

Combined Forecasts: What to Do When One Model Isn't Good Enough

Ed Blair, Michael Leonard, Bruce Elsheimer, SAS Institute Inc.

ABSTRACT

SAS® High-Performance Forecasting 4.1 offers a new, innovative process for automatically combining forecasts. Forecast combination, also called ensemble forecasting, is the subject of many academic papers in statistical and forecasting journals; it is a known technique for improving forecast accuracy and reducing variability of the resulting forecasts. By integrating these methods into a single software system, SAS High-Performance Forecasting 4.1 surpasses the functionality of any existing software system that incorporates this capability. This paper describes this new capability and includes examples that demonstrate the use and benefits of this new forecast combination process.

INTRODUCTION

The M-Competition is a periodic public competition of forecasting methods. This competition pits researchers and their methodologies against one another to forecast a collection of time series data. The goal of M-Competition is to evaluate the skills of researchers by the accuracy of their predictions. Makridakis and Hibon (2000) wrote the following as one of their four conclusions about the results of a recent M3 competition: "The accuracy of the combination of various methods outperforms, on average, the specific methods being combined and does well in comparison with other methods."

The lesson from this statement is that a combination of forecasts from simple models can add substantial value in terms of enhancing the quality of the forecasts produced, but the statement also admits that combinations might not always perform better than a suitably crafted model. With SAS High-Performance Forecasting 4.1, these combinations can be generated automatically and integrated into a well-proven framework for model selection and forecasting. The value of the tool is enhanced, and overall forecast quality can be improved by the use of combined forecasts.

The remaining sections of this paper discuss several aspects of forecast combination and provide examples that demonstrate the use of combined forecasts in the SAS High-Performance Forecasting procedures. First, there is a short case-study to explore why forecast combinations can be beneficial. That is followed by a brief summary of the mathematical details of forecast combination, a discussion of the forecast combination process flow, and a discussion of the model selection process as extended to incorporate combined forecasts. Several examples follow that demonstrate different aspects of combined forecast usage. These include SAS statements and samples of the output they produce.

For readers who are not familiar with the SAS High Performance Forecasting product, see the *SAS High-Performance Forecasting 4.1: User's Guide* for information about the various SAS procedures used in the examples. It is beyond the scope of this paper to describe these procedures and their use except as they pertain to the new features related to forecast combinations.

WHY COMBINE FORECASTS?

When might the combined method of forecasting outperform other forecasting methods? The time series for monthly oil production for the Central region in the Gulf of Mexico (2006) makes a compelling case for combined forecasts. The oil production data span the period from 1996 to 2006. PROC HPFDIAGNOSE is run to generate an autoregressive integrated moving average (ARIMA) model and an exponential smoothing model (ESM) for this series, and then the HPFENGINE procedure is run to select the best model based on a holdout sample of six months.

Figure 1 shows the results of the PROC HPFENGINE run. Although both of these models might be sufficiently accurate expressions of the data-generating process, would some weighted average of the forecasts produce a better-quality forecast?

Figure 1 Central Oil Production Model Selection

Central region oil production Generated models		
The HPFENGINE Procedure		
Model Selection Criterion = MAPE		
Model	Statistic	Selected
diag54	22.763945	No
diag55	18.745997	Yes
Model Selection Criterion = MAPE		
Model	Label	
diag54	ARIMA: oil ~ P = 1 D = (1,12) Q = ((1) (12,24)) NOINT	
diag55	Winters Method (Multiplicative)	

To answer this question, you can include a simple average of the forecasts from the generated ARIMA and ESM models in the mix and repeat the model selection process.

Figure 2 shows the results of the PROC HPFENGINE run with a combined forecast included. The combined forecast, judged by its mean absolute percentage error (MAPE) statistic over the last six months of the time span, is substantially better than either of its individual contributors. The combined forecast improves forecast accuracy by 58% relative to the best forecast from the time series models alone. Averaging the two contributors has effectively incorporated dynamics into the forecast that neither model alone could provide in isolation.

Figure 2 Model Selection with Combination Included

Central region oil production Generated models+Combination		
The HPFENGINE Procedure		
Model Selection Criterion = MAPE		
Model	Statistic	Selected
DIAG0	22.763945	No
DIAG1	18.745997	No
COMB	7.795160	Yes
Model Selection Criterion = MAPE		
Model	Label	
DIAG0	ARIMA: oil ~ P = 1 D = (1,12) Q = ((1) (12,24)) NOINT	
DIAG1	Winters Method (Multiplicative)	
COMB	AVG (DIAG0,DIAG1)	

A comparison of the forecasts and prediction errors for the respective models and their combination provides some useful insight. Figure 3 overlays the individual forecasts on a single plot with an exploded view at the end of the time span, while Figure 4 overlays the individual forecast prediction errors on a single plot with an exploded view at the end of the time span.

