

By gathering the prediction errors by the lead index, *l*, for all holdback indices, *b* = 1, …, *B*, you can evaluate forecast

performance by lead time. In other words, you can evaluate a model's *l*-step-ahead forecast performance across various forecast origins.

You can compute the lead performance statistics for a particular lead time, *l*. For $Y_l = \{y_{t-b+l}\}_{b=1}^{B}$ and

$\hat{Y}_l^{(m)} = \{\hat{y}_{T-b+l|T-b}^{(m)}\}_{b=1}^{B}$, let $SOF(Y_l, \hat{Y}_l^{(m)})$ represent the *lead performance statistics* for the *m*th time series model for the lead index, *l*.

Because the rolling simulations are specific to each model under consideration, you can "test drive" each model before using it.

SAS FORECAST SERVER PROCEDURES IMPLEMENTATION

You can use the HPFENGINE procedure, a procedure in the SAS Forecast Server Procedures product to perform rolling simulations. You can use the BACK= option in tandem with the LEAD= option to generate the predictions you need to compute the lead performance statistics. The following statements illustrate how you can use the HPFENGINE procedure to perform rolling simulations for holdback indices and lead time indices 1, 2, and 3.

```
proc hpfengine data=sashelp.air out=_NULL_ outfor=outfor1 LEAD=1 BACK=1;
id date interval=month;
forecast air;
run;

proc hpfengine data=sashelp.air out=_NULL_ outfor=outfor2 LEAD=2 BACK=2;
id date interval=month;
forecast air;
run;

proc hpfengine data=sashelp.air out=_NULL_ outfor=outfor3 LEAD=3 BACK=3;
id date interval=month;
forecast air;
run;

data rolling;
merge sashelp.air(firstobs=142)
outfor1(keep=date predict rename=predict=Back1 firstobs=144)
outfor2(keep=date predict rename=predict=Back2 firstobs=143)
outfor3(keep=date predict rename=predict=Back3 firstobs=142);
by date;
run;

proc print data=rolling; run;
```

Notice that the LEAD= and BACK= options vary with each successive PROC HPFENGINE statement. Executing this code produces printed output that is shown in Table 1.

Obs	DATE	AIR	Back1	Back2	Back3
1	OCT60	461	.	.	443.171
2	NOV60	390	.	393.431	388.404
3	DEC60	432	432.288	433.458	427.905

Table 1: Rolling Simulations for Lead Time (1, 2, and 3)

The first diagonal in Table 1 represents lead index 1 (the one-step-ahead predictions), the second diagonal in the table represents lead index 2 (the two-step-ahead predictions), and the third diagonal in the table represents lead

index 3 (the three-step-ahead predictions). By comparing the actual values (AIR) to the multi-step-ahead predictions, you can compute the multi-step-ahead prediction errors for each lead index (1, 2, and 3). You can compute various statistics from these comparisons to measure ex-ante forecast performance.

SAS FORECAST STUDIO IMPLEMENTATION

This section shows how you can use SAS Forecast Studio to create hierarchical time series forecasts and perform rolling simulations in a hierarchical context. Much more complicated analyses are possible.

HIERARCHICAL PROJECT CREATION IN FORECAST STUDIO

First, you create a hierarchical project, as shown in the following steps:

