

Chapter 1

The HPF Procedure

Chapter Contents

OVERVIEW	5
GETTING STARTED	6
SYNTAX	9
Functional Summary	9
PROC HPF Statement	11
BY Statement	14
FORECAST Statement	14
ID Statement	19
IDM Statement	22
Smoothing Model Specification Options for IDM Statement	24
Smoothing Model Parameter Specification Options	27
Smoothing Model Forecast Bounds Options	27
DETAILS	28
Accumulation	28
Missing Value Interpretation	30
Diagnostic Tests	30
Model Selection	30
Transformations	31
Parameter Estimation	31
Missing Value Modeling Issues	31
Forecasting	31
Inverse Transformations	31
Statistics of Fit	32
Forecast Summation	32
Comparison to the Time Series Forecasting System	32
Data Set Output	32
OUT= Data Set	32
OUTEST= Data Set	33
OUTFOR= Data Set	33
OUTSTAT= Data Set	34
OUTSUM= Data Set	35
OUTSEASON= Data Set	36
OUTTREND= Data Set	37
Printed Output	37

4 ♦ Chapter 1. The HPF Procedure

ODS Table Names	39
ODS Graphics (Experimental)	39
EXAMPLES	42
Example 1.1. Automatic Forecasting of Time Series Data	42
Example 1.2. Automatic Forecasting of Transactional Data	44
Example 1.3. Specifying the Forecasting Model	45
Example 1.4. Extending the Independent Variables for Multivariate Forecasts	46
Example 1.5. Forecasting Intermittent Time Series Data	47
Example 1.6. Illustration of ODS Graphics (Experimental)	49
REFERENCES	52

Chapter 1

The HPF Procedure

Overview

The HPF (High-Performance Forecasting) procedure provides a quick and automatic way to generate forecasts for many time series or transactional data in one step. The procedure can forecast millions of series at a time, with the series organized into separate variables or across BY groups.

- For typical time series, you can use the following smoothing models:
 - Simple
 - Double
 - Linear
 - Damped Trend
 - Seasonal
 - Winters Method (additive and multiplicative)
- Additionally, transformed versions of these models are provided:
 - Log
 - Square Root
 - Logistic
 - Box-Cox
- For intermittent time series (series where a large number of values are zero-valued), you can use an intermittent demand model such as Croston's Method and Average Demand Model.

Experimental graphics are now available with the HPF procedure. For more information, see the [“ODS Graphics”](#) section on page 39.

All parameters associated with the forecast model are optimized based on the data. Optionally, the HPF procedure can select the appropriate smoothing model for you using holdout sample analysis based on one of several model selection criteria.

The HPF procedure writes the time series extrapolated by the forecasts, the series summary statistics, the forecasts and confidence limits, the parameter estimates, and the fit statistics to output data sets. The HPF procedure optionally produces printed output for these results utilizing the Output Delivery System (ODS).

The HPF procedure can forecast both time series data, whose observations are equally spaced by a specific time interval (e.g., monthly, weekly), or transactional data, whose observations are not spaced with respect to any particular time interval. Internet, inventory, sales, and similar data are typical examples of transactional data. For transactional data, the data is accumulated based on a specified time interval to form a

time series. The HPF procedure can also perform trend and seasonal analysis on transactional data.

Additionally, the Time Series Forecasting System of SAS/ETS software can be used to interactively develop forecasting models, estimate the model parameters, evaluate the models' ability to forecast, and display these results graphically. Refer to Chapter 33, "Overview of the Time Series Forecasting System," (*SAS/ETS User's Guide*) for details.

Also, the EXPAND procedure can be used for the frequency conversion and transformations of time series. Refer to Chapter 16, "The EXPAND Procedure," (*SAS/ETS User's Guide*) for details.

Getting Started

The HPF procedure is simple to use for someone who is new to forecasting, and yet at the same time it is powerful for the experienced professional forecaster who needs to generate a large number of forecasts automatically. It can provide results in output data sets or in other output formats using the Output Delivery System (ODS). The following examples are more fully illustrated in the "Examples" section on page 42.

Given an input data set that contains numerous time series variables recorded at a specific frequency, the HPF procedure can automatically forecast the series as follows:

```
PROC HPF DATA=<input-data-set> OUT=<output-data-set>;
  ID <time-ID-variable> INTERVAL=<frequency>;
  FORECAST <time-series-variables>;
RUN;
```

For example, suppose that the input data set SALES contains numerous sales data recorded monthly, the variable that represents time is DATE, and the forecasts are to be recorded in the output data set NEXTYEAR. The HPF procedure could be used as follows:

```
proc hpf data=sales out=nextyear;
  id date interval=month;
  forecast _ALL_;
run;
```

The above statements automatically select the best fitting model, generate forecasts for every numeric variable in the input data set (SALES) for the next twelve months, and store these forecasts in the output data set (NEXTYEAR). Other output data sets can be specified to store the parameter estimates, forecasts, statistics of fit, and summary data.

If you want to print the forecasts using the Output Delivery System (ODS), then you need to add PRINT=FORECASTS:

```
proc hpf data=sales out=nextyear print=forecasts;
  id date interval=month;
  forecast _ALL_;
run;
```

Other results can be specified to output the parameter estimates, forecasts, statistics of fit, and summary data using ODS.

The HPF procedure can forecast both time series data, whose observations are equally spaced by a specific time interval (e.g., monthly, weekly), or transactional data, whose observations are not spaced with respect to any particular time interval.

Given an input data set containing transactional variables not recorded at any specific frequency, the HPF procedure accumulates the data to a specific time interval and forecasts the accumulated series as follows:

```
PROC HPF DATA=<input-data-set> OUT=<output-data-set>;
  ID <time-ID-variable> INTERVAL=<frequency>
  ACCUMULATE=<accumulation>;
  FORECAST <time-series-variables>;
RUN;
```

For example, suppose that the input data set **WEBSITES** contains three variables (**BOATS**, **CARS**, **PLANES**), that are Internet data recorded on no particular time interval, and the variable that represents time is **TIME**, which records the time of the Web hit. The forecasts for the total daily values are to be recorded in the output data set **NEXTWEEK**. The HPF procedure could be used as follows:

```
proc hpf data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast boats cars planes;
run;
```

The above statements accumulate the data into a daily time series and automatically generate forecasts for the **BOATS**, **CARS**, and **PLANES** variables in the input data set (**WEBSITES**) for the next seven days and store the forecasts in the output data set (**NEXTWEEK**).

The HPF procedure can specify a particular forecast model or select from several candidate models based on a selection criterion. The HPF procedure also supports transformed models and holdout sample analysis.

Using the previous **WEBSITES** example, suppose that you want to forecast the **BOATS** variable using the best seasonal forecasting model that minimizes the mean absolute percent error (MAPE), the **CARS** variable using the best nonseasonal forecasting model that minimizes the mean square error (MSE) using holdout sample analysis on the last five days, and the **PLANES** variable using the Log Winters Method (additive). The HPF procedure could be used as follows:

```

proc hpf data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast boats / model=bests select=mape;
  forecast cars / model=bestn select=mse holdout=5;
  forecast planes / model=addwinters transform=log;
run;

```

The above statements demonstrate how each variable in the input data set can be modeled differently and how several candidate models can be specified and selected based on holdout sample analysis or the entire range of data.

The HPF procedure is also useful in extending independent variables in (auto) regression models where future values of the independent variable are needed to predict the dependent variable.

Using the WEBSITES example, suppose that you want to forecast the ENGINES variable using the BOATS, CARS, and PLANES variable as regressor variables. Since future values of the BOATS, CARS, and PLANES are needed, the HPF procedure can be used to extend these variables in the future:

```

proc hpf data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast engines / model=none;
  forecast boats / model=bests select=mape;
  forecast cars / model=bestn select=mse holdout=5;
  forecast planes / model=addwinters transform=log;
run;

proc autoreg data= nextweek;
  model engines = boats cars planes;
  output out=enginehits p=predicted;
run;

```

The above HPF procedure statements generate forecasts for BOATS, CARS, and PLANES in the input data set (WEBSITES) for the next seven days and extend the variable ENGINES with missing values. The output data set (NEXTWEEK) of the PROC HPF statement is used as an input data set for the PROC AUTOREG statement. The output data set of PROC AUTOREG contains the forecast of the variable ENGINE based on the regression model with the variables BOATS, CARS, and PLANES as regressors. See the AUTOREG procedure for details on autoregression.

The HPF procedure can also forecast intermittent time series (series where a large number of values are zero-valued). Typical time series forecasting techniques are less effective in forecasting intermittent time series.

For example, suppose that the input data set INVENTORY contains three variables (TIRES, HUBCAPS, LUGBOLTS) that are demand data recorded on no particular time interval, the variable that represents time is DATE, and the forecasts for the total weekly values are to be recorded in the output data set NEXTMONTH. The models requested are intermittent demand models, which can be specified as MODEL=IDM

option. Two intermittent demand models are compared, Croston Model and Average Demand Model. The HPF procedure could be used as follows:

```
proc hpf data=inventory out=nextmonth lead=4 print=forecasts;
  id date interval=week accumulate=total;
  forecast tires hubcaps lugbolts / model=idm;
run;
```

In the above example, the total demand for inventory items is accumulated on a weekly basis and forecasts are generated that recommend future stocking levels.

Syntax

The following statements are used with the HPF procedure.

```
PROC HPF options;  
  BY variables;  
  FORECAST variable-list / options;  
  ID variable INTERVAL= interval options;  
  IDM options;
```

Functional Summary

The statements and options controlling the HPF procedure are summarized in the following table.

