

Paper 329-2011

Extending SAS[®] High-Performance Forecasting Using User-Defined Models and External Forecasts

Udo Sglavo, SAS Institute Inc., Cary, NC

ABSTRACT

For automatic forecasting of large numbers of time series, SAS[®] High-Performance Forecasting (part of SAS[®] Forecast Server) provides access to the most robust models available today. However as an experienced forecaster, you might have the desire to extend the model families provided. SAS High-Performance Forecasting allows you to extend the models in three ways:

- Custom repositories: You can create your own model repository.
- External models: Your forecasts are provided by methods that are external to the system.
- User-defined models: You are adding forecasting methods that are not provided by SAS High-Performance Forecasting.

In this paper, the use of custom repositories, external models, and user-defined models in SAS High-Performance Forecasting will be discussed, and four easy-to-follow examples will be provided.

LARGE-SCALE AUTOMATIC FORECASTING

The scale of business forecasting problems is increasing. One example is in retail, where more rigorous forecasting and planning processes and statistical tools are beginning to augment the “art” of merchandising. Retailers wish to know what products will sell at which price points, which promotions will be the most effective, and what is the best clearance strategy when a product is out of season. These questions all have a basis in forecasting.

Given the relatively small number of forecasters in most organizations, a large degree of automation is often required in order to complete the forecasting process in the time available in each planning period. SAS[®] Forecast Server provides large-scale automatic forecasting coupled with analytical sophistication which makes the forecasting process more manageable.

SAS Forecast Server can be accessed in different ways depending on your forecasting skills and SAS knowledge. SAS Forecast Server includes a Java-based graphical user interface (called SAS Forecast Studio) and a programmable batch environment (SAS High-Performance Forecasting) that give you the ability to produce and automate the forecasting process in the way that suits you best.

AUTOMATIC FORECASTING INFORMATION

In his SAS white paper “Large-Scale Automatic Forecasting Using Inputs and Calendar Events,” author Michael Leonard discusses an automatic forecasting information flow.

To automatically forecast a single time series, the time series model must be diagnosed, selected, and specified to obtain a selected model used for forecasting. These abstract statistical concepts must be made concrete and persist in a computer’s storage.

According to the author, SAS High-Performance Forecasting is using the following information to define its model repositories:

MODEL SPECIFICATION

A model specification is a concrete implementation of the abstract concepts related to a theoretical statistical time series model. Given a time series and a model specification, a forecast for the time series can be generated by applying the abstract statistical concepts associated with the model specification.

MODEL SELECTION LIST

A model selection list specifies a list of candidate model specifications and how to choose which model specification is best suited to forecast a particular time series. Given a time series and an appropriate model selection list, a forecasting model can be automatically selected for the time series. Because the model selection process is applied to each series individually, the process may select a different model for different series and may select a different model for a given time series with the passage of time as more data is collected.

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

SAS HIGH-PERFORMANCE FORECASTING

You can consider SAS High-Performance Forecasting as the forecasting engine of SAS Forecast Server. If you prefer, SAS Forecast Server automatically selects the appropriate model for each item being forecast. You can choose the desired level of automation to balance both speed and accuracy objectives.

There are numerous types of forecasting models that you can use. For automatic forecasting of large numbers of time series, SAS High-Performance Forecasting provides you with access to most the robust forecasting model techniques. The goal is not to have you manually choose the very best model for forecasting each time series. The goal is to provide you a list of candidate models that will forecast the large majority of the time series well.

HPFDIAGNOSE PROCEDURE

SAS High-Performance Forecasting provides a comprehensive diagnostic engine (called the HPFDIAGNOSE procedure) to automatically identify a univariate time series model. This procedure automatically diagnoses the statistical characteristics of each time series and identifies appropriate models. The models that the HPFDIAGNOSE procedure considers for each time series include

- autoregressive integrated moving average with exogenous inputs (ARIMAX) models
- exponential smoothing models
- unobserved components models
- intermittent demand models

Log transformation and stationarity tests are automatically performed. The ARIMAX model diagnostics find the autoregressive (AR) and moving average (MA) orders, detect outliers, and select the best input variables. The unobserved components model (UCM) diagnostics find the best components and select the best input variables.