Step 1   Select the forecasting environment, specify the project name, and optionally provide a project description.
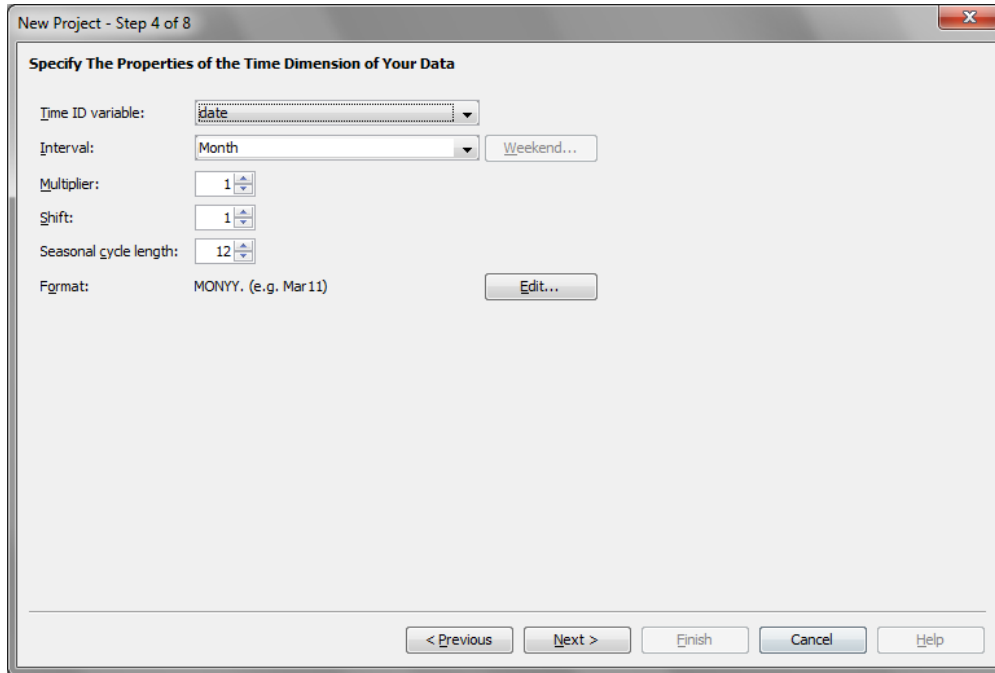
Step 2   From the **Name** list, select the input time series data set that you want to forecast.
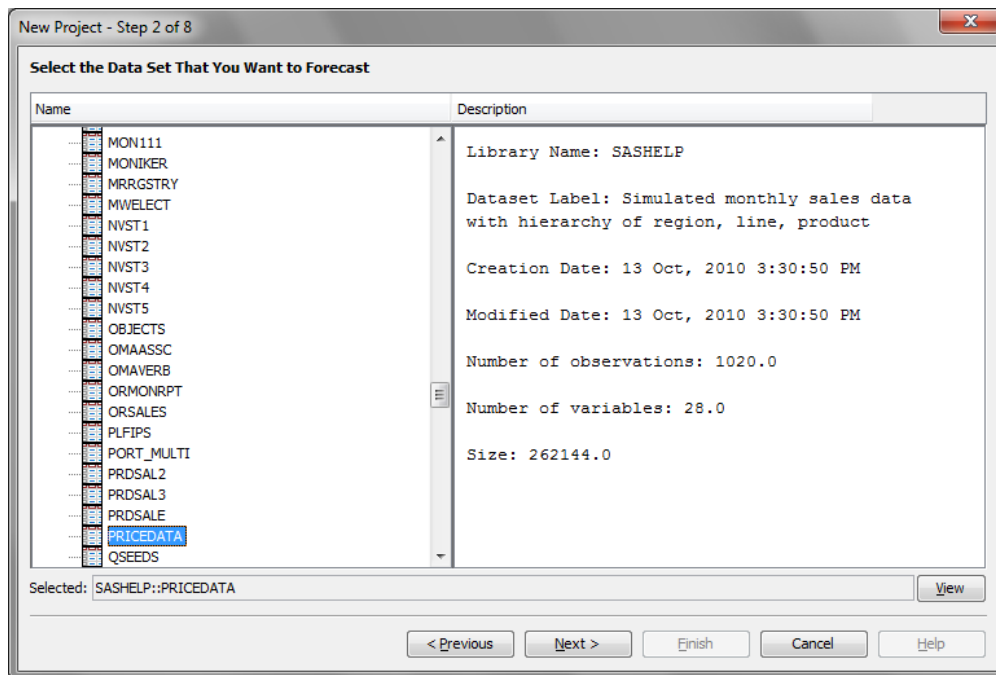
Step 3   From the **Available variables** list, select one or more variables that you want to use as classification (BY) variables, and click the right arrow to move them to the **Classification (BY) variables selected** list. Then use the up and down arrows to order them, and use the checkboxes to indicate whether to forecast a hierarchy.
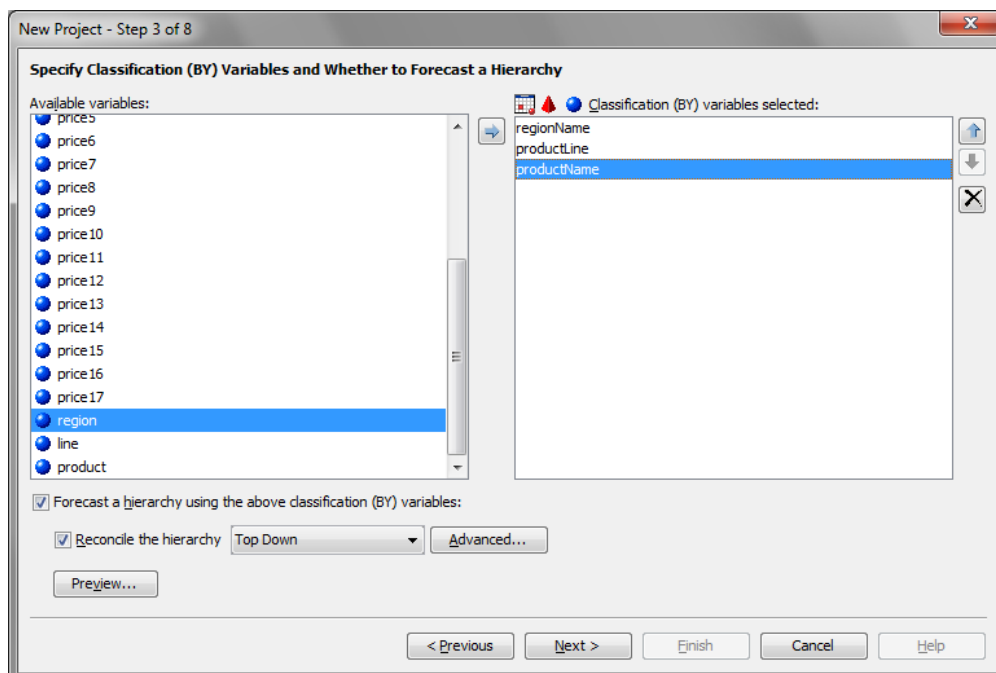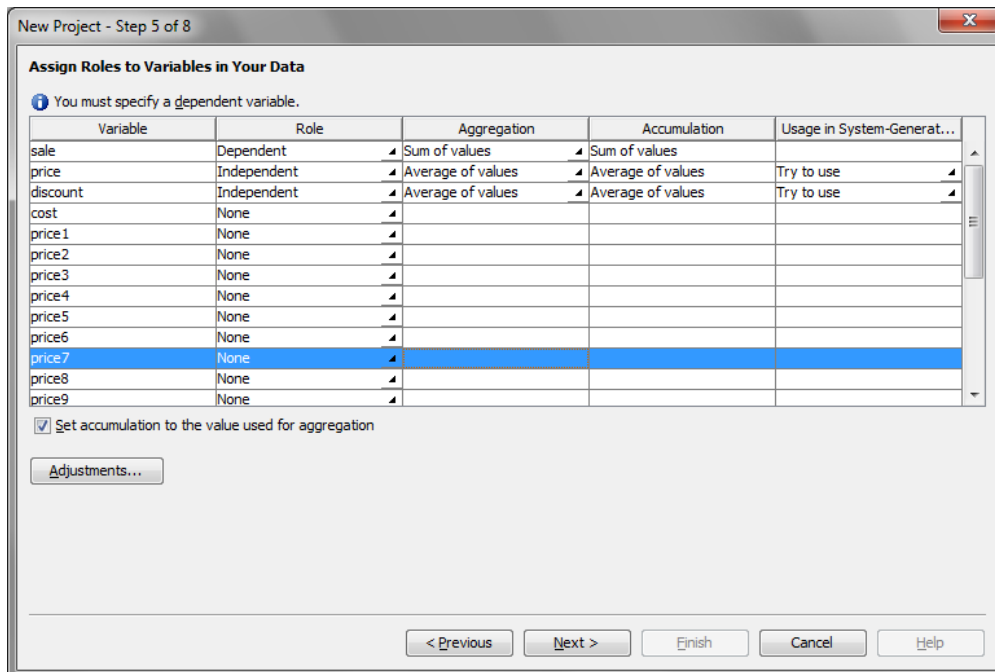
Step 4 Select the Time ID variable, and specify properties of the time dimension of the data.