Description	Statement	Option
Statements		
specify BY-group processing	BY	
specify variables to forecast	FORECAST	
specify the time ID variable	ID	
specify intermittent demand model	IDM	
Data Set Options		
specify the input data set	PROC HPF	DATA=
specify to output forecasts only	PROC HPF	NOOUTALL
specify the output data set	PROC HPF	OUT=
specify parameter output data set	PROC HPF	OUTEST=
specify forecast output data set	PROC HPF	OUTFOR=
specify seasonal statistics output data set	PROC HPF	OUTSEASON=
specify statistics output data set	PROC HPF	OUTSTAT=
specify summary output data set	PROC HPF	OUTSUM=
specify trend statistics output data set	PROC HPF	OUTTREND=
replace actual values held back	FORECAST	REPLACEBACK

Description	Statement	Option
replace missing values	FORECAST	REPLACEMISSING
use forecast forecast value to append	FORECAST	USE=
Accumulation and Seasonality Options		
specify accumulation frequency	ID	INTERVAL=
specify length of seasonal cycle	PROC HPF	SEASONALITY=
specify interval alignment	ID	ALIGN=
specify time ID variable values are not sorted	ID	NOTSORTED
specify starting time ID value	ID	START=
specify ending time ID value	ID	END=
specify accumulation statistic	ID, FORECAST	ACCUMULATE=
specify missing value interpretation	ID, FORECAST	SETMISSING=
specify zero value interpretation	ID, FORECAST	ZEROMISS=
Forecasting Horizon, Holdout, Holdback Options		
specify data to hold back	PROC HPF	BACK=
specify forecast holdout sample size	FORECAST	HOLDOUT=
specify forecast holdout sample percent	FORECAST	HOLDOUTPCT=
specify forecast horizon or lead	PROC HPF	LEAD=
specify horizon to start summation	PROC HPF	STARTSUM=
Forecasting Model and Selection Options		
specify confidence limit width	FORECAST	ALPHA=
specify intermittency	FORECAST	INTERMITTENT=
specify forecast model	FORECAST	MODEL=
specify median forecasts	FORECAST	MEDIAN
specify backcast initialization	FORECAST	NBACKCAST=
specify seasonality test	FORECAST	SEASONTEST=
specify model selection criterion	FORECAST	SELECT=
specify model transformation	FORECAST	TRANSFORM=
Intermittent Demand Model Options		
specify model for average demand	IDM	AVERAGE=
specify the base value	IDM	BASE=
specify model for demand intervals	IDM	INTERVAL=
specify model for demand sizes	IDM	SIZE=
Printing and Plotting Control Options		
specify graphical output	PROC HPF	PLOT=
specify printed output	PROC HPF	PRINT=
specify detailed printed output	PROC HPF	PRINTDETAILS

Description	Statement	Option
Miscellaneous Options		
specify that analysis variables are processed in sorted order	PROC HPF	<code>SORTNAMES</code>
limits error and warning messages	PROC HPF	<code>MAXERROR=</code>

PROC HPF Statement

PROC HPF *options*;

The following options can be used in the PROC HPF statement.

BACK= *n*

specifies the number of observations before the end of the data that the multistep forecasts are to begin. The default is BACK=0.

DATA= *SAS-data-set*

names the SAS data set containing the input data for the procedure to forecast. If the DATA= option is not specified, the most recently created SAS data set is used.

LEAD= *n*

specifies the number of periods ahead to forecast (forecast lead or horizon). The default is LEAD=12.

The LEAD= value is relative to the last observation in the input data set and not to the last nonmissing observation of a particular series. Thus, if a series has missing values at the end, the actual number of forecasts computed for that series will be greater than the LEAD= value.

MAXERROR= *number*

limits the number of warning and error messages produced during the execution of the procedure to the specified value. The default is MAXERRORS=50. This option is particularly useful in BY-group processing where it can be used to suppress the recurring messages.

NOOUTALL

specifies that only forecasts are written to the OUT= and OUTFOR= data sets. The NOOUTALL option includes only the final forecast observations in the output data sets, not the one-step forecasts for the data before the forecast period.

The OUT= and OUTFOR= data set will only contain the forecast results starting at the next period following the last observation to the forecast horizon specified by the LEAD= option.

OUT= *SAS-data-set*

names the output data set to contain the forecasts of the variables specified in the subsequent FORECAST statements. If an ID variable is specified, it will also be included in the OUT= data set. The values are accumulated based on the ACCUMULATE=

option and forecasts are appended to these values based on the FORECAST statement USE= option. The OUT= data set is particularly useful in extending the independent variables when forecasting dependent series associated with (auto) regression models. If the OUT= option is not specified, a default output data set DATAn is created. If you do not want the OUT= data set created, then use OUT=_NULL_.

OUTEST= SAS-data-set

names the output data set to contain the model parameter estimates and the associated test statistics and probability values. The OUTEST= data set is particularly useful for evaluating the significance of the model parameters and understanding the model dynamics.

OUTFOR= SAS-data-set

names the output data set to contain the forecast time series components (actual, predicted, lower confidence limit, upper confidence limit, prediction error, prediction standard error). The OUTFOR= data set is particularly useful for displaying the forecasts in tabular or graphical form.

OUTSEASON= SAS-data-set

names the output data set to contain the seasonal statistics. The statistics are computed for each season as specified by the ID statement INTERVAL= option or the SEASONALITY= option. The OUTSEASON= data set is particularly useful when analyzing transactional data for seasonal variations.

OUTSTAT= SAS-data-set

names the output data set to contain the statistics of fit (or goodness-of-fit statistics). The OUTSTAT= data set is particularly useful for evaluating how well the model fits the series. The statistics of fit are based on the entire range of the time series regardless of whether the HOLDOUT= option is specified.

OUTSUM= SAS-data-set

names the output data set to contain the summary statistics and the forecast summation. The summary statistics are based on the accumulated time series when the ACCUMULATE= or SETMISSING= options are specified. The forecast summations are based on the LEAD=, STARTSUM=, and USE= options. The OUTSUM= data set is particularly useful when forecasting large numbers of series and a summary of the results are needed.

OUTTREND= SAS-data-set

names the output data set to contain the trend statistics. The statistics are computed for each time period as specified by the ID statement INTERVAL= option. The OUTTREND= data set is particularly useful when analyzing transactional data for trends.

PRINT= option | (options)

specifies the printed output desired. By default, the HPF procedure produces no printed output. The following printing options are available:

ESTIMATES	prints the results of parameter estimation. (OUTEST= data set)
FORECASTS	prints the forecasts. (OUTFOR= data set)

PERFORMANCE	prints the performance statistics for each forecast.
PERFORMANCESUMMARY	prints the performance summary for each BY group.
PERFORMANCEOVERALL	prints the performance summary for all of the BY groups.
SEASONS	prints the seasonal statistics. (OUTSEASON= data set)
STATISTICS	prints the statistics of fit. (OUTSTAT= data set)
STATES	prints the backcast, initial, and final states.
SUMMARY	prints the summary statistics for the accumulated time series. (OUTSUM= data set)
TRENDS	prints the trend statistics. (OUTTREND= data set)
ALL	Same as PRINT=(ESTIMATES FORECASTS STATISTICS SUMMARY). PRINT=(ALL,TRENDS,SEASONS) prints all of the options listed above.

For example, PRINT=FORECASTS prints the forecasts, PRINT=(ESTIMATES, FORECASTS) prints the parameter estimates and the forecasts, and PRINT=ALL prints all of the above output.

The PRINT= option produces printed output for these results utilizing the Output Delivery System (ODS). The PRINT= option produces results similar to the data sets listed next to the above options in parenthesis.

PRINTDETAILS

specifies that output requested with the PRINT= option be printed in greater detail.

SEASONALITY= *number*

specifies the length of the seasonal cycle. For example, SEASONALITY=3 means that every group of three observations forms a seasonal cycle. The SEASONALITY= option is applicable only for seasonal forecasting models. By default, the length of the seasonal cycle is one (no seasonality) or the length implied by the INTERVAL= option specified in the ID statement. For example, INTERVAL=MONTH implies that the length of the seasonal cycle is twelve.

SORTNAMES

specifies that the variables specified in the FORECAST statements are processed in sorted order.

STARTSUM= *n*

specifies the starting forecast lead (or horizon) for which to begin summation of the forecasts specified by the LEAD= option. The STARTSUM= value must be less than the LEAD= value. The default is STARTSUM=1, that is, the sum from the one-step ahead forecast to the multistep forecast specified by the LEAD= option.

The prediction standard errors of the summation of forecasts takes into account the correlation between the multistep forecasts. The DETAILS section describes the STARTSUM= option in more detail.

BY Statement

BY *variables*;

A BY statement can be used with PROC HPF to obtain separate analyses for groups of observations defined by the BY variables.

FORECAST Statement

FORECAST *variable-list / options*;

The FORECAST statement lists the numeric variables in the DATA= data set whose accumulated values represent time series to be modeled and forecast. The options specify which forecast model is to be used or how the forecast model is selected from several possible candidate models.

A data set variable can be specified in only one FORECAST statement. Any number of FORECAST statements can be used. The following options can be used with the FORECAST statement.

ACCUMULATE= *option*

specifies how the data set observations are accumulated within each time period for the variables listed in the FORECAST statement. If the ACCUMULATE= option is not specified in the FORECAST statement, accumulation is determined by the ACCUMULATE= option of the ID statement. See the ID statement ACCUMULATE= option for more details.

ALPHA= *number*

specifies the significance level to use in computing the confidence limits of the forecast. The ALPHA=value must be between 0 and 1. The default is ALPHA=0.05, which produces 95% confidence intervals.

HOLDOUT= *n*

specifies the size of the holdout sample to be used for model selection. The holdout sample is a subset of actual time series ending at the last nonmissing observation. If the ACCUMULATE= option is specified, the holdout sample is based on the accumulated series. If the holdout sample is not specified, the full range of the actual time series is used for model selection.

For each candidate model specified, the holdout sample is excluded from the initial model fit and forecasts are made within the holdout sample time range. Then, for each candidate model specified, the statistic of fit specified by the SELECT= option is computed using only the observations in the holdout sample. Finally, the candidate model, which performs best in the holdout sample, based on this statistic, is selected to forecast the actual time series.

The HOLDOUT= option is only used to select the best forecasting model from a list of candidate models. After the best model is selected, the full range of the actual time series is used for subsequent model fitting and forecasting. It is possible that one model will outperform another model in the holdout sample but perform less well when the entire range of the actual series is used.

If MODEL=BESTALL and HOLDOUT= options are used together, the last one hundred observations are used to determine whether the series is intermittent. If the series determined not to be intermittent, holdout sample analysis will be used to select the smoothing model.

HOLDOUTPCT= *number*

specifies the size of the holdout sample as a percentage of the length of the time series. If HOLDOUT=5 and HOLDOUTPCT=10, the size of the holdout sample is $\min(5, 0.1T)$ where T is the length of the time series with beginning and ending missing values removed. The default is 100 (100%).

INTERMITTENT= *number*

specifies a number greater than one which is used to determine whether or not a time series is intermittent. If the average demand interval is greater than this number then the series is assumed to be intermittent. This option is used with MODEL=BESTALL option. The default is INTERMITTENT=1.25.

MEDIAN

specifies that the median forecast values are to be estimated. Forecasts can be based on the mean or median. By default the mean value is provided. If no transformation is applied to the actual series using the TRANSFORM= option, the mean and median forecast values are identical.