As a result of running the HPFDIAGNOSE procedure, model specifications for each model family and a model selection list are created and stored in an automatic model repository.

CUSTOM MODEL REPOSITORIES

In addition to automatic model repository created by the HPFDIAGNOSE procedure, you can create your own model repository using the HPFxxxSPEC (where xxx stands for ARIMA, UCM, ESM or IDM) and the HPFSELECT procedures. Your custom model repository stores information about time series models (model specification) and how to select an appropriate time series model (model selection list) when given a particular time series. You are in full control of formulating the models and setting specific options on the model selection process (such as specifying the selection criteria).

EXTERNAL AND USER-DEFINED MODELS

In addition to the previously described general classes of exponential smoothing models (ESM), unobserved component models (UCM), autoregressive integrated moving average with exogenous inputs models (ARIMAX), and intermittent demand models (IDM), SAS High-Performance Forecasting allows for external models and user-defined models.

External models are used for forecasts that are provided external to the system. These external forecasts may have originated from an external statistical model from another software package, may have been provided by an outside organization (such as a marketing organization or government agency), or may be based on judgment.

User-defined models are external models that are implemented with the SAS programming language or the C programming language by the user of SAS High-Performance Forecasting. For these models, you can even create your own computational algorithms to generate the forecasts. These models are considered external because they were not implemented in SAS High-Performance Forecasting.

EXAMPLES

To illustrate how you can extend SAS High-Performance Forecasting, four easy-to-follow examples will be provided. The focus of these examples is to demonstrate the required steps and not so much the complexity of the models that are created. In fact, you will find that we are extending SAS High-Performance Forecasting with a moving average model of three observations. Furthermore, you will see how we implement this method first using a custom model repository, then using an external forecast, and finally using a user-defined model. Eventually, you will see how these three approaches can be combined in one repository.

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

As an example data set, we will use the airline passenger series, given as Series G in Box and Jenkins (1976). This series is often used in time series literature as an example of a non-stationary seasonal time series. This series is a monthly series consisting of the number of airline passengers in the United States who traveled during the years 1949 to 1960. Its main features are a steady rise in the number of passengers from year to year and the seasonal variation in the numbers during any given year. The airline passenger series is available in SASHELP library as the AIR data set.

Using a moving average model of three observations is by no means an appropriate model for the AIR data set. In fact, the HPFDIAGNOSE procedure will be able to come up with the so-called AIRLINE model or a Winter's exponential smoothing model on log-transformed data for you automatically. However, given the wide-spread use of the airline data and the simplicity of the moving average model, they make ideal candidates for illustrating the required steps for extending SAS High-Performance Forecasting.

EXAMPLE 1: CUSTOM REPOSITORY

In our first example, we will use existing SAS High-Performance Forecasting functionality to create a moving average of three observations (MA3) model. As a first step, we need to define the model specification. As a MA3 model can be expressed as an ARIMA model with fixed estimates, we are employing the HPFARIMASPEC procedure.

```
proc hpfarimaspec modelrepository=work.custom_repository
  specname=custom_MA3;
  forecast symbol=y noint
  p=3 ar=(0.3333333333 0.3333333333 0.3333333333);
  estimate noest;
run;
```

A couple of things to note in the previous code:

- We are storing the model specification in a temporary custom repository called work.custom_repository.
- The specification name of our model is MA3.
- The FORECAST statement allows us to specify the MA3 model in detail. In particular, we are providing the three specifications of the AR components that are 1/3 each.
- By using an ESTIMATE NOEST statement, we tell the procedure not to estimate any parameters but to take our parameter specifications into account.

As the next step, we will need to build our model selection list using the HPFSELECT procedure:

```
proc hpfselect modelrepository=work.custom_repository
  selectname=custom_select
  selectlabel="Custom MA 3 Model";
  diagnose seasonstest=none;
  specification custom_MA3;
run;
```

A couple of things to note in this code:

- We need to make sure that the model selection list is in the same custom repository as the model specification (custom_repository).
- The name of our selection list is custom_select.
- The SPECIFICATION statement points to our model specification (custom_MA3).
- By default, SAS High-Performance Forecasting runs a seasonality test that we are disabling with the DIAGNOSE statement. This statement is required to make sure SAS High-Performance Forecasting is not dropping our inadequate MA3 model and is using a seasonal model instead.