Figure 3 Forecast Plots

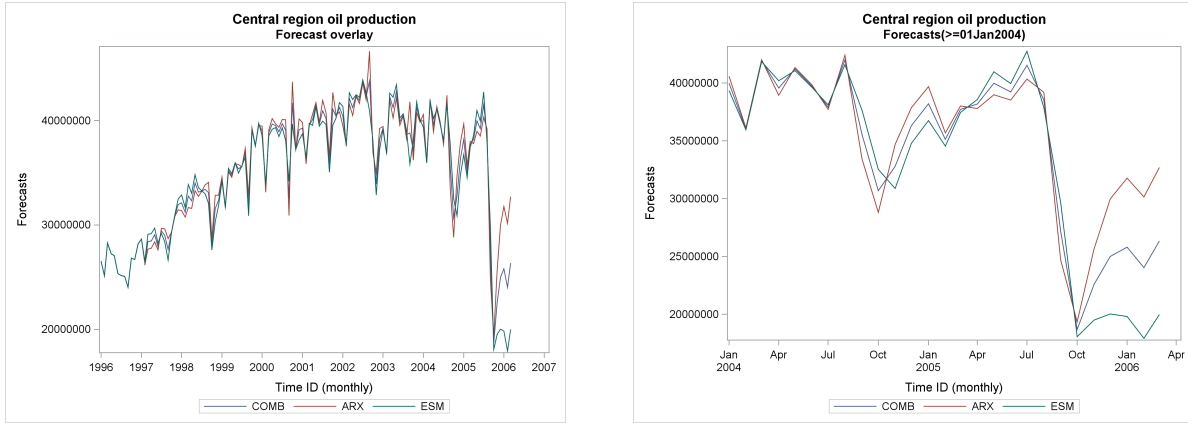
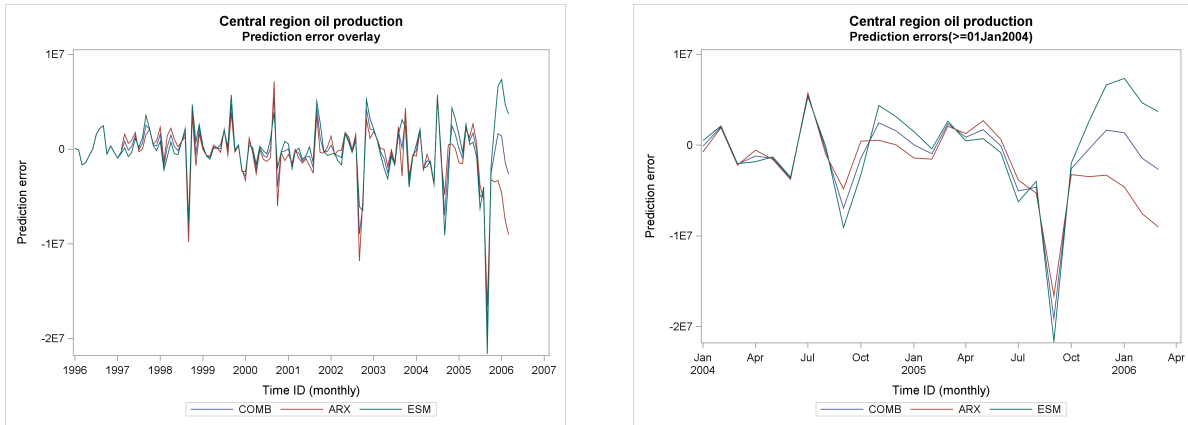


Figure 4 Prediction Error Plots



The forecasts from the ARIMA and ESM models start to diverge towards the end of the time span. Each is responding differently to the abrupt change in the Central region oil production that starts to occur around August 2005. One forecast is overestimating, and the other is underestimating. The simple average serves to correct for these different dynamics in the respective candidate models. Over the early history of the evolution, you can see that there is little difference between the combined forecast and its contributors.

The value of the technique of combining forecasts is well-demonstrated in this test case by its ability to improve predictive accuracy as measured by the user-chosen selection criterion. Not every time series prediction benefits from combination. The power of this technique becomes apparent when you consider that SAS High-Performance Forecasting 4.1 allows for combination methods to be applied to large-scale time series forecasting of hierarchically structured data. The software makes it possible to generate combinations for inclusion into its model selection process in an automated fashion. In all cases, combined forecasts must prove their worth by their performance in comparison to other forecasts in the model selection process.

FORECAST COMBINATION MATHEMATICAL DETAILS

Given a time series $y_t, t = 1, \dots, T$, and set of M forecasting models denoted by $F_i(), i = 1, \dots, M$, with possible inputs (not shown), presume that

$$y_t = F_i(Y_{t-1}) + \epsilon_{i,t}$$

where Y_{t-1} denotes a vector of time series observations up to time $t - 1$ and the following conditions are assumed for $\epsilon_{i,t}$:

$$\begin{aligned} E[\epsilon_{i,t}] &= 0 \\ E[\epsilon_{i,t}\epsilon_{j,s}] &= 0, t \neq s \\ E[\epsilon_{i,t}\epsilon_{j,t}] &= \rho_{i,j,t}\sigma_{i,t}\sigma_{j,t} \end{aligned}$$

where $\sigma_{i,t}$ denotes the prediction standard error for models i and j at time t , and $\rho_{i,j,t}$ denotes the error correlation between models i and j at time t .

Forecasting results from the set of models $F_i()$ are presumed to be available at the outset of the combination processing. They are generated on demand from running the models and consuming their results to generate the forecast combination.

Independently, each model $F_i()$ is fit to produce a fitted forecast model that is denoted $\hat{F}_i()$. Define the following notation for the results of $\hat{F}_i()$:

$\hat{y}_{i,t}$ denotes the prediction for fitted model $\hat{F}_i()$ at time t .

$e_{i,t} = y_t - \hat{y}_{i,t}$ denotes the prediction error for fitted model $\hat{F}_i()$ at time t .

$\hat{\sigma}_{i,t}^2 = \text{Var}(e_{i,t})$ denotes the prediction error variance for fitted model $\hat{F}_i()$ at time t .

$\hat{\sigma}_{i,t} = \sqrt{\text{Var}(e_{i,t})}$ denotes the prediction standard error for fitted model $\hat{F}_i()$ at time t .

The combined forecast, denoted \hat{y}_c , uses combination weights w_i to combine the fitted model forecasts as a weighted average of the predicted values with possible restrictions on the combination weights. More precisely,

$\hat{y}_{c,t} = \sum_{i=1}^M w_i \hat{y}_{i,t}$ denotes the combined forecast at time t .

The combination weights can correspond to a simple average, you can specify them, or they can be estimated using a variety of methods from the fitted forecast results. For the present discussion, the estimated weights are denoted \hat{w}_i regardless of how they are obtained. This leads to the following expressions for the combined forecast and its prediction error:

$\hat{y}_{c,t} = \sum_{i=1}^M \hat{w}_i \hat{y}_{i,t}$ denotes the combined forecast prediction at time t .

$e_{c,t} = y_t - \hat{y}_{c,t}$ denotes the combined forecast prediction error at time t .