MODEL= *model-name*

specifies the forecasting model to be used to forecast the actual time series. A single model can be specified or a group of candidate models can be specified. If a group of models is specified, the model used to forecast the accumulated time series is selected based on the SELECT= option and the HOLDOUT= option. The default is MODEL=BEST. The following forecasting models are provided:

NONE	No forecast. The accumulated time series is appended with missing values in the OUT= data set. This option is particularly useful when the results stored in the OUT= data set are subsequently used in (auto) regression analysis where forecasts of the independent variables are needed to forecast the dependent variable.
SIMPLE	Simple (Single) Exponential Smoothing
DOUBLE	Double (Brown) Exponential Smoothing
LINEAR	Linear (Holt) Exponential Smoothing
DAMPTREND	Damped Trend Exponential Smoothing
SEASONAL	Seasonal Exponential Smoothing
WINTERS	Winters Multiplicative Method
ADDWINTERS	Winters Additive Method
BEST	Best Candidate Smoothing Model (SIMPLE, DOUBLE, LINEAR, DAMPTREND, SEASONAL, WINTERS, ADDWINTERS)
BESTN	Best Candidate Nonseasonal Smoothing Model (SIMPLE, DOUBLE, LINEAR, DAMPTREND)

BESTS	Best Candidate Seasonal Smoothing Model (SEASONAL, WINTERS, ADDWINTERS)
IDM CROSTON	Intermittent Demand Model such as Croston's Method or Average Demand Model. An intermittent time series is one whose values are mostly zero.
BESTALL	Best Candidate Model (IDM, BEST)

The BEST, BESTN, and BESTS options specify a group of models by which the HOLDOUT= option and SELECT= option are used to select the model used to forecast the accumulated time series based on holdout sample analysis. Transformed versions of the above smoothing models can be specified using the TRANSFORM= option.

The BESTALL option specifies that if the series is intermittent, an intermittent demand model such as Croston's Method or Average Demand Model (MODEL=IDM) is selected; otherwise, the best smoothing model is selected (MODEL=BEST). Intermittency is determined by the INTERMITTENT= option.

The documentation for [Chapter 2, "Forecasting Process Details,"](#) describes the above smoothing models and intermittent models in greater detail.

NBACKCAST= *n*

specifies the number of observations used to initialize the backcast states. The default is the entire series.

REPLACEBACK

specifies that actual values excluded by the BACK= option are replaced with one-step-ahead forecasts in the OUT= data set.

REPLACEMISSING

specifies that embedded missing actual values are replaced with one-step-ahead forecasts in the OUT= data set.

SEASONTEST= *option*

specifies the options related to the seasonality test. This option is used with MODEL=BEST and MODEL=BESTALL option.

The following options are provided:

SEASONTEST=NONE no test

SEASONTEST=(SIGLEVEL=number) Significance probability value.

Series with strong seasonality have small test probabilities. SEASONTEST=(SIGLEVEL=0) always implies seasonality. SEASONTEST=(SIGLEVEL=1) always implies no seasonality. The default is SEASONTEST=(SIGLEVEL=0.01).

SELECT= *option*

specifies the model selection criterion (statistic of fit) to be used to select from several candidate models. This option would often be used in conjunction with the

HOLDOUT= option. The default is SELECT=RMSE. The following statistics of fit are provided:

SSE	Sum of Square Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
UMSE	Unbiased Mean Square Error
URMSE	Unbiased Root Mean Square Error
MAXPE	Maximum Percent Error
MINPE	Minimum Percent Error
MPE	Mean Percent Error
MAPE	Mean Absolute Percent Error
MDAPE	Median Percent Error
GMAPE	Geometric Mean Percent Error
MINPPE	Minimum Predictive Percent Error
MAXPPE	Maximum Predictive Percent Error
MSPPE	Mean Predictive Percent Error
MAPPE	Symmetric Mean Absolute Predictive Percent Error
MDAPPE	Median Predictive Percent Error
GMAPPE	Geometric Mean Predictive Percent Error
MINSPE	Minimum Symmetric Percent Error
MAXSPE	Maximum Symmetric Percent Error
MSPE	Mean Symmetric Percent Error
SMAPE	Symmetric Mean Absolute Percent Error
MDASPE	Median Symmetric Percent Error
GMASPE	Geometric Mean Symmetric Percent Error
MINRE	Minimum Relative Error
MAXRE	Maximum Relative Error
MRE	Mean Relative Error
MRAE	Mean Relative Absolute Error
MDRAE	Median Relative Absolute Error
GMRAE	Geometric Mean Relative Absolute Error
MAXERR	Maximum Error
MINERR	Minimum Error
ME	Mean Error
MAE	Mean Absolute Error
RSQUARE	R-Square

ADJRSQ	Adjusted R-Square
AADJRSQ	Amemiya's Adjusted R-Square
RWRSQ	Random Walk R-Square
AIC	Akaike Information Criterion
SBC	Schwarz Bayesian Information Criterion
APC	Amemiya's Prediction Criterion

SETMISSING= option | number

specifies how missing values (either actual or accumulated) are assigned in the accumulated time series for variables listed in the FORECAST statement. If the SETMISSING= option is not specified in the FORECAST statement, missing values are set based on the SETMISSING= option of the ID statement. See the ID statement SETMISSING= option for more details.

TRANSFORM= option

specifies the time series transformation to be applied to the actual time series. The following transformations are provided:

NONE	No transformation is applied. This option is the default.
LOG	Logarithmic transformation
SQRT	Square-root transformation
LOGISTIC	Logistic transformation
BOXCOX(<i>n</i>)	Box-Cox transformation with parameter number where number is between -5 and 5
AUTO	Automatically choose between NONE and LOG based on model selection criteria.

When the TRANSFORM= option is specified the time series must be strictly positive. Once the time series is transformed, the model parameters are estimated using the transformed series. The forecasts of the transformed series are then computed, and finally, the transformed series forecasts are inverse transformed. The inverse transform produces either mean or median forecasts depending on whether the MEDIAN option is specified.

The TRANSFORM= option is not applicable when MODEL=IDM is specified.

USE= option

specifies which forecast values are appended to the actual values in the OUT= and OUTSUM= data sets. The following USE= options are provided:

PREDICT	The predicted values are appended to the actual values. This option is the default.
LOWER	The lower confidence limit values are appended to the actual values.

UPPER The upper confidence limit values are appended to the actual values.

Thus, the **USE=** option enables the **OUT=** and **OUTSUM=** data sets to be used for worst/best/average/median case decisions.

ZEROMISS= option

specifies how beginning and/or ending zero values (either actual or accumulated) are interpreted in the accumulated time series for variables listed in the **FORECAST** statement. If the **ZEROMISS=** option is not specified in the **FORECAST** statement, missing values are set based on the **ZEROMISS=** option of the **ID** statement. See the **ID** statement **ZEROMISS=** option for more details.

ID Statement

ID variable *INTERVAL= interval options;*

The **ID** statement names a numeric variable that identifies observations in the input and output data sets. The **ID** variable's values are assumed to be SAS date, time, or datetime values. In addition, the **ID** statement specifies the (desired) frequency associated with the actual time series. The **ID** statement options also specify how the observations are accumulated and how the time **ID** values are aligned to form the actual time series. The information specified affects all variables specified in subsequent **FORECAST** statements. If the **ID** statement is specified, the **INTERVAL=** option must also be specified. If an **ID** statement is not specified, the observation number, with respect to the **BY** group, is used as the time **ID**.

The following options can be used with the **ID** statement.

ACCUMULATE= option

specifies how the data set observations are accumulated within each time period. The frequency (width of each time interval) is specified by the **INTERVAL=** option. The **ID** variable contains the time **ID** values. Each time **ID** variable value corresponds to a specific time period. The accumulated values form the actual time series, which is used in subsequent model fitting and forecasting.

The **ACCUMULATE=** option is particularly useful when there are zero or more than one input observations coinciding with a particular time period (e.g., transactional data). The **EXPAND** procedure offers additional frequency conversions and transformations that can also be useful in creating a time series.

The following options determine how the observations are accumulated within each time period based on the **ID** variable and the frequency specified by the **INTERVAL=** option:

NONE No accumulation occurs; the **ID** variable values must be equally spaced with respect to the frequency. This is the default option.

TOTAL	Observations are accumulated based on the total sum of their values.
AVERAGE AVG	Observations are accumulated based on the average of their values.
MINIMUM MIN	Observations are accumulated based on the minimum of their values.
MEDIAN MED	Observations are accumulated based on the median of their values.
MAXIMUM MAX	Observations are accumulated based on the maximum of their values.
N	Observations are accumulated based on the number of nonmissing observations.
NMISS	Observations are accumulated based on the number of missing observations.
NOBS	Observations are accumulated based on the number of observations.
FIRST	Observations are accumulated based on the first of their values.
LAST	Observations are accumulated based on the last of their values.
STDDEV STD	Observations are accumulated based on the standard deviation of their values.
CSS	Observations are accumulated based on the corrected sum of squares of their values.
USS	Observations are accumulated based on the uncorrected sum of squares of their values.

If the ACCUMULATE= option is specified, the SETMISSING= option is useful for specifying how accumulated missing values are treated. If missing values should be interpreted as zero, then SETMISSING=0 should be used. The DETAILS section describes accumulation in greater detail.

ALIGN= option

controls the alignment of SAS dates used to identify output observations. The ALIGN= option accepts the following values: BEGINNING | BEG | B, MIDDLE | MID | M, and ENDING | END | E. BEGINNING is the default.

END= option

specifies a SAS date, datetime, or time value that represents the end of the data. If the last time ID variable value is less than the END= value, the series is extended with missing values. If the last time ID variable value is greater than the END= value, the series is truncated. For example, END="&sysdate"D uses the automatic macro variable SYSDATE to extend or truncate the series to the current date. This option and the START= option can be used to ensure that data associated with each BY group contains the same number of observations.

INTERVAL= *interval*

specifies the frequency of the input time series. For example, if the input data set consists of quarterly observations, then INTERVAL=QTR should be used. If the SEASONALITY= option is not specified, the length of the seasonal cycle is implied from the INTERVAL= option. For example, INTERVAL=QTR implies a seasonal cycle of length 4. If the ACCUMULATE= option is also specified, the INTERVAL= option determines the time periods for the accumulation of observations.

The basic intervals are YEAR, SEMIYEAR, QTR, MONTH, SEMIMONTH, TENDAY, WEEK, WEEKDAY, DAY, HOUR, MINUTE, SECOND. Refer to SAS/ETS User's Guide chapter on Date Interval, Foremats, and Functions for the intervals that can be specified.