Step 5 Assign roles to variables in the data. Select the dependent variable, independent variables, adjustment, and reporting variables.

Step 6 Specify the data preparation options.

Step 7 Specify the number of future periods to forecast (the forecast horizon).

New Project - Step 7 of 8

Finish Creating Project

The project is now set up and forecasts can be produced. Before producing forecasts, there are several optional steps that you may take:

Change the number of periods to forecast(horizon): [12]

Change other forecasting settings
Create and import events

< Previous   Next >   Finish   Cancel   Help

Step 8 Finish the wizard to produce forecasts of the hierarchical data to be displayed in SAS Forecast Studio. Default models are created by the Forecast Server Procedures for each node in the hierarchy that you defined.

New Project - Step 8 of 8

Optional Project Actions

Select action to perform on pressing "Finish":

⦿ Produce forecasts and display in SAS Forecast Studio. For each series the following steps will be performed:

-Diagnose: The series will be analyzed for features such as trends, seasonality, and cyclicity, and system-generated models appropriate to the series will be created.If you have specified independent variables or events, these will be taken into account for system-generated ARIMA and UCM models. They will be used in models if they improve performance.

-Select: The system-generated models, along with any other models you have included, will be evaluated against the data and the best-performing model selected as the forecast model.

-Fit: A final set of parameters of the forecast model will be estimated.

-Forecast: The forecast model will be used to produce a forecast.

After forecasts have been produced for each series, the hierarchy will be reconciled.

◯ Save SAS code to produce forecasts later

< Previous   Next >   Finish   Cancel   Help

Step 9 After your Forecast Studio project is created, the Forecast Summary dialog box appears. Close this dialog box.
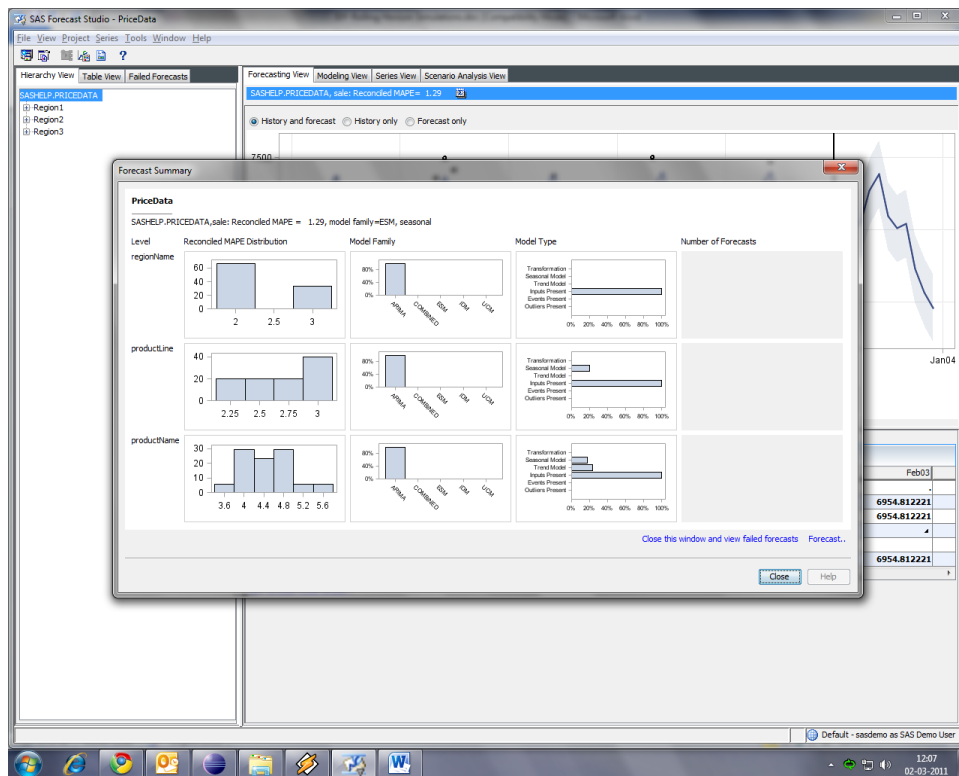
Step 10 Select **Modeling View** to display the list of automatically generated models for the root node in the hierarchy.
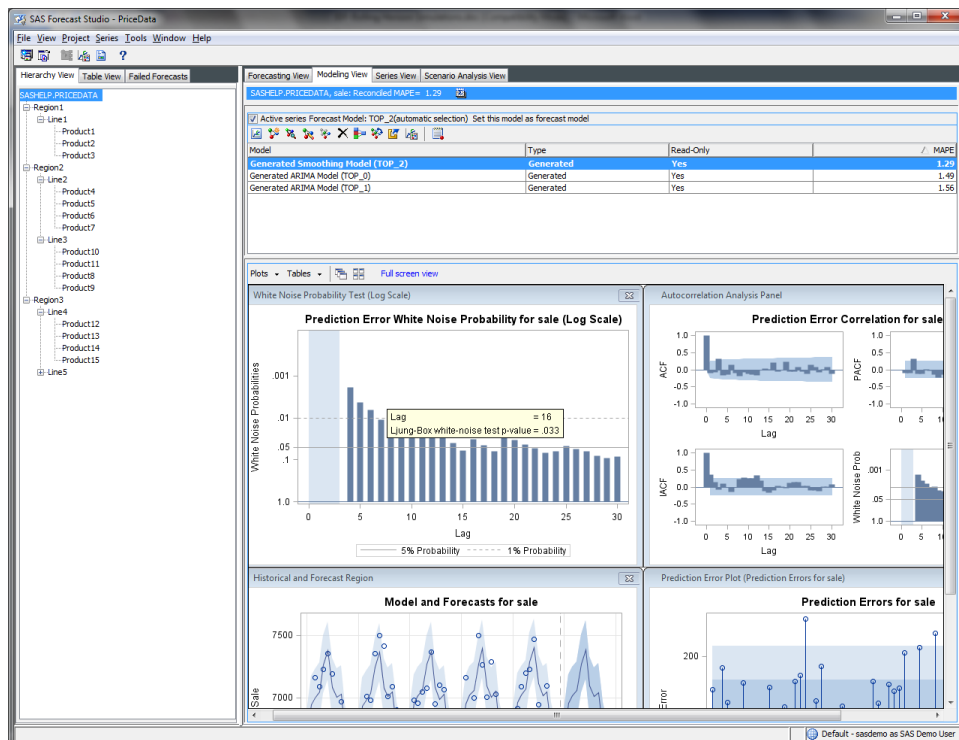
ROLLING HORIZON SIMULATIONS TO EVALUATE THE EX-ANTE FORECAST PERFORMANCE OF THE MODEL