The prediction error variances are most generally defined by

$$\text{Var}(e_{c,t}) = \text{Var}(y_t - \hat{y}_{c,t})$$

When you assume or impose the constraint $\sum_{i=1}^M \hat{w}_i = 1$, simplification results in

$$\begin{aligned} \text{Var}(e_{c,t}) &= \text{Var}\left(\sum_{i=1}^M \hat{w}_i e_{i,t}\right) \\ &= \hat{W}_t^T \Sigma_t \hat{W}_t \end{aligned}$$

where \hat{W}_t denotes the fitted weight vector $M \times 1$, and $\Sigma_t = \text{Var}(E_t)$ with $E_t = (e_{1,t}, e_{2,t}, \dots, e_{M,t})^T$ denotes the ensemble prediction error vector at time t .

Alternatively, this can be expressed as

$$\begin{aligned} \text{Var}(e_{c,t}) &= \sum_{i=1}^M \hat{w}_i^2 \text{Var}(e_{i,t}) + 2 \sum_{j < i} \hat{w}_i \hat{w}_j \text{Cov}(e_{i,t}, e_{j,t}) \\ &= \sum_{i=1}^M \hat{w}_i^2 \sigma_{i,t}^2 + 2 \sum_{j < i} \hat{w}_i \hat{w}_j \rho_{i,j,t} \sigma_{i,t} \sigma_{j,t} \end{aligned}$$

If the prediction errors for the individual forecasts are not correlated at time t , then the combined forecast prediction error variance and standard error series are simply

$$\text{Var}(e_{c,t}) = \sum_{i=1}^M \hat{w}_i^2 \sigma_{i,t}^2$$

$$\hat{\sigma}_t = \sqrt{\sum_{i=1}^M \hat{w}_i^2 \sigma_{i,t}^2}$$

For computation, the $\hat{\sigma}_{i,t}$ values from the fitted models \hat{F}_i are used as estimates of $\sigma_{i,t}$.

FORECAST COMBINATION PROCESS

As described in the previous section, forecast combination operates in the context of a set of forecast models, $\{F_i()\}_{i=1}^M$, and uses the results of the set of the corresponding fitted models $\{\hat{F}_i()\}_{i=1}^M$. That discussion of the mathematical details of forecast combination presumes all of the forecasts in the set $F_i()$ are used in the combination. The set of candidate forecasts that are considered during a forecast combination can in fact be a superset of the active set of forecasts that are finally combined. Denote the candidate set of forecasts for consideration as $\mathcal{F}(0)$. Elements of this set are some collection of forecasts $F_i(), i = 1, \dots, M$. As you will see shortly, the set of candidate models in the combination process can be defined through the HPFSELECT procedure or can be automatically generated from the HPFDIAGNOSE procedure. The set of candidate models so defined is referred to as a combined model list.

The forecast combination process follows a sequence of steps that operates on and within a set of candidate forecasts. In the abstract, step i in this process takes a set $\mathcal{F}(i-1)$ as input and produces a set $\mathcal{F}(i)$ as output to the next step. Some steps function solely to potentially reduce the candidate set $\mathcal{F}(i-1)$ by the application of some form of test. The initial candidate set $\mathcal{F}(0)$ is reduced to a final set of forecasts that are combined in the final step of the process.

The steps in the forecast combination process are listed here. Subsequent sections provide summaries for some of these to support the examples used in this paper. You can find more detailed information in Chapter 17 of the *SAS High-Performance Forecasting 4.1: User's Guide*.

1. Candidate exclusion tests
2. Encompassing tests
3. Weight estimation
4. Forecast combination and confidence intervals

Candidate Exclusion Tests

The time series y_t to be modeled and forecast has certain intrinsic properties (such as seasonality and trend) that are determined by diagnostic tests if enabled. Some candidate models might not have matching characteristics; they are excluded from further consideration. As a result, a possibly reduced set of candidate forecasts $\mathcal{F}(1)$ is generated. If $\mathcal{F}(1)$ is empty, no further processing is performed and the combined forecast produces no result.

You can request that the candidates $\mathcal{F}(1)$ be subjected to forecast quality tests. For each $F_i() \in \mathcal{F}(1)$, two forms of missing value tests can be performed on its predicted series $\hat{y}_{i,t}, t = 1, \dots, T$:

- Forecast fit region percentage missing: You can specify a test for missing forecast values over the fit region of the $\hat{y}_{i,t}$ series as a threshold, P_{miss} . Let p_i denote the percentage missing from $\hat{y}_{i,t}$ over the fit region. If $p_i \geq P_{miss}$, then the forecast for candidate $F_i()$ is omitted from $\mathcal{F}(2)$.
- Forecast horizon percentage missing: You can specify a test for missing values in $\hat{y}_{i,t}$ over the forecast horizon as a threshold, H_{miss} . The forecast horizon is the range of time indices that correspond to multistep forecasts denoted by $[t_b^H, t_e^H]$ with span $H = t_e^H - t_b^H$. Let h_i denote the percentage of missing values from $\hat{y}_{i,t}$ over the forecast horizon. If $h_i \geq H_{miss}$, then the forecast for candidate $F_i()$ is omitted from $\mathcal{F}(2)$.

The end result of this step is the set of candidates $\mathcal{F}(2) \subseteq \mathcal{F}(1)$.

Combination Weight Methods

This step assigns weights to the forecasts that are present in the set $\mathcal{F}(3)$ as directed by the combined model list. Forecast combination weights $\hat{w}_i, \forall i \in \mathcal{F}(3)$, are determined by one of the following methods:

- simple average
- user-defined weights

- rank weights
- ranked user weights
- root mean squared error (RMSE) weights
- corrected Akaike's information criterion (AICC) weights
- ordinary least squares (OLS) weights
- restricted least squares weights
- least absolute deviation (LAD) weights

Provision is made for weight assignment to reduce the set of forecasts based on method-specific criteria. For example, for some estimation methods, a weight might be nonestimable, and the corresponding forecast is eliminated from use in the combined forecast. For other methods, you can define a criterion to exclude weights that fall below a given threshold. When this criterion is satisfied, the corresponding forecasts are eliminated from use in the combined forecast. This step produces the set $\mathcal{F}(4)$, which contains the final set of candidates to be combined along with their respective weights $\hat{w}_i, i \in \mathcal{F}(4)$.