NOTSORTED

specifies that the time ID values are not in sorted order. The HPF procedure will sort the data with respect to the time ID prior to analysis.

SETMISSING= *option* | *number*

specifies how missing values (either actual or accumulated) are assigned in the accumulated time series. If a number is specified, missing values are set to number. If a missing value indicates an unknown value, this option should not be used. If a missing value indicates no value, a SETMISSING=0 should be used. You would typically use SETMISSING=0 for transactional data because no recorded data usually implies no activity. The following options can also be used to determine how missing values are assigned:

MISSING	Missing values are set to missing. This is the default option.
AVERAGE AVG	Missing values are set to the accumulated average value.
MINIMUM MIN	Missing values are set to the accumulated minimum value.
MEDIAN MED	Missing values are set to the accumulated median value.
MAXIMUM MAX	Missing values are set to the accumulated maximum value.
FIRST	Missing values are set to the accumulated first nonmissing value.
LAST	Missing values are set to the accumulated last nonmissing value.
PREVIOUS PREV	Missing values are set to the previous accumulated nonmissing value. Missing values at the beginning of the accumulated series remain missing.
NEXT	Missing values are set to the next accumulated nonmissing value. Missing values at the end of the accumulated series remain missing.

If SETMISSING=MISSING is specified and the MODEL= option specifies a smoothing model, the missing observations are smoothed over. If MODEL=IDM

is specified, missing values are assumed to be periods of no demand, that is, SETMISSING=MISSING is equivalent to SETMISSING=0.

START= option

specifies a SAS date, datetime, or time value that represents the beginning of the data. If the first time ID variable value is greater than the START= value, the series is prepended with missing values. If the first time ID variable value is less than the START= value, the series is truncated. This option and the END= option can be used to ensure that data associated with each by group contains the same number of observations.

ZEROMISS= option

specifies how beginning and/or ending zero values (either actual or accumulated) are interpreted in the accumulated time series. The following options can also be used to determine how beginning and/or ending zero values are assigned:

NONE	Beginning and/or ending zeros unchanged. This is the default.
LEFT	Beginning zeros are set to missing.
RIGHT	Ending zeros are set to missing.
BOTH	Both beginning and ending zeros are set to missing.

If the accumulated series is all missing and/or zero the series is not changed.

IDM Statement

IDM options;

The IDM statement is used to specify an intermittent demand model. An intermittent demand series can be analyzed in two ways: individually modeling both demand interval and size component or jointly modeling these components using the average demand component (demand size divided by demand interval). The IDM statement specifies the two smoothing models to be used to forecast the demand interval component (INTERVAL= option) and the demand size component (SIZE= option), or specifies the single smoothing model to be used to forecast the average demand component (AVERAGE= option). Therefore, two smoothing models (INTERVAL= and SIZE= options) must be specified or one smoothing model (AVERAGE= option) must be specified. Only one statement can be specified.

The following examples illustrate typical uses of the IDM statement:

```
/* default specification */
idm;

/* demand interval model only specification */
idm interval=(transform=log);

/* demand size model only specification */
idm size=(method=linear);
```

```

/* Croston's Method */
idm interval=(method=simple)
    size      =(method=simple);

/* Log Croston's Method */
idm interval=(method=simple transform=log)
    size      =(method=simple transform=log);

/* average demand model specification */
idm average=(method=bestn);

```

The default specification uses both the INTERVAL= option and SIZE= option defaults for the decomposed (Croston's) demand model and the AVERAGE= option defaults for the average demand model.

The following example illustrates how to automatically choose the decomposed demand model using MAPE as the model selection criterion:

```

idm interval=(method=simple transform=auto select=mape)
    size      =(method=simple transform=auto select=mape);
forecast sales / model=idm select=mape;

```

The preceding fits two forecast models (simple and log simple exponential smoothing) to both the demand interval and size components. The forecast model that results in the lowest in-sample MAPE for each component is used to forecast the component.

The following example illustrates how to automatically choose the average demand model using MAPE as the model selection criterion:

```

idm average=(method=simple transform=auto select=mape);
forecast sales / model=idm;

```

The preceding fits two forecast models (simple and log simple exponential smoothing) to the average demand component. The forecast model that results in the lowest in-sample MAPE is used to forecast the component.

Combining the above two examples, the following example illustrates how to automatically choose between the decomposed demand model and the average demand model using MAPE as the model selection criterion:

```

idm interval=(method=simple transform=auto select=mape)
    size      =(method=simple transform=auto select=mape)
    average   =(method=simple transform=auto select=mape);
forecast sales / model=idm select=mape;

```

The preceding automatically selects between the decomposed demand model and the average demand model as described previously. The forecast model that results in the lowest in-sample MAPE is used to forecast the series.

The following options can be specified in the IDM statement:

INTERVAL=(smoothing-model-options)

specifies the smoothing model used to forecast the demand interval component. See smoothing model specification options described below.

SIZE=(smoothing-model-options)

specifies the smoothing model used to forecast the demand size component. See smoothing model specification options described below.

AVERAGE=(smoothing-model-options)

specifies the smoothing model used to forecast the demand average component. See smoothing model specification options described below.

BASE=AUTO | number

specifies the base value of the time series used to determine the demand series components. The demand series components are determined based on the departures from this base value. If a base value is specified, this value is used to determine the demand series components. If BASE=AUTO is specified, the time series properties are used to automatically adjust the time series. For the common definition of Croston's Method use BASE=0 which defines departures based on zero. The default is BASE=0.

Given a time series, y_t , and base value, b , the time series is adjusted by the base value to create the base adjusted time series, $x_t = y_t - b$. Demands are assumed to occur when the base adjusted series is nonzero (or when the time series, y_t , departs from the base value, b).

When BASE=AUTO, the base value is automatically determined by the time series median, minimum, and maximum value and the INTERMITTENT= option value of the FORECAST statement.

Smoothing Model Specification Options for IDM Statement

The smoothing model options describe how to forecast the demand interval, size, and average demand components (INTERVAL= option, SIZE= option, and AVERAGE= option).

If the smoothing model options are not specified, the following are the defaults for the demand interval, size, and average components.

```
interval=(transform=auto method=bestn
          levelrest=(0.0001 0.9999)
          trendrest=(0.0001 0.9999)
          damprest=(0.0001 0.9999) select=rmse bounds=(1,.));

size      =(transform=auto method=bestn
          levelrest=(0.0001 0.9999)
          trendrest=(0.0001 0.9999)
          damprest=(0.0001 0.9999) select=rmse);

average   =(transform=auto method=bestn
          levelrest=(0.0001 0.9999)
```

```
trendrest=(0.0001 0.9999)
damprest =(0.0001 0.9999) select=rmse);
```

The above smoothing model options provide the typical automation in intermittent demand model selection.

The following describes the smoothing model options:

TRANSFORM= *option*

specifies the time series transformation to be applied to the demand component. The following transformations are provided:

NONE	No transformation is applied.
LOG	Logarithmic transformation
SQRT	Square-root transformation
LOGISTIC	Logistic transformation
BOXCOX(<i>n</i>)	Box-Cox transformation with parameter number where number is between -5 and 5
AUTO	Automatically choose between NONE and LOG based on model selection criteria. This option is the default.

When the TRANSFORM= option is specified, the demand component must be strictly positive. Once the demand component is transformed, the model parameters are estimated using the transformed component. The forecasts of the transformed component are then computed, and finally, the transformed component forecasts are inverse transformed. The inverse transform produces either mean or median forecasts depending on whether the MEDIAN option is specified.

MEDIAN

specifies that the median forecast values are to be estimated. Forecasts can be based on the mean or median. By default the mean value is provided. If no transformation is applied to the actual series using the TRANSFORM= option, the mean and median component forecast values are identical.

METHOD= *method-name*

specifies the forecasting model to be used to forecast the demand component. A single model can be specified or a group of candidate models can be specified. If a group of models is specified, the model used to forecast the accumulated time series is selected based on the SELECT= option of the IDM statement and the HOLDOUT= option of the FORECAST statement. The default is METHOD=BESTN. The following forecasting models are provided:

SIMPLE	Simple (Single) Exponential Smoothing
DOUBLE	Double (Brown) Exponential Smoothing
LINEAR	Linear (Holt) Exponential Smoothing
DAMPTREND	Damped Trend Exponential Smoothing

BESTN Best Candidate Nonseasonal Smoothing Model (SIMPLE, DOUBLE, LINEAR, DAMPTREND)

NOSTABLE

specifies that the smoothing model parameters are not restricted to the additive invertible region of the parameter space. By default, the smoothing model parameters are restricted to be inside the additive invertible region of the parameter space.

LEVELPARG= *number*

specifies the level weight parameter initial value. See the smoothing model parameter specifications options below.

LEVELREST=(*number,number*)

specifies the level weight parameter restrictions. See the smoothing model parameter specifications options below.

TRENDPARG= *number*

specifies the trend weight parameter initial value. See the smoothing model parameter specifications options below.

TRENDREST=(*number,number*)

specifies the trend weight parameter restrictions. See the smoothing model parameter specifications options below.

DAMPPARG= *number*

specifies the damping weight parameter initial value. See the smoothing model parameter specifications options below.

DAMPREST=(*number,number*)

specifies the damping weight parameter restrictions. See the smoothing model parameter specifications options below.

NOEST

specifies that the smoothing model parameters are fixed values. To use this option, all of the smoothing model parameters must be explicitly specified. By default, the smoothing model parameters are optimized.

BOUNDS=(*number,number*)

Specifies the component forecast bound. See the smoothing model forecast bounds below.

SELECT= *option*

specifies the model selection criterion (statistic of fit) to be used to select from several candidate models. This option would often be used in conjunction with the HOLDOUT= option specified in the FORECAST statement. The default is SELECT=RMSE. The statistics of fit provided are the same as those provided in the FORECAST statement.

Smoothing Model Parameter Specification Options

The parameter options are used to specify smoothing model parameters. If the parameter restrictions are not specified the default is (0.0001 0.9999), which implies that the parameters are restricted between 0.0001 and 0.9999. Parameters and their restrictions are required to be greater than or equal to -1 and less than or equal to 2. Missing values indicate no lower and/or upper restriction. If the parameter initial values are not specified, the optimizer uses a grid search to find an appropriate initial value.

Smoothing Model Forecast Bounds Options

Specifies the demand component forecast bounds. The forecast bounds restrict the component forecasts. The default for demand interval forecasts is BOUNDS=1. The lower bound for the demand interval forecast must be greater than one. The default for demand size forecasts is BOUNDS=(.,.) or no bounds. The demand size forecasts bounds are applied after the forecast is adjusted by the base value.