After you have created a project, you can use rolling simulations analysis to evaluate the ex-ante forecast performance of a model in order to better evaluate future model performance.

Step 1   In the **Model** table select the model for which you want to evaluate performance.
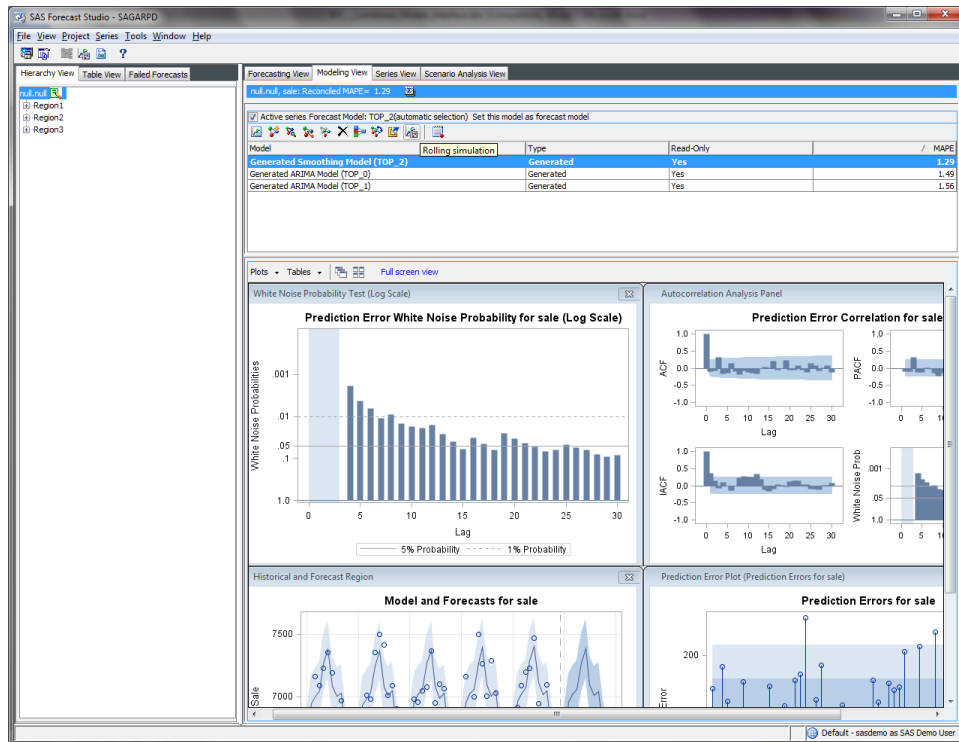
Step 2 In the Rolling Simulations dialog box, you can optionally specify the maximum number of out-of-sample observations for the simulation. On the Simulations tab, you can graphically view the predictions of the model at various forecast origins and compare them to the actual values. Note that Table 1 is the transposed form of the table shown in the Rolling Simulations dialog box. The bold numbers in the table represent out-of-sample predictions.
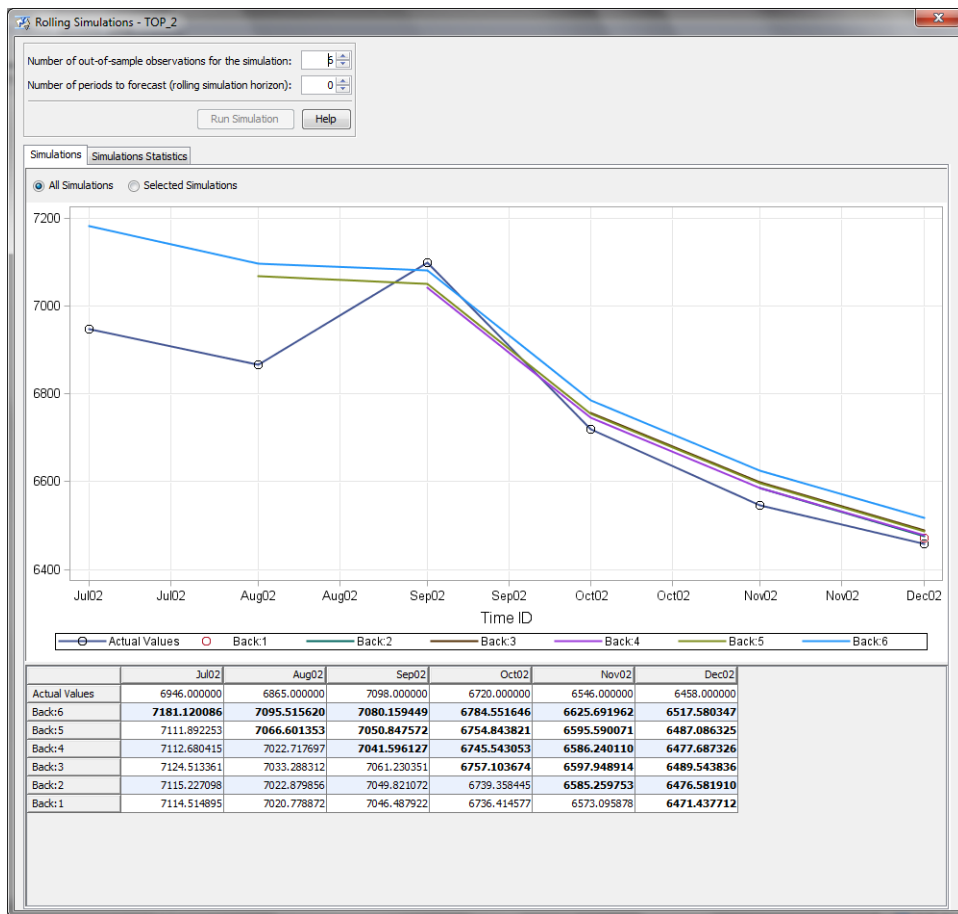
Step 3   In the Rolling Simulations dialog box, you can optionally specify the number of out-of-sample observations for the simulation and the number of periods to forecast (rolling simulation horizon). On the Simulations tab, you can graphically view the predictions of the model at various forecast origins and compare them to the actual values. In this step, the number of periods to forecast is 3.