After defining the model specification and the model selection list, our custom repository is ready to use. Note that we have only one model available. Hence, there is no need to pick and choose a model from a wide range of candidates. In general, a custom repository can feature many different models.

In order to apply our model, we will need to run the HPFENGINE procedure. Note that we need to point the procedure to our repository (custom_repository) and the model selection list (custom_select). We will not only ask the procedure to create a data set (which contains our forecasts also referred to as predictions with confidence limits), but we will also create a forecast plot.

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

To get an idea of the forecasting accuracy, we are also using an out-of-sample period of 1 (specified by BACK=1), and we are forecasting one period ahead (specified by LEAD=1).

```
ods graphics on;
proc hpfengine data=sashelp.air
  modelrepository=custom_repository
  globalselection=custom_select
  out=_null_
  outfor=outfor1
  back=1
  lead=1
  plot=forecasts;
  id date interval=month;
  forecast air;
run;
ods graphics off;
```

Here is the output from our first example. Note that our MA3 model forecasts 453 (rounded) for our lead time in Dec 1960.

VIEWTABLE: Forecasts for SASHELP.AIR								
	Variable Name	DATE	Actual Values	Predicted Values	Lower Confidence Limits	Upper Confidence Limits	Prediction Errors	Prediction Standard Errors
137	AIR	MAY1960	472	423.66666243	325.45027883	521.88304603	48.33333757	50.111320603
138	AIR	JUN1960	535	450.66666216	352.45027856	548.88304576	84.33333784	50.111320603
139	AIR	JUL1960	622	489.33332844	391.11694484	587.54971204	132.66667156	50.111320603
140	AIR	AUG1960	606	542.99999457	444.78361097	641.21637817	63.00000543	50.111320603
141	AIR	SEP1960	508	587.66666079	489.45027719	685.88304439	-79.66666079	50.111320603
142	AIR	OCT1960	461	578.66666088	480.45027728	676.88304448	-117.6666609	50.111320603
143	AIR	NOV1960	390	524.99999475	426.78361115	623.21637835	-134.9999947	50.111320603
144	AIR	DEC1960	432	452.99999547	354.78361187	551.21637907	-20.99999547	50.111320603

Table 1 Forecasts Data Set for the Custom Repository Example

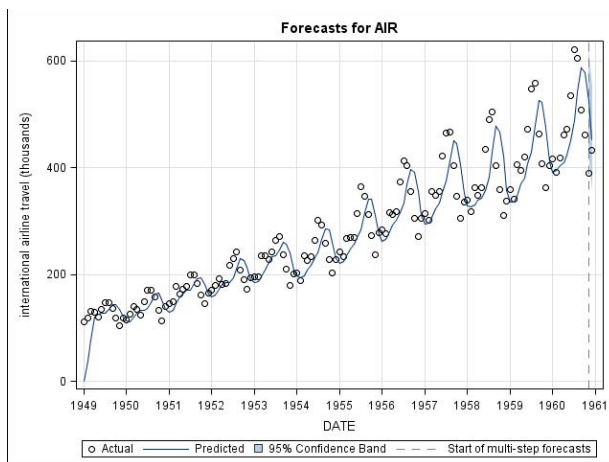


Figure 1 Forecast Plot for the Custom Repository Example

EXAMPLE 2: EXTERNAL FORECAST

External models are used for forecasts that are provided using features external to SAS High-Performance Forecasting. In this example, we will use the EXPAND procedure of SAS/ETS® software to come up with a MA3 model. Note that we could have used a DATA step to come up with these forecasts. In general, all we need to do is to store our forecasts in a SAS data set. These forecasts could come from any source as long as they are available as a SAS data set.

```
proc expand data=sashelp.air out=external_MA3 method=none;
  id date;
  convert air=predict / transformout=(lag 1 movave 3);
run;
```

Note that our MA3 forecasts are stored in a new variable called Predict in the EXTERNAL_M3 data set.