Details

The HPF procedure can be used to perform trend and seasonal analysis on transactional data. For trend analysis, various sample statistics are computed for each time period defined by the time ID variable and INTERVAL= option. For seasonal analysis, various sample statistics are computed for each season defined by the INTERVAL= or the SEASONALITY= option. For example, suppose the transactional data ranges from June 1990 to January 2000, then the trend statistics are computed for every month: June 1990, July 1990, ..., January 2000. The seasonal statistics are computed for each season: January, February, ..., December.

The HPF procedure can be used to forecast time series data as well as transactional data. If the data is transactional, then the procedure must first accumulate the data into a time series before it can be forecast. The procedure uses the following sequential steps to produce forecasts, with the options that control the step listed to the right:

- | | |
|---------------------------------|--|
| 1. Accumulation | ACCUMULATE= option |
| 2. Missing Value Interpretation | SETMISSING= option |
| 3. Diagnostic Tests | INTERMITTENT= and SEASONTTEST= options |
| 4. Model Selection | MODEL=, HOLDOUT=, HOLDOUTPCT=, and SELECT= options |
| 5. Transformations | TRANSFORM= option |
| 6. Parameter Estimation | MODEL= option |
| 7. Forecasting | MODEL= and LEAD= options |
| 8. Inverse Transformation | TRANSFORM= and MEDIAN options |
| 9. Statistics of Fit | SELECT= option |
| 10. Summation of Forecasts | LEAD= and STARTSUM= options |

Each of the above steps is described below.

Accumulation

If the ACCUMULATE= option is specified, data set observations are accumulated within each time period. The frequency (width of each time interval) is specified by the INTERVAL= option. The ID variable contains the time ID values. Each time ID value corresponds to a specific time period. Accumulation is particularly useful when the input data set contains transactional data, whose observations are not spaced with respect to any particular time interval. The accumulated values form the actual time series, which is used in subsequent analyses.

For example, suppose a data set contains the following observations:

```

19MAR1999    10
19MAR1999    30
11MAY1999    50
12MAY1999    20
23MAY1999    20

```

If the INTERVAL=MONTH is specified, all of the above observations fall within three time periods of March 1999, April 1999, and May 1999. The observations are accumulated within each time period as follows:

If the ACCUMULATE=NONE option is specified, an error is generated because the ID variable values are not equally spaced with respect to the specified frequency (MONTH).

If the ACCUMULATE=TOTAL option is specified:

O1MAR1999	40
O1APR1999	.
O1MAY1999	90

If the ACCUMULATE=AVERAGE option is specified:

O1MAR1999	20
O1APR1999	.
O1MAY1999	30

If the ACCUMULATE=MINIMUM option is specified:

O1MAR1999	10
O1APR1999	.
O1MAY1999	20

If the ACCUMULATE=MEDIAN option is specified:

O1MAR1999	20
O1APR1999	.
O1MAY1999	20

If the ACCUMULATE=MAXIMUM option is specified:

O1MAR1999	30
O1APR1999	.
O1MAY1999	50

If the ACCUMULATE=FIRST option is specified:

O1MAR1999	10
O1APR1999	.
O1MAY1999	50

If the ACCUMULATE=LAST option is specified:

```

O1MAR1999    30
O1APR1999    .
O1MAY1999    20

```

If the ACCUMULATE=STDDEV option is specified:

```

O1MAR1999    14.14
O1APR1999    .
O1MAY1999    17.32

```

As can be seen from the above examples, even though the data set observations contained no missing values, the accumulated time series may have missing values.

Missing Value Interpretation

Sometimes missing values should be interpreted as unknown values. The forecasting models used by the HPF procedure can effectively handle missing values (see the “[Missing Value Modeling Issues](#)” section on page 31). But sometimes missing values are known, such as when missing values are created from accumulation and no observations should be interpreted as no (zero) value. In the former case, the SETMISSING= option can be used to interpret how missing values are treated. The SETMISSING=0 option should be used when missing observations are to be treated as no (zero) values. In other cases, missing values should be interpreted as global values, such as minimum or maximum values of the accumulated series. The accumulated and interpreted time series is used in subsequent analyses.

Diagnostic Tests

The INTERMITTENT= option set the thresholds for categorizing a series as intermittent or non-intermittent. The SEASONTEST= option set the thresholds for categorizing a series as seasonal or non-seasonal.

Model Selection

When more than one candidate model is specified, forecasts for each candidate model are compared using the model selection criterion specified by the SELECT= option. The selection criterion is computed using the multistep forecasts in the holdout sample range if the HOLDOUT= or HOLDOUTPCT= options are specified, or the one-step ahead forecasts for the full range of the time series if the HOLDOUT= and HOLDOUTPCT= options are not specified. The candidate model with the best selection criterion is selected to forecast the time series.

Transformations

If the TRANSFORM= option is specified, the time series is transformed prior to model parameter estimation and forecasting. Only strictly positive series can be transformed. An error is generated when the TRANSFORM= option is used with a nonpositive series.

Parameter Estimation

All parameters associated with the model are optimized based on the data with the default parameter restrictions imposed. If the TRANSFORM= option is specified, the transformed time series data are used to estimate the model parameters.

Missing Value Modeling Issues

The treatment of missing values varies with the forecasting model. For the smoothing models, missing values after the start of the series are replaced with one-step-ahead predicted values, and the predicted values are applied to the smoothing equations. See [Chapter 2, “Forecasting Process Details,”](#) for greater detail on how missing values are treated in the smoothing models. For MODEL=IDM, specified missing values are assumed to be periods of no demand.

The treatment of missing values can also be specified by the user with the SETMISSING= option, which changes the missing values prior to modeling.

Even though all of the observed data are nonmissing, using the ACCUMULATE= option can create missing values in the accumulated series.

Forecasting

Once the model parameters are estimated, one-step ahead forecasts are generated for the full range of the actual (optionally transformed) time series data, and multistep forecasts are generated from the end of the observed time series to the future time period after the last observation specified by the LEAD= option. If there are missing values at the end of the time series, the forecast horizon will be greater than that specified by the LEAD= option.

Inverse Transformations

If the TRANSFORM= option is specified, the forecasts of the transformed time series are inverse transformed. By default, the mean (expected) forecasts are generated. If the MEDIAN option is specified, the median forecasts are generated.

Statistics of Fit

The statistics of fit (or goodness-of-fit statistics) are computed by comparing the actual time series data and the generated forecasts. If the `TRANSFORM=` option was specified, the statistics of fit are based on the inverse transformed forecasts.

Forecast Summation

The multistep forecasts generated by the above steps are summed from the `STARTSUM=` number to the `LEAD=` number. For example, if `STARTSUM=4` and `LEAD=6`, the 4-step through 6-step ahead forecasts are summed. The predictions are simply summed. However, the prediction error variance of this sum is computed taking into account the correlation between the individual predictions. The upper and lower confidence limits for the sum of the predictions is then computed based on the prediction error variance of the sum.

The forecast summation is particularly useful when it is desirable to model in one frequency yet the forecast of interest is another frequency. For example, if a time series has a monthly frequency (`INTERVAL=MONTH`) and you want a forecast for the third and fourth future months, a forecast summation for the third and fourth month can be obtained by specifying `STARTSUM=3` and `LEAD=4`.

Variance-related computations are only computed when no transformation is specified (`TRANSFORM=NONE`).

Comparison to the Time Series Forecasting System

With the exception of Model Selection, the techniques used in the HPF procedure are identical to the Time Series Forecasting System of SAS/ETS software. For Model Parameter Estimation, the default parameter restrictions are imposed.

Data Set Output

The HPF procedure can create the `OUT=`, `OUTEST=`, `OUTFOR=`, `OUTSTAT=`, `OUTSUM=`, `OUTSEASON=`, and `OUTTREND=` data sets. In general, these data sets will contain the variables listed in the `BY` statement. In general, if a forecasting step related to an output data step fails, the values of this step are not recorded or are set to missing in the related output data set, and appropriate error and/or warning messages are recorded in the log.

OUT= Data Set

The `OUT=` data set contains the variables specified in the `BY`, `ID`, and `FORECAST` statements. If the `ID` statement is specified, the `ID` variable values are aligned and extended based on the `ALIGN=` and `INTERVAL=` options. The values of the variables specified in the `FORECAST` statements are accumulated based on the `ACCUMULATE=` option and missing values are interpreted based on the `SETMISSING=` option. If the `REPLACEMISSING` option is specified, missing values embedded missing values are replaced by the one step-ahead forecasts.

These variable values are then extrapolated based on their forecasts or extended with missing values when the MODEL=NONE option is specified. If USE=LOWER is specified, the variable is extrapolated with the lower confidence limits; if USE=UPPER, the variable is extrapolated using the upper confidence limits; otherwise, the variable values are extrapolated with the predicted values. If the TRANSFORM= option is specified, the predicted values will contain either mean or median forecasts depending on whether or not the MEDIAN option is specified.

If any of the forecasting steps fail for particular variable, the variable values are extended by missing values.

OUTEST= Data Set

The OUTEST= data set contains the variables specified in the BY statement as well as the variables listed below. For variables listed in FORECAST statements where the option MODEL=NONE is specified, no observations are recorded for these variables. For variables listed in FORECAST statements where the option MODEL=NONE is not specified, the following variables contain observations related to the parameter estimation step:

NAME	Variable name
MODEL	Forecasting Model
TRANSFORM	Transformation
PARAM	Parameter Name
EST	Parameter Estimate
STDERR	Standard Errors
TVALUE	<i>t</i> -Values
PVALUE	Probability Values

If the parameter estimation step fails for a particular variable, no observations are recorded.

OUTFOR= Data Set

The OUTFOR= data set contains the variables specified in the BY statement as well as the variables listed below. For variables listed in FORECAST statements where the option MODEL=NONE is specified, no observations are recorded for these variables. For variables listed in FORECAST statements where the option MODEL=NONE is not specified, the following variables contain observations related to the forecasting step:

NAME	Variable name
TIMEID	Time ID values
PREDICT	Predicted Values
STD	Prediction Standard Errors

LOWER	Lower Confidence Limits
UPPER	Upper Confidence Limits
ERROR	Prediction Errors

If the forecasting step fails for a particular variable, no observations are recorded. If the TRANSFORM= option is specified, the values in the variables listed above are the inverse transform forecasts. If the MEDIAN option is specified, the median forecasts are stored; otherwise, the mean forecasts are stored.