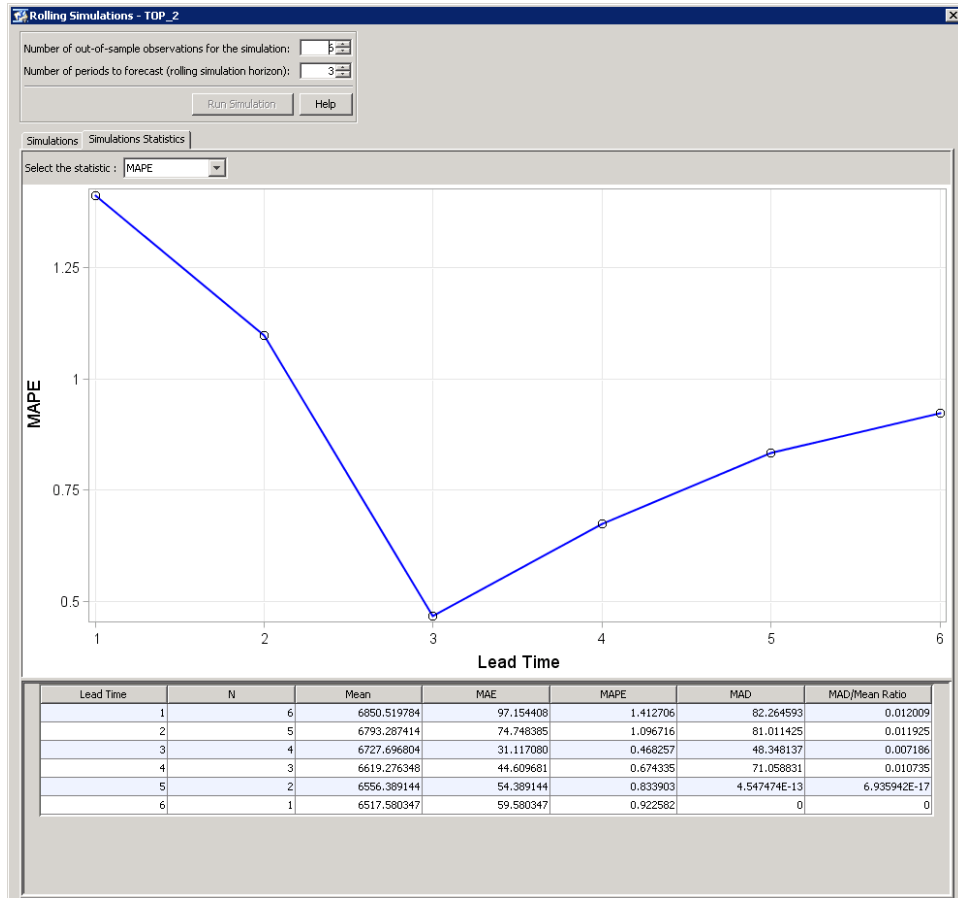|	Jul02	Aug02	Sep02	Oct02	Nov02	Dec02	Jan03	Feb03
Actual Values	6946.000000	6865.000000	7098.000000	6720.000000	6546.000000	6458.000000		
Back:6	7181.120086	7095.515620	7080.159449	6784.551646	6625.691962	6517.580347	6734.556447	7009.539780
Back:5	7111.892253	7066.601353	7050.847572	6754.843821	6595.590071	6487.086325	6701.357996	6975.954304
Back:4	7112.680415	7022.717697	7041.596127	6745.543053	6586.240110	6477.687326	6689.835674	6964.383417
Back:3	7124.513361	7033.288312	7061.230351	6757.103674	6597.948914	6489.543836	6702.263650	6976.957689
Back:2	7115.227098	7022.879856	7049.821072	6739.358445	6585.259753	6476.581910	6688.786488	6963.210362
Back:1	7114.514895	7020.778872	7046.487922	6736.414577	6573.095878	6471.437712	6683.083554	6957.358289

Step 4 On the Simulation Statistics tab, you can analyze various performance measures with respect to the lead time. In this step, MAPE is the selected statistic.

Lead Time	N	Mean	MAE	MAPE	MAD	MAD/Mean Ratio
1	6	6850.519784	97.154408	1.412706	82.264593	0.012009
2	5	6793.287414	74.748385	1.096716	81.011425	0.011925
3	4	6727.696804	31.117080	0.468257	48.348137	0.007186
4	3	6619.276348	44.609681	0.674335	71.058831	0.010735
5	2	6556.389144	54.389144	0.833903	4.547474E-13	6.935942E-17
6	1	6517.580347	59.580347	0.922582	0	0