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

In order for SAS High-Performance Forecasting to identify the external forecasts, we will use the HPFEXMSPEC procedure. We are specifying a specification name (external_MA3) and a model repository (work.external_repository).

```
proc hpfexmspec modelrepository=work.external_repository specname=external_MA3;
  exm;
run;
```

As before, the next step is to define a model selection list using the HPFSELECT procedure.

```
proc hpfselect modelrepository=work.external_repository
  selectname=external_select
  selectlabel="External MA 3";
  specification external_MA3/exmmap(predict=predict);
run;
```

Different than before, we will need to tell the procedure how to reference the external forecasts. Using the EXMMAP option, we specify the name of the variable (Predict) that contains the forecasts. If we also have information about confidence intervals and standard deviation, we could provide this information as well.

```
ods graphics on;
proc hpfengine data=external_MA3
  modelrepository=work.external_repository
  globalselection=external_select
  out=_null_
  outfor=outfor2
  back=1
  lead=1
  plot=forecasts;
  id date interval=month;
  forecast air;
  external predict;
run;
ods graphics off;
```

Note that we need to use the data set that contains the external forecasts (not SASHELP.AIR). In our case, this data set was created using the EXPAND procedure and is called external_MA3. The external statement tells the HPFENGINE procedure where to look for the external forecasts.

As expected, the resulting outputs are similar to those in Example 1. Again the MA3 prediction is 453 for Dec1960.

VIEWTABLE: Forecasts for WORK.EXTERNAL_MA3								
	Variable Name	DATE	Actual Values	Predicted Values	Lower Confidence Limits	Upper Confidence Limits	Prediction Errors	Prediction Standard Errors
137	AIR	MAY1960	472	423.66666667	329.48261378	517.85071956	48.333333333	48.053971211
138	AIR	JUN1960	535	450.66666667	356.48261378	544.85071956	84.333333333	48.053971211
139	AIR	JUL1960	622	489.33333333	395.14928044	583.51738622	132.66666667	48.053971211
140	AIR	AUG1960	606	543	448.81594711	637.18405289	63	48.053971211
141	AIR	SEP1960	508	587.66666667	493.48261378	681.85071956	-79.66666667	48.053971211
142	AIR	OCT1960	461	578.66666667	484.48261378	672.85071956	-117.66666667	48.053971211
143	AIR	NOV1960	390	525	430.81594711	619.18405289	-135	48.053971211
144	AIR	DEC1960	432	453	358.81594711	547.18405289	-21	48.053971211

Table 2 Forecasts Data Set for the External Forecast Example

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

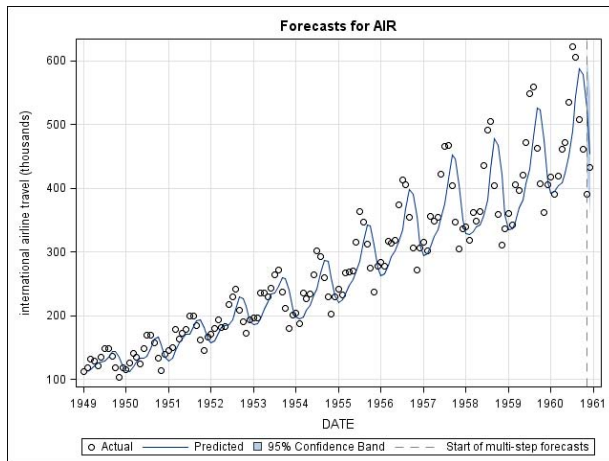


Figure 2 Forecast Plot for the External Forecast Example

EXAMPLE 3: USER-DEFINED MODEL

SAS High-Performance Forecasting allows you to add user-defined models. These models are also external models in a way, but they are implemented with the SAS programming language or the C programming language. Different from what we referred to as external models, you are creating your own computational algorithms to generate forecasts and run them as part of the SAS High-Performance Forecasting framework.

In this example, we will use the FCMP procedure to create our own computation of a MA3 model (using SAS programming language). The SAS Function Compiler (FCMP) procedure enables you to create, test, and store SAS functions, CALL routines, and subroutines before you use them in other SAS procedures or DATA steps. You can call PROC FCMP functions and CALL routines from the DATA step just as you would any other SAS function, CALL routine, or subroutine. This feature enables you to more easily read, write, and maintain code with independent and reusable subroutines.