OUTSTAT= Data Set

The OUTSTAT= data set contains the variables specified in the BY statement as well as the variables listed below. For variables listed in FORECAST statements where the option MODEL=NONE is specified, no observations are recorded for these variables. For variables listed in FORECAST statements where the option MODEL=NONE is not specified, the following variables contain observations related to the statistics of fit step:

<u>_NAME_</u>	Variable name
SSE	Sum of Square Error
MSE	Mean Square Error
UMSE	Unbiased Mean Square Error
RMSE	Root Mean Square Error
URMSE	Unbiased Root Mean Square Error
MAPE	Mean Absolute Percent Error
MAE	Mean Absolute Error
RSQUARE	R-Square
ADJRSQ	Adjusted R-Square
AADJRSQ	Amemiya's Adjusted R-Square
RWRSQ	Random Walk R-Square
AIC	Akaike Information Criterion
SBC	Schwarz Bayesian Information Criterion
APC	Amemiya's Prediction Criterion
MAXERR	Maximum Error
MINERR	Minimum Error
MINPE	Maximum Percent Error
MAXPE	Minimum Percent Error
ME	Mean Error
MPE	Mean Percent Error

If the statistics of fit step fails for particular variable, no observations are recorded. If the TRANSFORM= option is specified, the values in the variables listed above are computed based on the inverse transform forecasts. If the MEDIAN option is specified, the median forecasts are the basis; otherwise, the mean forecasts are the basis.

OUTSUM= Data Set

The OUTSUM= data set contains the variables specified in the BY statement as well as the variables listed below. The OUTSUM= data set records the summary statistics for each variable specified in a FORECAST statement. For variables listed in FORECAST statements where the option MODEL=NONE is specified, the values related to forecasts are set to missing. For variables listed in FORECAST statements where the option MODEL=NONE is not specified, the forecast values are set based on the USE= option.

Variables related to summary statistics are based on the ACCUMULATE= and SETMISSING= options:

NAME	Variable name
STATUS	Forecasting Status. Nonzero values imply that no forecast was generated for the series.
NOBS	Number of Observations
N	Number of Nonmissing Observations
NMISS	Number of Missing Observations
MIN	Minimum Value
MAX	Maximum Value
MEAN	Mean Value
STDDEV	Standard Deviation

Variables related to forecast summation are based on the LEAD= and STARTSUM= options:

PREDICT	Forecast Summation Predicted Values
STD	Forecast Summation Prediction Standard Errors
LOWER	Forecast Summation Lower Confidence Limits
UPPER	Forecast Summation Upper Confidence Limits

Variance-related computations are only computed when no transformation is specified (TRANSFORM=NONE).

Variables related to multistep forecast based on the LEAD= and USE= options:

`_LEAD n _` Multistep Forecast (n ranges from one to `LEAD=number`). If `USE=LOWER`, this variable will contain the lower confidence limits; if `USE=UPPER`, this variable will contain the upper confidence limits; otherwise, this variable will contain the predicted values.

If the forecast step fails for a particular variable, the variables related to forecasting are set to missing. The `OUTSUM=` data set contains both a summary of the (accumulated) time series and optionally its forecasts for all series.

OUTSEASON= Data Set

The `OUTSEASON=` data set contains the variables specified in the `BY` statement as well as the variables listed below. The `OUTSEASON=` data set records the seasonal statistics for each variable specified in a `FORECAST` statement.

Variables related to seasonal statistics are based on the `INTERVAL=` or `SEASONALITY=` options:

<code>_NAME_</code>	Variable name
<code>_TIMEID_</code>	Time ID values
<code>_SEASON_</code>	Seasonal index
<code>NOBS</code>	Number of Observations
<code>N</code>	Number of Nonmissing Observations
<code>NMISS</code>	Number of Missing Observations
<code>MIN</code>	Minimum Value
<code>MAX</code>	Maximum Value
<code>RANGE</code>	Range Value
<code>SUM</code>	Summation Value
<code>MEAN</code>	Mean Value
<code>STDDEV</code>	Standard Deviation
<code>CSS</code>	Corrected Sum of Squares
<code>USS</code>	Uncorrected Sum of Squares
<code>MEDIAN</code>	Median Value

The above statistics are computed for each season.

OUTTREND= Data Set

The OUTTREND= data set contains the variables specified in the BY statement as well as the variables listed below. The OUTTREND= data set records the trend statistics for each variable specified in a FORECAST statement.

Variables related to trend statistics are based on the INTERVAL= and SEASONALITY= options:

NAME	Variable name
TIMEID	Time ID values
SEASON	Seasonal index
NOBS	Number of Observations
N	Number of Nonmissing Observations
NMISS	Number of Missing Observations
MIN	Minimum Value
MAX	Maximum Value
RANGE	Range Value
SUM	Summation Value
MEAN	Mean Value
STDDEV	Standard Deviation
CSS	Corrected Sum of Squares
USS	Uncorrected Sum of Squares
MEDIAN	Median Value

The above statistics are computed for each time period.

Printed Output

The HPF procedure optionally produces printed output for these results utilizing the Output Delivery System (ODS). By default, the procedure produces no printed output. All output is controlled by the PRINT= and PRINTDETAILS options associated with the PROC HPF statement. In general, if a forecasting step related to printed output fails, the values of this step are not printed and appropriate error and/or warning messages are recorded in the log. The printed output is similar to the output data set and these similarities are described below.

PRINT=SUMMARY

prints the summary statistics and forecast summaries similar to the OUTSUM= data set.

PRINT=ESTIMATES

prints the parameter estimates similar to the OUTEST= data set.

PRINT=FORECASTS

prints the forecasts similar to the OUTFOR= data set. For MODEL=IDM, a table containing demand series is also printed.

If the MODEL=IDM option is specified, the demand series predictions table is also printed. This table is based on the demand index (when demands occurred).

PRINT=PERFORMANCE

prints the performance statistics.

PRINT=PERFORMANCESUMMARY

prints the performance summary for each BY group.

PRINT=PERFORMANCEOVERALL

prints the performance summary for all BY groups.

PRINT=STATES

prints the backcast, initial, and final smoothed states.

PRINT=SEASONS

prints the seasonal statistics similar to the OUTSEASON= data set.

PRINT=STATISTICS

prints the statistics of fit similar to the OUTSTAT= data set.

PRINT=TRENDS

Prints the trend statistics similar to the OUTTREND= data set.

PRINTDETAILS

The PRINTDETAILS option is the opposite of the NOOUTALL option.

Specifically, if PRINT=FORECASTS and the PRINTDETAILS options are specified, the one-step ahead forecasts, throughout the range of the data, are printed as well as the information related to a specific forecasting model such as the smoothing states. If the PRINTDETAILS option is not specified, only the multistep forecasts are printed.

ODS Table Names

The table below relates the PRINT= options to ODS tables:

Table 1.1. ODS Tables Produced in PROC HPF

ODS Table Name	Description	Option
ODS Tables Created by the PRINT=SUMMARY option		
DescStats	Descriptive Statistics	
DemandSummary	Demand Summary	MODEL=IDM option only
ForecastSummary	Forecast Summary	
ForecastSummmation	Forecast Summation	
ODS Tables Created by the PRINT=ESTIMATES option		
ModelSelection	Model Selection	
ParameterEstimates	Parameter Estimates	
ODS Tables Created by the PRINT=FORECASTS option		
Forecasts	Forecast	
Demands	Demands	MODEL=IDM option only
ODS Tables Created by the PRINT=PERFORMANCE option		
Performance	Performance Statistics	
ODS Tables Created by the PRINT=PERFORMANCESUMMARY option		
PerformanceSummary	Performance Summary	
ODS Tables Created by the PRINT=PERFORMANCEOVERALL option		
PerformanceSummary	Performance Overall	
ODS Tables Created by the PRINT=SEASONS option		
SeasonStatistics	Seasonal Statistics	
ODS Tables Created by the PRINT=STATES option		
SmoothedStates	Smoothed States	
DemandStates	Demand States	MODEL=IDM option only
ODS Tables Created by the PRINT=STATISTICS option		
FitStatistics	Statistics of Fit	
ODS Tables Created by the PRINT=TRENDS option		
TrendStatistics	Trend Statistics	

The ODS table ForecastSummary is related to all time series within a BY group. The other tables are related to a single series within a BY group.

ODS Graphics (Experimental)

This section describes the use of ODS for creating graphics with the HPF procedure. These graphics are experimental in this release, meaning that both the graphical results and the syntax for specifying them are subject to change in a future release.

To request these graphs, you must specify the ODS GRAPHICS statement. In addition, you can specify the PLOT= option in the HPF statement according to the

following syntax. For more information on the ODS GRAPHICS statement, refer to Chapter 9, “Statistical Graphics Using ODS” (*SAS/ETS User’s Guide*).

PLOT= option | (options)

specifies the graphical output desired. By default, the HPF procedure produces no graphical output. The following printing options are available:

ERRORS	plots prediction error time series graphics.
ACF	plots prediction error autocorrelation function graphics.
PACF	plots prediction error partial autocorrelation function graphics.
IACF	plots prediction error inverse autocorrelation function graphics.
WN	plots white noise graphics.
MODELS	plots model graphics.
FORECASTS	plots forecast graphics.
MODELFORECASTSONLY	plots forecast graphics with confidence limits in the data range.
FORECASTSONLY	plots the forecast in the forecast horizon only.
LEVELS	plots smoothed level component graphics.
SEASONS	plots smoothed seasonal component graphics.
TRENDS	plots smoothed trend (slope) component graphics.
ALL	Same as specifying all of the above PLOT= options.

For example, PLOT=FORECASTS plots the forecasts for each series. The PLOT= option produces printed output for these results utilizing the Output Delivery System (ODS). The PLOT= statement is experimental for this release of SAS.

ODS Graph Names

PROC HPF assigns a name to each graph it creates using ODS. You can use these names to reference the graphs when using ODS. The names are listed in [Table 1.2](#).

To request these graphs, you must specify the ODS GRAPHICS statement. In addition, you can specify the **PLOT=** option in the HPF statement. For more information on the ODS GRAPHICS statement, refer to Chapter 9, “Statistical Graphics Using ODS” (*SAS/ETS User’s Guide*).