Step 5   On the Simulations tab, you can switch the mode to graphically view the selected predictions of the model at various forecast horizons and compare them to the actual values.
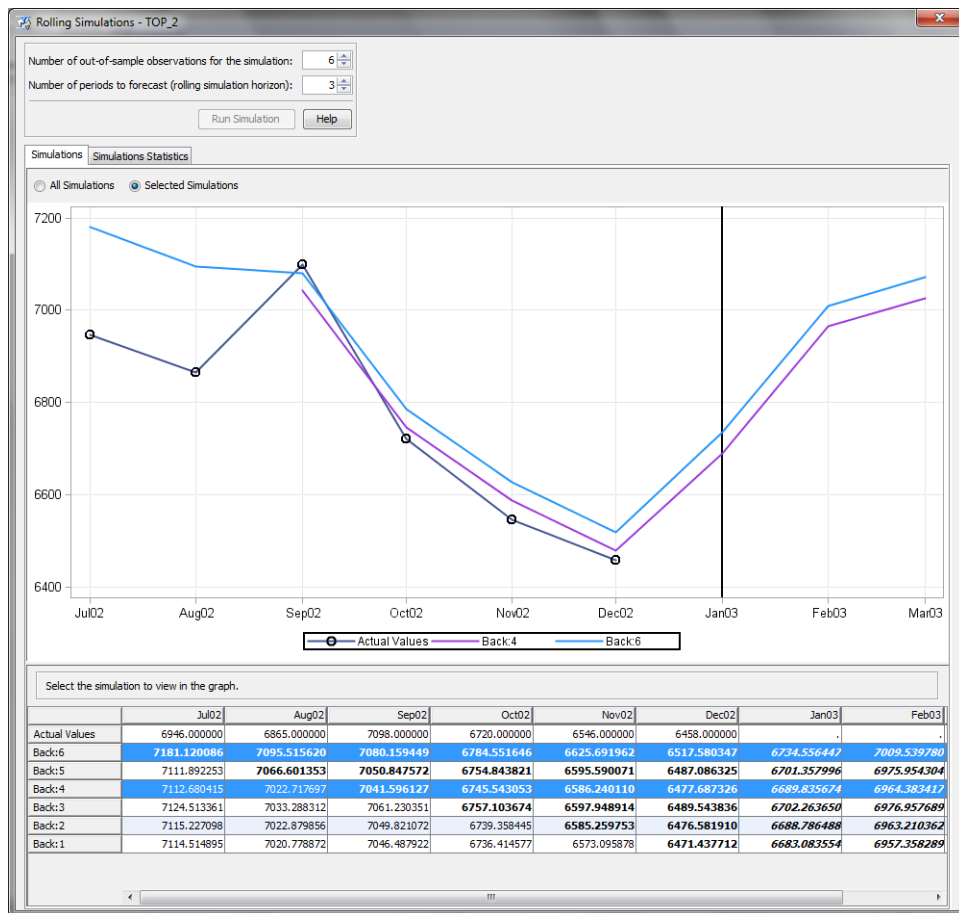
CONCLUSION

This paper illustrates the various statistics that are associated with time series models and with an automatic time series model selection process, and shows how you can use SAS Forecast Server Procedure and SAS Forecast Studio software to compute them. In particular, it shows how you can use rolling simulations to evaluate ex-ante forecast performance. References

Box, G. E. P., G. M.Jenkins, and G. C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Brockwell, P. J., and R. A. Davis. 1996. *Introduction to Time Series and Forecasting*. New York: Springer-Verlag.

Chatfield, C. 2000. *Time Series Models*. Boca Raton, FL: Chapman & Hall/CRC.

Fuller, W. A. 1995. *Introduction to Statistical Time Series*. New York: John Wiley & Sons, Inc.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Harvey, A. C. 1994. *Time Series Models*. Cambridge, MA: MIT Press.

Leonard, M. J. 2002. "Large-Scale Automatic Forecasting: Millions of Forecasts." International Symposium of Forecasting. Dublin.

Leonard, M. J. 2004. "Large-Scale Automatic Forecasting with Calendar Events and Inputs." International Symposium of Forecasting. Sydney.

Makridakis, S. G., S. C. Wheelwright, and R. J. Hyndman. 1997. *Forecasting: Methods and Applications*. New York: Wiley.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:	Michael Leonard
Organization:	SAS
Address:	SAS Campus Drive
City,	State ZIP: 27513
Work Phone:	919-531-6967
Fax:	
Email:	Michael.Leonard@sas.com
Web:	