Forecast Generation

This step performs the weighted combination of the forecasts in the set $\mathcal{F}(4)$. For each $t, t = 1, \dots, T$, the forecast combination produces a value for $\hat{y}_{c,t}$ according to the expression

$$\hat{y}_{c,t} = \sum_{i \in \mathcal{F}(4)} \hat{w}_{i,t} \hat{y}_{i,t}$$

where $\hat{y}_{i,t}$ denotes the prediction for fitted model $\hat{F}_i()$ at time t and $\hat{w}_{i,t}$ denotes the weight assigned to fitted model $\hat{F}_i()$ at time t .

The presence of missing values in the forecasts at a given time t affects the generation of the combined forecast $\hat{y}_{c,t}$. The resulting value is determined by one of two missing value treatment methods. You can specify this behavior in the combined model list for the forecast combination; if specified, it takes precedence over any default. The following briefly describes combination behavior at a given time t when one or more of the $\hat{y}_{i,t}$ values are missing:

- Rescale (MISSMODE=RESCALE): \hat{w}_i values associated with the nonmissing $\hat{y}_{i,t}$ values are dynamically normalized to satisfy the requirement for the sum of weights to equal 1.
- Missing (MISSMODE=MISSING): The combined forecast $\hat{y}_{c,t}$ is set to missing.

Forecast Confidence Intervals

The method of estimating prediction error variance can be specified in the combined model list for the forecast combination. In all cases, the estimated prediction error variance of the combined forecast uses the estimates of prediction error variance from the forecasts that are finally combined and the exact weights used in that combination at each time t . Recall the previous section's discussion of missing value treatment modes that affect the per-time index weight values. The combination weight estimates are treated as if they are known parameters in all the prediction error variance expressions.

You can specify one of the following values for the STDERR option in the COMBINE statement to select the prediction error variance estimation method:

STDERR=DIAG specifies that the variance computation assume that the forecast errors at time t are uncorrelated so that the simple diagonal form of Σ_t is used. This is the default method.

STDERR=ESTCORR specifies that the variance computation estimate correlations $\rho_{i,j,t}$ via the sample cross-correlation between $e_{i,t}$ and $e_{j,t}$ over the time span $t = 1, \dots, T$, where T denotes the last time index of the actual series y_t . Of course, using STDERR=ESTCORR implies that the error series $e_{i,t}$ and $e_{j,t}$ are assumed to be jointly stationary.

STDERR=ESTCORR(TAU= τ) is similar to STDERR=ESTCORR except that the cross-correlation estimates are localized to a time window of τ steps. The time span $t = 1, \dots, T$ is quantized into segments of τ steps working from T backwards for in-sample cross-correlation estimates. The cross-correlation estimates from the interval $[T - \tau, T]$ are used for the period of multistep forecasts that extend beyond time T .

MODEL SELECTION: FORECAST MODEL SELECTION GRAPH

SAS High-Performance Forecasting 4.1 defines a model selection mechanism that enables you to define a wide range of strategies. Pertinent facts of the model selection process follow:

- Forecasts arise from time series models and from lists of forecasts.
- Both models and lists have specification names.
- The HPFESMSPEC, HPFIDMSPEC, HPFARIMASPEC, HPFUCMSPEC, and HPFEXMSPEC procedures define time series model specifications.
- The HPFSELECT procedure defines lists.
- Lists reference their contributors by name.
- Lists have either model selection or combination semantics.
- You can specify a statistic-of-fit criterion that defines how forecasts in lists are ranked.
- List semantics dictate how ranked forecasts are used to produce the list's forecast.

Consequently, starting from a model selection list, a rooted directed acyclic graph (DAG) is realized by connecting the specification references together. This DAG is termed a forecast model selection graph. While this is a new concept in SAS High-Performance Forecasting 4.1, you can easily observe that it extends and generalizes the one-level model selection list that existed in previous releases. This more general decision topology is readily motivated from considering two simple use cases:

- selecting between the best-performing individual model forecast and a combined forecast
- combining the best-performing forecast from a model selection list with other forecasts

Supporting these two scenarios basically provides the framework for the generality of the forecast model selection graph. Figure 5 and Figure 6 depict both of these scenarios.