Table 1.2. ODS Graphics Produced by PROC HPF

ODS Graph Name	Plot Description	Statement	PLOT= Option
DemandErrorsPlot	Average Demand Errors	PROC HPF	PLOT=ERRORS
DemandForecastsPlot	Average Demand Forecasts	PROC HPF	PLOT=FORECASTS
DemandIntervalHistogram	Demand Interval Histogram	PROC HPF	PLOT=MODELS
DemandIntervalPlot	Demand Interval Forecast Plot	PROC HPF	PLOT=MODELS

Table 1.2. (continued)

ODS Graph Name	Plot Description	Statement	Option
DemandSizeHistogram	Demand Size Histogram	PROC HPF	PLOT=MODELS
DemandSizePlot	Demand Size Forecast Plot	PROC HPF	PLOT=MODELS
ErrorACFNORMPlot	Standardized autocorrelation of Prediction Errors	PROC HPF	PLOT=ACF
ErrorACFPlot	Autocorrelation of Prediction Errors	PROC HPF	PLOT=ACF
ErrorHistogram	Prediction Error Histogram	PROC HPF	PLOT=ERRORS
ErrorIACFNORMPlot	Standardized inverse autocorrelation of Prediction Errors	PROC HPF	PLOT=IACF
ErrorIACFPlot	Inverse autocorrelation of Prediction Errors	PROC HPF	PLOT=IACF
ErrorPACFNORMPlot	Standardized partial autocorrelation of Prediction Errors	PROC HPF	PLOT=PACF
ErrorPACFPlot	Partial autocorrelation of Prediction Errors	PROC HPF	PLOT=PACF
ErrorPlot	Plot of Prediction Errors	PROC HPF	PLOT=ERRORS
ErrorWhiteNoiseLogProbPlot	White noise log probability plot of Prediction Errors	PROC HPF	PLOT=WN
ErrorWhiteNoiseProbPlot	White noise probability plot of Prediction Errors	PROC HPF	PLOT=WN
ForecastsOnlyPlot	Forecasts Only Plot	PROC HPF	PLOT=FORECASTONLY
ForecastsPlot	Forecasts Plot	PROC HPF	PLOT=FORECAST
LevelStatePlot	Smoothed Level State Plot	PROC HPF	PLOT=LEVELS
ModelForecastsPlot	Model and Forecasts Plot	PROC HPF	PLOT=MODELFORECAST
ModelPlot	Model Plot	PROC HPF	PLOT=MODELS
SeasonStatePlot	Smoothed Season State Plot	PROC HPF	PLOT=SEASONS
StockingAveragePlot	Stocking Average Plot	PROC HPF	PLOT=FORECASTS
StockingLevelPlot	Stocking Level Plot	PROC HPF	PLOT=FORECASTS
TrendStatePlot	Smoothed Trend State Plot	PROC HPF	PLOT=TRENDS

Examples

Example 1.1. Automatic Forecasting of Time Series Data

This example illustrates how the HPF procedure can be used for the automatic forecasting of time series data. Retail sales data is used for this illustration.

The following DATA step creates a data set from data recorded monthly at numerous points of sales. The data set, SALES, will contain a variable DATE that represents time and a variable for each sales item. Each value of the DATE variable is recorded in ascending order and the values of each of the other variables represent a single time series:

```
data sales;
  format date date9.;
  input date date9. shoes socks laces dresses coats shirts ties
        belts hats blouses;
  datalines;
  ... data lines omitted ...
  ;
run;
```

The following HPF procedure statements automatically forecast each of the monthly time series.

```
proc hpf data=sales out=nextyear;
  id date interval=month;
  forecast _ALL_;
run;
```

The above statements automatically select the best fitting model and generate forecasts for every numeric variable in the input data set (SALES) for the next twelve months, and stores these forecasts in the output data set (NEXTYEAR).

The following GPLOT procedure statements plot the forecasts related to shoe sales:

```
title1 "Shoe Department Sales";
axis2 label=(a=-90 r=90 "items" );
symbol1 v = dot      i = join l = 1;
symbol2 v = star     i = join l = 2;
symbol3 v = circle   i = join l = 2;

proc gplot data=nextyear;
  plot shoes * date = 1
        socks * date = 2
        laces * date = 3 / overlay
  haxis= '01JAN1994'd to '01DEC2000'd by year
  href= '01JAN1999'd
  vaxis=axis2;
run;
```


The GPLOT procedure results are shown in Figure 1.1. The historical data is shown left the horizontal reference line and the forecasts for the next twelve monthly periods is shown to the right.

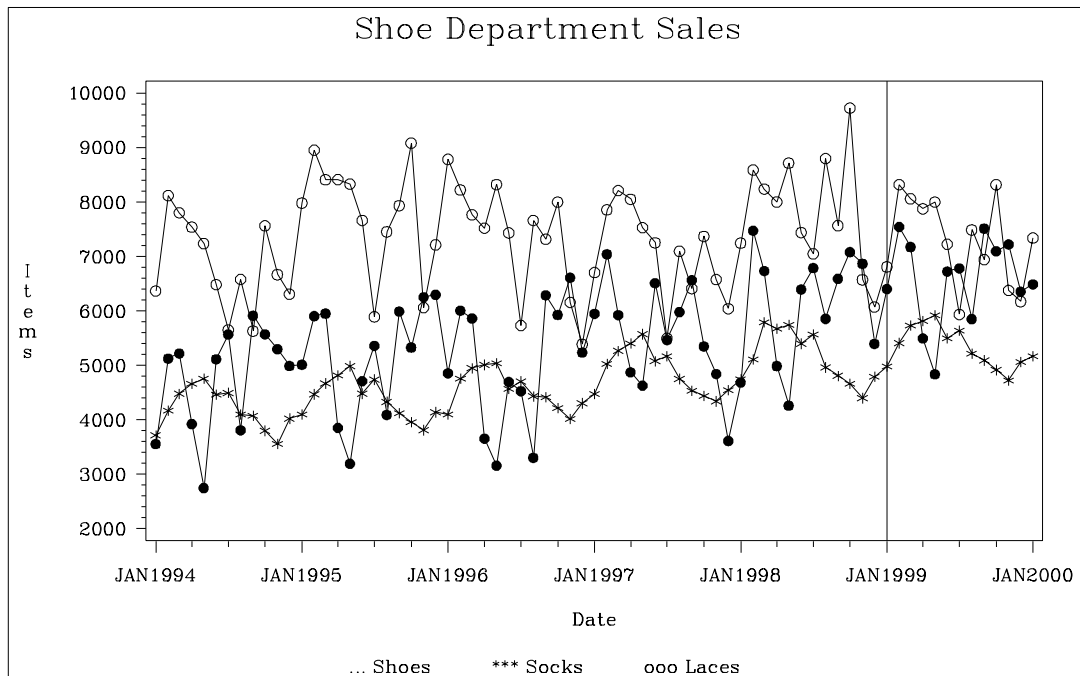


Figure 1.1. Retail Sales Forecast Plots

The following HPF procedure statements are identical to the statements above with the exception that the PRINT=FORECASTS option is specified:

```
proc hpf data=sales out=nextyear print=forecasts;
  id date interval=month;
  forecast _ALL_;
run;
```

In addition to automatically forecasting each of the monthly time series, the above statements print the forecasts using the Output Delivery System (ODS), which is partially shown in Output 1.1.1. This output shows the predictions, prediction standard errors and the upper and lower confidence limits for the next twelve monthly periods.

Output 1.1.1. Forecast Tables

Forecasts for Variable shoes					
Obs	Time	Forecasts	Standard Error	95% Confidence Limits	
62	FEB1999	7548.0041	607.5238	6357.2792	8738.7289
63	MAR1999	7177.1472	699.4400	5806.2701	8548.0244
64	APR1999	5497.5595	780.7609	3967.2964	7027.8227
65	MAY1999	4838.2001	854.5169	3163.3778	6513.0224
66	JUN1999	6728.4521	922.5244	4920.3375	8536.5668
67	JUL1999	6786.1094	985.9738	4853.6362	8718.5826
68	AUG1999	5853.9650	1045.6953	3804.4399	7903.4900
69	SEP1999	7517.0144	1102.2949	5356.5561	9677.4728
70	OCT1999	7100.2489	1156.2315	4834.0769	9366.4210
71	NOV1999	7224.6449	1207.8618	4857.2793	9592.0106
72	DEC1999	6357.1556	1257.4701	3892.5594	8821.7518
73	JAN2000	6492.2657	1305.2871	3933.9500	9050.5815

Example 1.2. Automatic Forecasting of Transactional Data

This example illustrates how the HPF procedure can be used to automatically forecast transactional data. Internet data is used for this illustration.

The following DATA step creates a data set from data recorded at several Internet Web sites. The data set, WEBSITES, will contain a variable TIME that represents time and the variables ENGINE, BOATS, CARS, and PLANES that represent Internet Web site data. Each value of the TIME variable is recorded in ascending order, and the values of each of the other variables represent a transactional data series.

```
data websites;
  format time datetime.;
  input time datetime. boats cars planes engines;
  datalines;
  ... data lines omitted ...
  ;
run;
```

The following HPF procedure statements automatically forecast each of the transactional data series:

```
proc hpf data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast boats cars planes;
run;
```

The above statements accumulate the data into a daily time series and automatically generate forecasts for the BOATS, CARS, and PLANES variables in the input data set (WEBSITES) for the next week and stores the forecasts in the output data set (NEXTWEEK).

The following GPLOT procedure statements plot the forecasts related to the Internet data:

```

title1 "Website Data";
axis2 label=(a=-90 r=90 "Websites" );
symbol1 v = dot      i = join l = 1;
symbol2 v = star     i = join l = 2;
symbol3 v = circle   i = join l = 2;

proc gplot data=nextweek;
  plot boats * time = 1
        cars * time = 2
        planes * time = 3 / overlay
  haxis= '13MAR2000:00:00:00'dt to
        '18APR2000:00:00:00'dt by dtweek
  href= '11APR2000:00:00:00'dt
  vaxis=axis2;
run;

```

The Gplot procedure results are shown in Figure 1.2. The historical data is shown to the left of the horizontal reference line and the forecasts for the next twelve monthly periods are shown to the right.