We will not illustrate how to use the C programming language in this paper. However, you will find the following example in the *SAS High-Performance Forecasting 3.1: User's Guide*:

http://support.sas.com/documentation/cdl/en/hpfug/62015/HTML/default/hpfug_hpfuserdefined_sect003.htm

```
proc fcmp outlib=work.hpfengine.funcs;
  subroutine move_avg3(act[*], pred[*]);
    outargs pred;
    actlen = DIM(act);
    predlen = DIM(pred);
    pred[1] = 0;
    pred[2] = act[1]/3.0;
    pred[3] = (act[1] + act[2])/3.0;
    do i=4 to actlen+1;
      pred[i] = (act[i-1] + act[i-2] + act[i-3])/3.0;
    end;
    do i=actlen+2 to predlen;
      pred[i] = (pred[i-1] + pred[i-2] + pred[i-3])/3.0;
    end;
  endsub;
run;
```

By using the CMPLIB option, you tell SAS High-Performance Forecasting where to locate your new routine. Note that PROC FCMP stores its output in a SAS data set. In this example, the data set is called work.hpfengine.

```
options cmplib=work.hpfengine;
```

We are using the HPFEXMSPEC procedure to create a specification name (user_MA3) and a model repository (work.user_repository).

```
proc hpfexmspec modelrepository=work.user_repository specname=user_MA3;
  exm;
```

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

```
run;
```

Using the EXMFUNC option on the SPECIFICATION statement in PROC HPFSELECT, we are now able to call our user-defined function (move_avg3). As defined earlier, we are passing the actual values of our series to our function (specified by the _ACTUAL_ option), and the function returns the predictions (specified by the _PREDICT_ option).

```
proc hpfselect modelrepository=work.user_repository
  selectname=user_select;
  diagnose seautest=none;
  specification user_MA3 / exmfunc('move_avg3(_actual_ _predict_ )');
run;
```

The main difference to Example 2 is that we are not feeding the HPFENGINE procedure with forecasts that were created before running the procedure. This time we are creating these forecasts on-the-fly by running our newly defined function. Note that the syntax for the HPFENGINE procedure is almost similar to Example 1.

```
ods graphics on;
proc hpfengine data=sashelp.air
  modelrepository=work.user_repository
  globalselection=user_select
  out=_null_
  outfor=outfor3
  back=1
  lead=1
  plot=forecasts;
  id date interval=month;
  forecast air;
run;
ods graphics off;
```

The results are again similar to those in Example 1 and 2. The MA3 prediction is 453 for Dec1960.

VIEWTABLE: Forecasts for SASHELP.AIR

	Variable Name	DATE	Actual Values	Predicted Values	Lower Confidence Limits	Upper Confidence Limits	Prediction Errors	Prediction Standard Errors
137	AIR	MAY1960	472	423.6666667	326.76328552	520.57004781	48.33333333	49.441409083
138	AIR	JUN1960	535	450.6666667	353.76328552	547.57004781	84.33333333	49.441409083
139	AIR	JUL1960	622	489.3333333	392.42995219	586.23671448	132.66666667	49.441409083
140	AIR	AUG1960	606	543	446.09661885	639.90338115	63	49.441409083
141	AIR	SEP1960	508	587.6666667	490.76328552	684.57004781	-79.66666667	49.441409083
142	AIR	OCT1960	461	578.6666667	481.76328552	675.57004781	-117.66666667	49.441409083
143	AIR	NOV1960	390	525	428.09661885	621.90338115	-135	49.441409083
144	AIR	DEC1960	432	453	356.09661885	549.90338115	-21	49.441409083