Figure 5 Select between Models and Combination

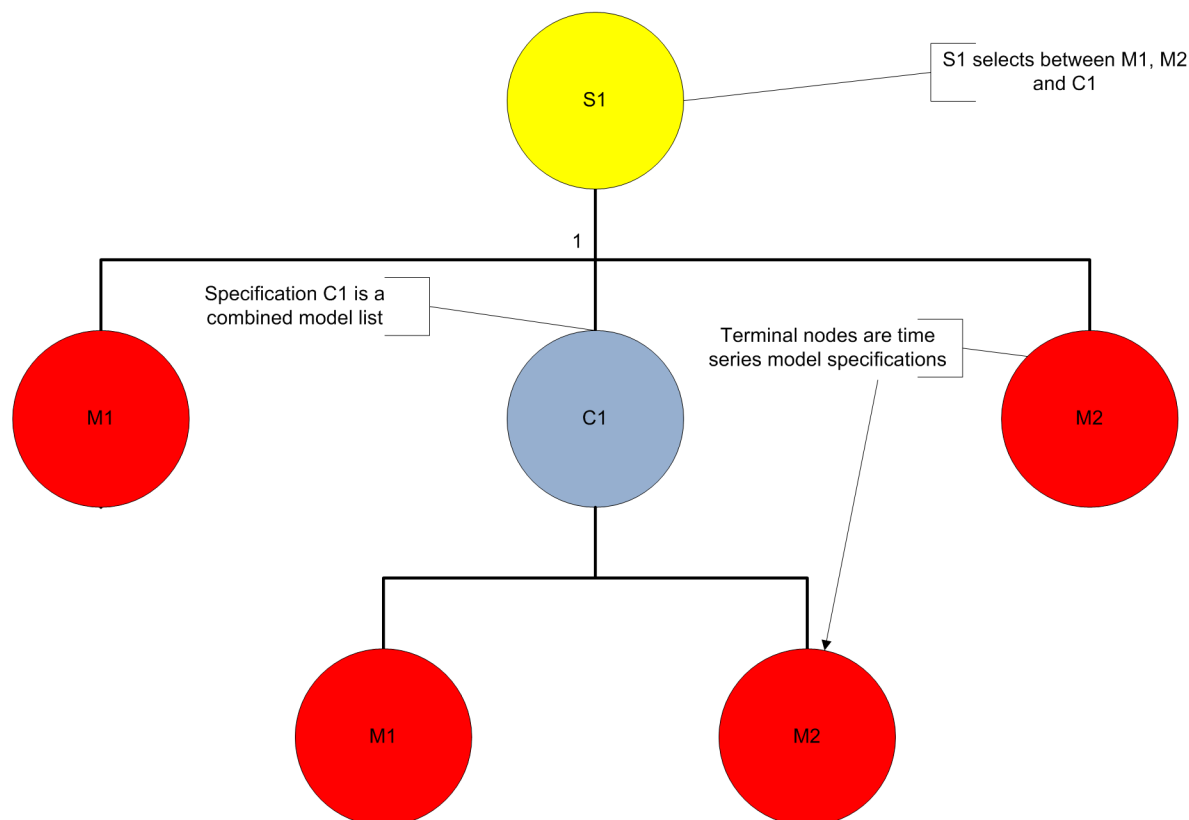
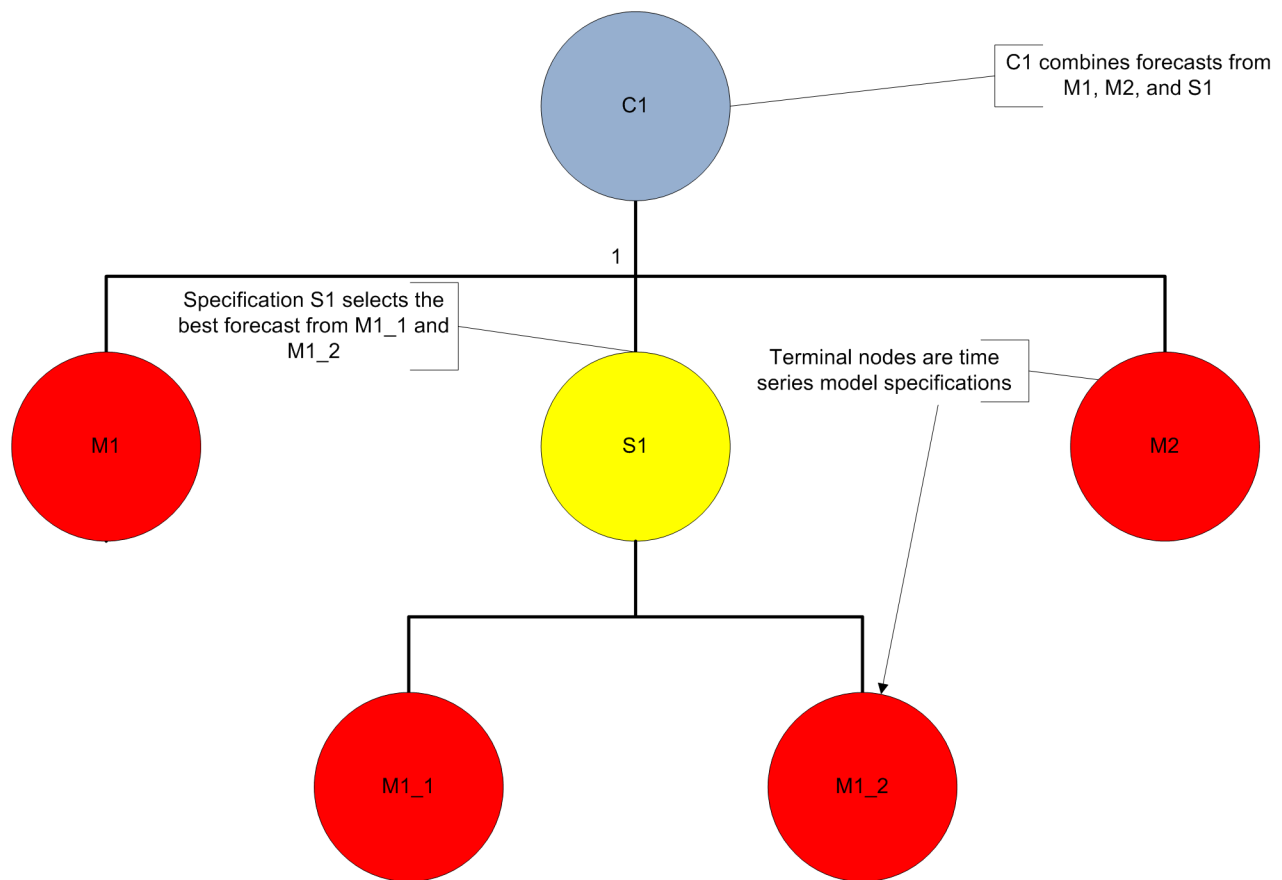


Figure 6 Combine Models and Selected Model



A SIMPLE EXAMPLE

This simple example demonstrates the use of a combined model list that uses a simple average of the Box-Jenkins (1976) AIRLINE model and the best seasonal exponential smoothing model (BESTS) to forecast monthly totals for international passengers from the data set Sashelp.Air. First, use the HPFARIMASPEC and HPFESMSPEC procedures to define the model specifications for the AIRLINE and BESTS models:

```

proc hpfarimaspec rep=work.rep specname=airline;
    forecast symbol=y dif=(1,s) q=(1)(1)s noint;
run;

proc hpfesmspec rep=work.rep specname=bests;
    esm method=bests;
run;

```

Then use the HPFSELECT procedure to define the combined model list named COMBAVG to compute the simple average of the forecasts from the AIRLINE and BESTS models (all other combination options are defaulted):

```

proc hpfselect rep=work.rep name=combavg label="AVG(AIRLINE,BESTS)";
    combine method=average;
    spec airline bests;
run;

```

Then define a model selection list AIRSELECT to select between the forecasts from the AIRLINE model and the combined forecast given by COMBAVG. A holdout region of six months is used to select the best forecast using MAPE as the statistic of fit for selection.

```

proc hpfselect rep=work.rep name=airselect;
    select holdout=6 criterion=mape;
    spec combavg airline;
run;

```


Finally, use the HPFENGINE procedure to evaluate the model selection list that is named AIRSELECT. There is no indication in the following PROC HPFENGINE statements that a combined model list is being used. PROC HPFENGINE constructs the forecast model selection graph as directed by the GLOBALSELECTION= option and evaluates it for the series Air from Sashelp.Air. The use of a combined forecast is defined by the reference to the COMBAVG specification from the model selection list AIRSELECT as one of the candidates to include for evaluation.

```
proc hpfengine data=sashelp.air
    rep=work.rep globalselection=airselect
    print=(select estimates)
    plot=(components);
    id date interval=month;
    forecast air;
run;
```

In the model selection results in [Figure 7](#), you can see that the combination of the AIRLINE model forecast with the ESM forecast produces a better selection statistic of fit than the AIRLINE model alone.

Figure 7 Model Selection Results

The HPFENGINE Procedure					
Model Selection Criterion = MAPE					
Model	Statistic	Selected	Label		
COMBAVG	2.3528210	Yes	AVG(AIRLINE,BESTS)		
AIRLINE	3.7973928	No	ARIMA: Y ~ D = (1,s) Q = ((1)(1)s) NOINT		

[Figure 8](#) through [Figure 10](#) show the parameter estimates for all of the models that contribute to the combination in addition to those for the combined forecast proper.

Figure 8 Parameter Estimates for AIRLINE Model

Parameter Estimates for AIRLINE Model					
Component	Parameter	Estimate	Standard Error	t Value	Approx Pr > t
AIR	MA1_1	0.30945	0.08428	3.67	0.0004
AIR	MA2_12	0.11279	0.10170	1.11	0.2695

Figure 9 Parameter Estimates for BESTS Model

Parameter Estimates for BESTS Model					
Component	Parameter	Estimate	Standard Error	t Value	Approx Pr > t
AIR	Level Weight	0.30728	0.03153	9.74	<.0001
AIR	Trend Weight	0.0010000	0.0030237	0.33	0.7413
AIR	Seasonal Weight	0.87493	0.07769	11.26	<.0001

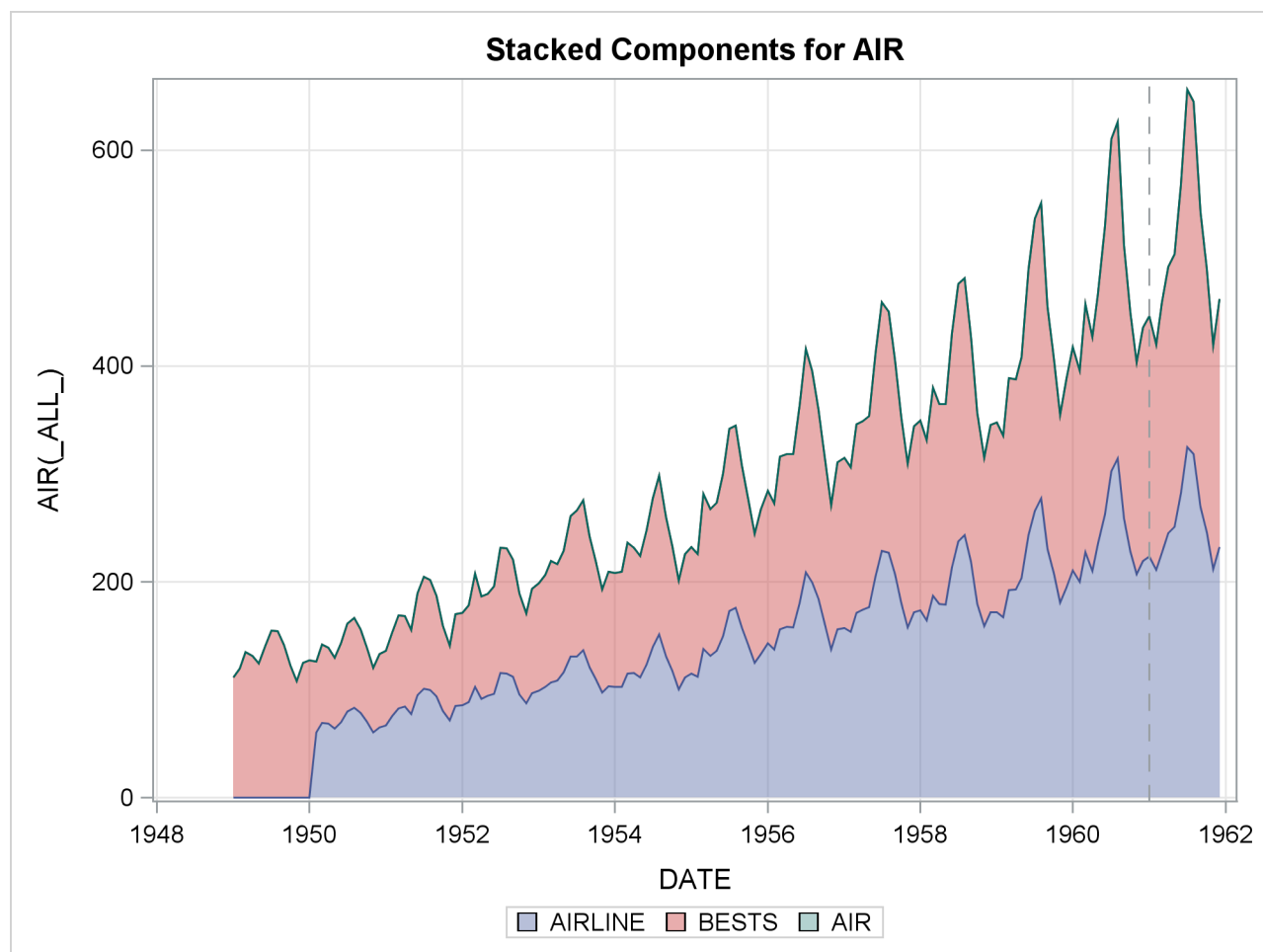
Figure 10 Parameter Estimates for COMBAVG Model