Table 3 Forecasts Data Set for the User-Defined Forecast Example

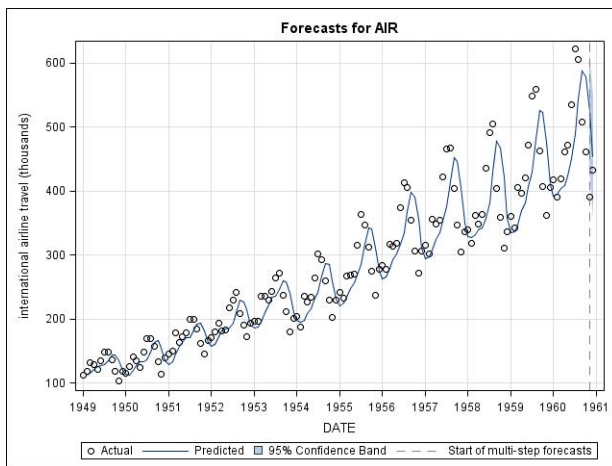


Figure 3 Forecast Plot for the User-Defined Forecast Example

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

EXAMPLE 4: COMBINING ALL THREE APPROACHES

You may wonder if it is possible to combine all three approaches in one PROC HPFENGINE run. The last example will illustrate this. As you will find a lot of the statements made in earlier examples are also true for this example. The only difference is the possibility of combining individual model repositories. The ability to combine model repositories makes SAS High-Performance Forecasting very powerful, because you have complete control over how to organize and run individual model repositories. In fact, you can also combine your repositories with the repository that is created by PROC HPFDIAGNOSE, even though this task is not outlined in this paper.

In order to combine existing model repositories into one, we will use the CATNAME statement. Our new combined repository is now called MA3_repository. Of course, it contains three times the same model but with different ways to compute the forecasts.

```
catname MA3_repository (work.custom_repository work.external_repository
work.user_repository);
```

Now, the HPFSELECT procedure is combining the approaches outlined in Example 1, 2, and 3.

```
proc hpfselect modelrepository=work.MA3_repository
selectname=MA3_select
selectlabel="MA3";
diagnose seasontest=none;
specification external_MA3/exmmap(predict=predict);
specification custom_MA3;
specification user_MA3 / exmfunc('move_avg3(_actual_ _predict_ )');
run;
```

As we have now more than one model in our repository, we will let PROC HPFENGINE pick the model which fits our data best. We will use a hold-out sample to pick a model (HOLDOUT=1) and print details about the model selection process.

```
proc hpfengine data=external_MA3
modelrepository=work.MA3_repository
globalselection=MA3_select
task=select(holdout=1)
out=_null_
outfor=outfor4
back=1
lead=1
print=select;
id date interval=month setmissing=0;
forecast air;
external predict;
run;
```

Model Selection Criterion = MAPE			
Model	Statistic	Selected	Label
EXTERNAL_MA3	34.615385	No	External Model: Perfect Autocorrelation
CUSTOM_MA3	34.615383	Yes	ARIMA: Y ~ P = 3 NOINT
USER_MA3	34.615385	No	External Model: Perfect Autocorrelation

Output 1 Output from Print Statement

Note that by default the model selection criterion is MAPE. The differences of the MAPE values are due to the way we have implemented our forecasting models. As expected, all three models are performing in a similar way.

Extending SAS® High-Performance Forecasting Using User-Defined Models and External Forecasts, continued

CONCLUSION

For automatic forecasting of large numbers of time series, SAS High-Performance Forecasting provides access to the most robust models for forecasting, such as exponential smoothing models, unobserved component models, autoregressive integrated moving average models, and intermittent demand models. As illustrated, SAS High-Performance Forecasting also allows experienced forecasters to extend the models in three ways:

- Custom repositories: You can create your own custom model repository.
- External models: Your forecasts are provided by methods that are external to the system.
- User-defined models: You are adding forecasting methods that are not provided by SAS High-Performance Forecasting.

By providing this ultimate flexibility, SAS High-Performance Forecasting can be considered as one of the most complete automatic forecasting engines available today.

REFERENCES

- SAS Institute Inc. 2004. SAS Institute white paper. "Large-Scale Automatic Forecasting with Inputs and Calendar Events." <http://www.sas.com/reg/wp/corp/3478>
- Box, George. E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control* Englewood Cliffs, NJ: Prentice Hall.
- SAS Institute Inc. 2009. *SAS High-Performance Forecasting 3.1: User's Guide*. Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

A special thanks to Michael Leonard, Meredith John, Mike Gilliland, all SAS Institute Inc., and Snurre Jensen from SAS Institute in Denmark for providing technical and other information used in this paper.

CONTACT INFORMATION

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

Udo Sglavo
SAS Campus Drive
SAS Institute Inc.
E-mail: Udo.Sglavo@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.