Parameter Estimates for COMBAVG Model					
Component	Parameter	Estimate	Standard Error	t Value	Approx Pr > t
AIRLINE	WEIGHT	0.50000	.	.	.
BESTS	WEIGHT	0.50000	.	.	.

The PLOT=COMPONENTS option for PROC HPFENGINE produces a graph that includes the weighted forecast components that contribute to the final combined forecast. The weighted components are plotted in a stack on the same set of axes with the weighted forecast components arranged from bottom to top in the order of their combination to

yield the final combined forecast. The graph reflects the result of the default MISSMODE=RESCALE treatment for the leading missing values in the AIRLINE model forecast where only forecasts from the BESTS model are available.

Figure 11 Combined Forecast Components



COMPARING COMBINATION WEIGHT METHODS

Extending the previous example, you can compare the results of different combination methods in a single run of the HPFENGINE procedure. The following statements compare the forecasts that result from simple averaging, AICC weights, RMSE weights, and LAD weights with the AIRLINE model forecast:

```
proc hpfselect rep=work.rep name=combavg label="AVG(AIRLINE,BESTS)";
  combine method=average;
  spec airline bests;
run;

proc hpfselect rep=work.rep name=combaicc label="AICC(AIRLINE,BESTS)";
  combine method=aicc;
  spec airline bests;
run;

proc hpfselect rep=work.rep name=combrmse label="RMSEWGT(AIRLINE,BESTS)";
  combine method=rmsewgt;
  spec airline bests;
run;

proc hpfselect rep=work.rep name=comblad label="LAD(AIRLINE,BESTS)";
  combine method=lad(errtype=ape);
  spec airline bests;
run;
```

```
proc hpfsselect rep=work.rep name=comparecombs;
  select holdout=6 criterion=mape;
  spec airline combavg combaicc combrmse comblad;
run;
```

From the model selection results in Figure 12, the LAD method produces a better selection statistic of fit than any of the other candidate forecasts. This should not be too surprising since the LAD objective effectively decides the combination weights to minimize absolute percentage error.

Figure 12 Model Selection Results

The HPFENGINE Procedure				
Model Selection Criterion = MAPE				
Model	Statistic	Selected	Label	
AIRLINE	3.7973928	No	ARIMA: $Y \sim D = (1,s) \quad Q = ((1)(1)s) \quad \text{NOINT}$	
COMBAVG	2.3528210	No	AVG (AIRLINE,BESTS)	
COMBAICC	3.7973928	No	AICC (AIRLINE,BESTS)	
COMBRMSE	2.3108299	No	RMSEWGT (AIRLINE,BESTS)	
COMBLAD	2.2461506	Yes	LAD (AIRLINE,BESTS)	

COMBINING JUDGMENTAL FORECASTS

Externally specified models enable forecasts from external sources to be included into the model selection process in the HPFENGINE procedure. Since ensemble models operate on forecasts, you can naturally expect a combination of forecasts from externally specified models to function in this setting. This simple example demonstrates such a scenario.

For the data set Sashelp.Pricedata, suppose you have sales forecasts that arise from the judgment of three independent sources. Suppose you want to compare a weighted combination of those judgmental forecasts against the sales forecast from the best exponential smoothing model (BESTESM) where the weighted combination assigns predetermined weights to the combination candidates. This example assumes that the external forecasts are contained in a separate data set Perm.External where the variables salefn denote the three judgmental forecasts to be combined. This example also demonstrates the use of the AUXDATA= option for supporting auxiliary data set access. This is another new feature in SAS High-Performance Forecasting 4.1.

First, you create three separate external models to identify each of the combined forecast contributors (named XF1, XF2, and XF3), and then you combine those with the combined model list COMBXF. Combination weights of 0.5, 0.3, and 0.2 are assigned to XF1, XF2, and XF3, respectively. You must map each external model specification to the data set variable that supplies its forecast. Finally, the model selection list XFSELECT defines the comparison between BESTESM and COMBXF over a holdout region of six months by using MAPE as the selection criterion.

```
proc hpfesmspec rep=work.rep specname=bestesm;
  esm method=best;
run;

proc hpfxmspec rep=work.rep specname=xf1;
  exm;
run;

proc hpfxmspec rep=work.rep specname=xf2;
  exm;
run;

proc hpfxmspec rep=work.rep specname=xf3;
  exm;
run;

proc hpfsselect rep=work.rep name=combxflabel="USERDEF (XF1,XF2,XF3) ";
  combine method=userdef(0.5,0.3,0.2);
  spec xf1 /exmmap(predict=salef1);
  spec xf2 /exmmap(predict=salef2);
  spec xf3 /exmmap(predict=salef3);
run;

proc hpfsselect rep=work.rep name=xfselect;
  select holdout=6 criterion=mape;
```

```
spec combxf bestesm;
run;
```

You now run the HPFENGINE procedure on the primary data set Sashelp.Pricedata with the external forecasts in Perm.External. PROC HPFENGINE constructs the forecast model selection graph as directed by the GLOBALSELECTION= option and runs it for each BY group in Sashelp.Pricedata. PROC HPFENGINE selects the best forecast for each BY group independently from the candidates that are evaluated in the forecast model selection graph.

```
proc hpfengine data=sashelp.pricedata
  auxdata=perm.external
  rep=work.rep globalselection=xfselect
  out=out3 outfor=for3
  outmodelinfo=minfo3
  outstatselect=oss3
  outstat=stat3
  outest=est3
  lead=12;
by region line product;
id date interval=month horizonstart='01jan2003'd end='01dec2003'd;
forecast sale;
external salef1 salef2 salef3;
run;
```

Sashelp.Pricedata contains 17 BY groups for the hierarchy Region, Line, Product. Figure 13 shows a summary of MAPE statistics for the selected models over the BY groups on the basis of whether a combined judgmental forecast was selected.

Figure 13 MAPE Summary Statistics for Combined versus Best ESM

Combined judgment vs. BESTESM						
The MEANS Procedure						
Analysis Variable : MAPE Mean Absolute Percent Error						
Comb Selected	N Obs	N	Mean	Std Dev	Minimum	Maximum
NO	12	12	6.8565985	2.9873591	3.4681708	12.4410259
YES	5	5	6.4816863	2.4262238	4.3111761	9.9540557

AUTOMATIC MODEL COMBINATION

For candidate models that are automatically generated from the HPFDIAGNOSE procedure, the addition of automatic model combination enables this technique to be used in the realm of large-scale automated time series forecasting for hierarchically structured data. This is the domain of the SAS High-Performance Forecasting product. The addition of a COMBINE statement to the PROC HPFDIAGNOSE statement block causes the HPFDIAGNOSE procedure to generate a combined model list for the individual time series models that it produces. In the context of hierarchical time series forecasting, this automatic generation feature becomes a convenient, if not essential, way to include an ensemble model in each BY group's model selection list. The forecast that results from the automatic combined model list enjoys no special treatment in the model selection process in the HPFENGINE procedure. It must stand on its own merits; when it proves best by the specified selection criterion, it is used as the selected forecast.

The following example demonstrates the use of automatic model combination for Sashelp.Pricedata. It generates candidate models to forecast the variable sale. The independent variable price is considered for use in the ARIMA models. Model selection uses the symmetric mean absolute percentage error (SMAPE) criterion over a holdout sample of six months.

```
proc hpfdiagnose
  data=sashelp.pricedata
  rep=work.rep
  outest=diagest
  holdout=4
  criterion=smape;
by region line product;
id date interval=month;
forecast sale;
```

```

input price/required=yes;
esm;
arimax method=scan;
combine method=avg
  encompass=ols(alpha=0.05)
  misspercent=25
  hormisspercent=10;
run;

```

Sashelp.Pricedata has a total of 17 BY groups at the specified hierarchy level. Figure 14 shows a summary of SMAPE statistics for the selected models over BY groups on the basis of whether a combined forecast was selected.

Figure 14 SMAPE Summary Statistics for Combined versus Single Model

Summary: Combination vs. Single model						
The MEANS Procedure						
Analysis Variable : SMAPE Mean Absolute Symmetric Percent Error						
Comb Selected	N Obs	N	Mean	Std Dev	Minimum	Maximum
NO	12	12	4.8474113	3.4928775	1.1789547	13.2997034
YES	5	5	4.3500032	1.2757722	2.7375724	5.8542727

Figure 15 shows a summary of SMAPE statistics for the BY groups where combined models were selected compared to the second-place model in the same BY group.

Figure 15 SMAPE Summary Statistics for Combined versus Best Alternate

Combination used: Combination versus Best Alternative						
The MEANS Procedure						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Comb	SMAPE	5	4.3500032	1.2757722	2.7375724	5.8542727
Second	SMAPE	5	4.9647232	1.4373792	3.2109125	7.0145075

With minimal SAS coding, the HPFDIAGNOSE procedure automatically generates ensemble models that are evaluated in the model selection process in the HPFENGINE procedure. The merits of these models are proven in this simple test case by the fact that they are chosen as the best model for 29% of the time series in the data set. When the combined forecast is selected as the best forecast, the average SMAPE is 0.61 better than the closest alternative. Figure 15 shows that the variability of the SMAPE criterion for the combined models is noticeably smaller than the SMAPE for the cases that do not select the combined forecast.

CONCLUSION

Ensemble models, in the form of forecast combination, are a significant new feature in SAS High-Performance Forecasting 4.1. Forecast combinations offer great potential benefit to the user with improved forecasts. A forecast combination process is described that embodies several distinct stages with the goal of improving the quality of the combined forecast through the elimination of poor-quality forecasts and redundant information. A variety of weight generation methods are provided to meet a wide range of user needs with a computationally robust forecast combination algorithm that gracefully handles missing values in the final set of combination candidates. Combined forecasts are included into a well-proven model selection mechanism in the HPFENGINE procedure. Model selection has been substantially extended and enhanced to use the forecast model selection graph concept, which enables ensemble models to use and operate on any forecast that is produced by any model or model list. Automatic generation of combined model lists is facilitated as an extension of the model generation capability of the HPFDIAGNOSE procedure. Several examples demonstrate the use of ensemble models in both the HPFENGINE and HPFDIAGNOSE procedures and the benefits obtained from their use.

References

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. Revised Edition. San Francisco: Holden-Day.

Makridakis, S. and Hibon, M. (2000). "The M3 Competition: Results, Conclusions, and Implications." *International Journal of Forecasting* 16, pp.451–476.

Minerals Management Service (2006). Monthly Oil and Gas Production, Gulf of Mexico, 1996–2006.

SAS Institute Inc. 2011. *SAS High-Performance Forecasting 4.1: User's Guide*. Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

The authors thank Mark Little and Kenneth Sanford of the Advanced Analytics Division at SAS Institute Inc. for their assistance in improving the organization and content of this paper. Thanks are also due to Anne Baxter and Ed Huddleston for their valuable editorial comments.

CONTACT INFORMATION

Ed Blair

SAS Institute Inc.

SAS Campus Drive, Cary, NC 27513

Work Phone: 919-531-6956, Fax: 919-677-4444

E-mail: Ed.Blair@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 registered trademarks or trademarks of their respective companies.