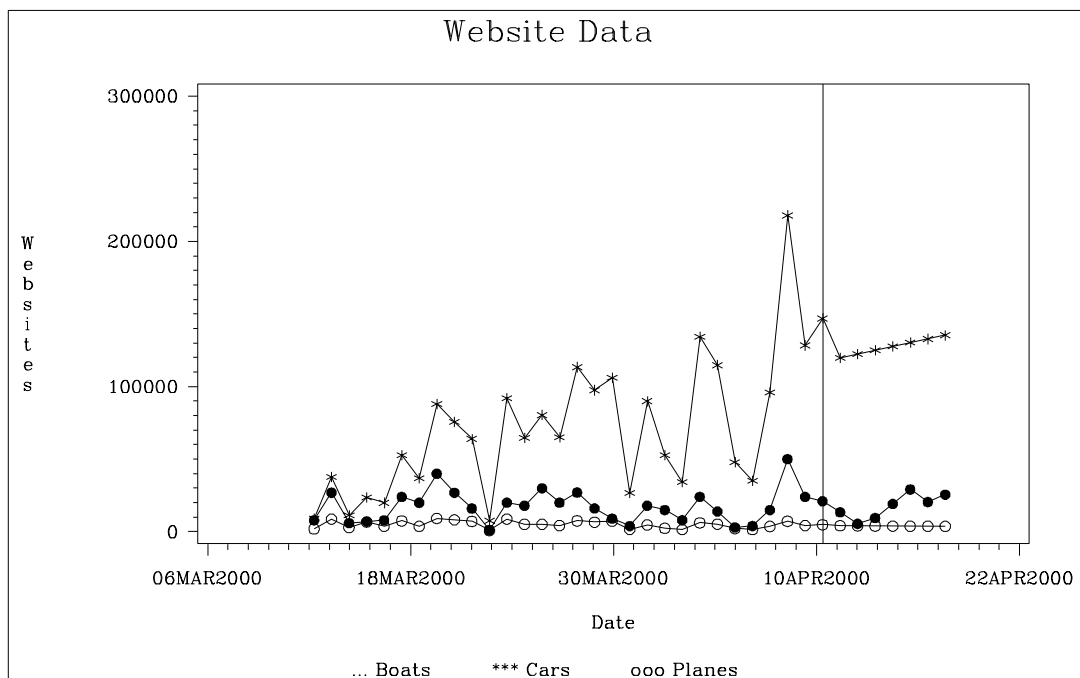


Figure 1.2. Internet Data Forecast Plots

Example 1.3. Specifying the Forecasting Model

In the previous example, the HPF procedure was used to automatically select the appropriate forecasting model using the root mean square error (RMSE) as the default selection criterion. This example illustrates how the HPF procedure can be used to more narrowly specify the possible candidate models. Internet data from the previous example are used for this illustration.

In this example, we will forecast the **BOATS** variable using the best seasonal forecasting model (**BESTS**) that minimizes the mean absolute percent error (**MAPE**), the **CARS** variable using the best nonseasonal forecasting model (**BESTN**) that minimizes the mean square error (**MSE**) using holdout sample analysis, and the **PLANES** variable using Log Winters (additive). The following HPF procedure statements forecast each of the transactional data series based on these requirements:

```
proc hpf data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast boats / model=bests select=mape;
  forecast cars / model=bestn select=mse holdout=5;
  forecast planes / model=addwinters transform=log;
run;
```

Example 1.4. Extending the Independent Variables for Multivariate Forecasts

In the previous example, the HPF procedure was used to forecast several transactional series variables using univariate models. This example illustrates how the HPF procedure can be used to extend the independent variables associated with a multiple regression forecasting problem. Specifically, PROC HPF is used to extend the independent variables for use in forecasting a regression model.

In this example, we will accumulate and forecast the **BOATS**, **CARS**, and **PLANES** variables as illustrated in the previous example. In addition, we will accumulate the **ENGINES** variable to form a time series that is then extended with missing values within the forecast horizon with the specification of **MODEL=NONE**.

```
proc hpf data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast engines / model=none;
  forecast boats / model=bests select=mape;
  forecast cars / model=bestn select=mse holdout=5;
  forecast planes / model=winters transform=log;
run;
```

The following AUTOREG procedure statements are used to forecast the **ENGINES** variable by regressing on the independent variables (**BOATS**, **CARS**, and **PLANES**).

```
proc autoreg data= nextweek;
  model engines = boats cars planes / noprint;
  output out=enginehits p=predicted;
run;
```

The output data set (**NEXTWEEK**) of the PROC HPF statement is used as an input data set for the PROC AUTOREG statement. The output data set of PROC AUTOREG contains the forecast of the variable **ENGINES** based on the regression model with the variables **BOATS**, **CARS**, and **PLANES** as regressors. See the AUTOREG procedure for details on autoregression models.

The following Gplot procedure statements plot the forecasts related to the ENGINES variable:

```
proc gplot data=enginehits;
  plot boats * time = 1
       cars  * time = 2
       planes * time = 3
       predicted * time = 4 / overlay
  haxis= '13MAR2000:00:00:00'dt to
         '18APR2000:00:00:00'dt by dtweek
  href= '11APR2000:00:00:00'dt
  vaxis=axis2;
run;
```

The Gplot procedure results are shown in Figure 1.3. The historical data is shown left the horizontal reference line and the forecasts for the next four weekly periods is shown to the right.

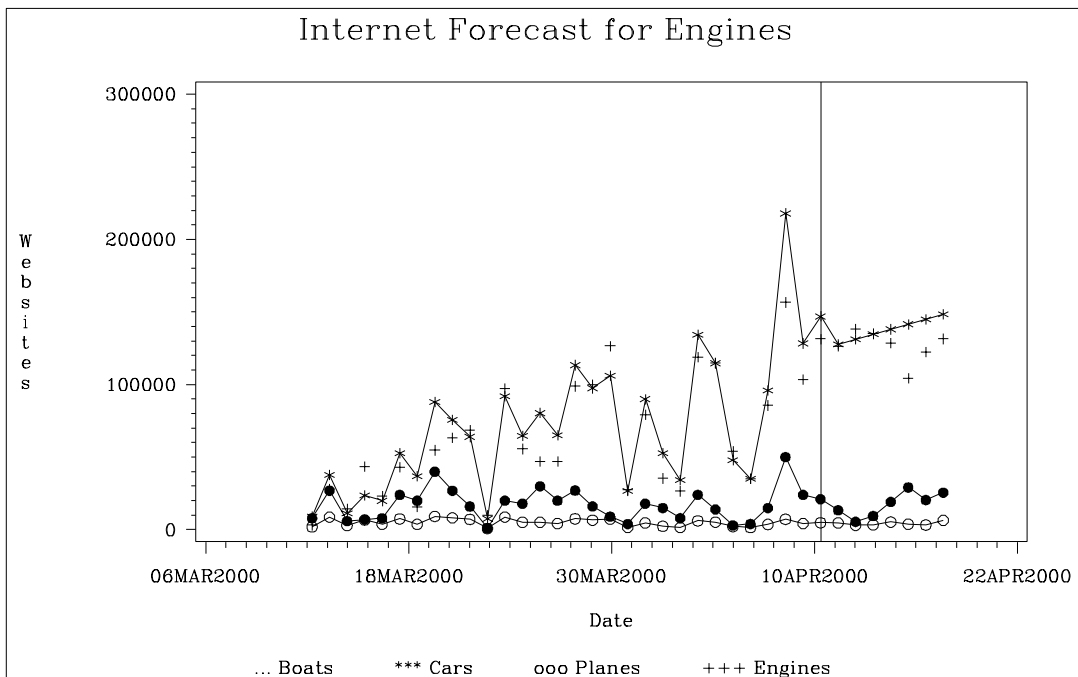


Figure 1.3. Internet Data Forecast Plots

Example 1.5. Forecasting Intermittent Time Series Data

This example illustrates how the HPF procedure can be used to forecast intermittent time series data. Inventory demand is used for this illustration.

The following DATA step creates a data set from inventory data recorded at no particular frequency. The data set, INVENTORY, will contain a variable DATE that represents time and the demand variables (TIRES, HUBCAPS, and LUGBOLTS), which represent inventory items. Each value of the DATE variable is recorded in ascending order, and the values of each of the other variables represent a transactional data series.

```
data inventory;
  format date date9.;
  input date date9. tires hubcaps lugbolts;
  datalines;
  ... data lines omitted ...
  ;
run;
```

The following HPF procedure statements forecast each of the transactional data series using and intermittent demand model:

```
proc hpf data=inventory out=nextmonth lead=4 print=forecasts;
  id date interval=week accumulate=total;
  forecast tires hubcaps lugbolts / model=idm;
run;
```

The above statements accumulate the data into a weekly time series, and generate forecasts for the TIRES, HUBCAPS, and LUGBOLTS variables in the input data set (INVENTORY) for the four weekly periods, and store the forecasts in the output data set (NEXTMONTH). The PRINT=FORECAST option produces the results partially shown in [Output 1.5.1](#). The first table records the demand series and predictions. The second table represents forecasts or recommended stocking levels.

Output 1.5.1. Forecast Tables

Demands for Variable tires						
Index	Demand	Time	Demand Intervals		Demand Size	
			Actual	Predicted	Actual	Predicted
1	Sun, 31 Aug 1997		14	14.0000	6.00000	4.42151
2	Sun, 26 Oct 1997		8	14.0000	4.00000	4.74216
3	Sun, 1 Mar 1998		18	12.8920	2.00000	4.59140
4	Sun, 26 Apr 1998		8	13.8353	2.00000	4.06498
5	Sun, 31 May 1998		5	12.7577	2.00000	3.64550
6	Sun, 27 Sep 1998		17	11.3251	6.00000	3.31123
7	Sun, 3 Jan 1999		14	12.3731	.	3.85743
Error = (Interval Actual)*(Predict) - (Size Actual)						
Demands for Variable tires						
Index	Demand	Time	Estimate of Mean Demand per Period			Error
			Predict	Std	95% Confidence Limits	
1	Sun, 31 Aug 1997		0.31582	0.11634	0.08780 0.54384	-1.578494
2	Sun, 26 Oct 1997		0.33873	0.11634	0.11071 0.56674	-1.290193
3	Sun, 1 Mar 1998		0.35614	0.11634	0.12813 0.58416	4.410578
4	Sun, 26 Apr 1998		0.29381	0.11634	0.06579 0.52183	0.350503
5	Sun, 31 May 1998		0.28575	0.11634	0.05773 0.51377	-0.571254
6	Sun, 27 Sep 1998		0.29238	0.11634	0.06436 0.52040	-1.029543
7	Sun, 3 Jan 1999		0.27553	0.11634	0.04751 0.50355	.
Error = (Interval Actual)*(Predict) - (Size Actual)						
Forecasts for Variable tires						
Obs	Time	Forecasts	Standard Error	95% Confidence Limits		
84	Sun, 3 Jan 1999	0.27553	0.11634	0.04751	0.50355	
85	Sun, 10 Jan 1999	0.27553	0.11634	0.04751	0.50355	
86	Sun, 17 Jan 1999	0.27553	0.11634	0.04751	0.50355	
87	Sun, 24 Jan 1999	0.27553	0.11634	0.04751	0.50355	

Example 1.6. Illustration of ODS Graphics (Experimental)

This example illustrates the use of experimental ODS graphics.

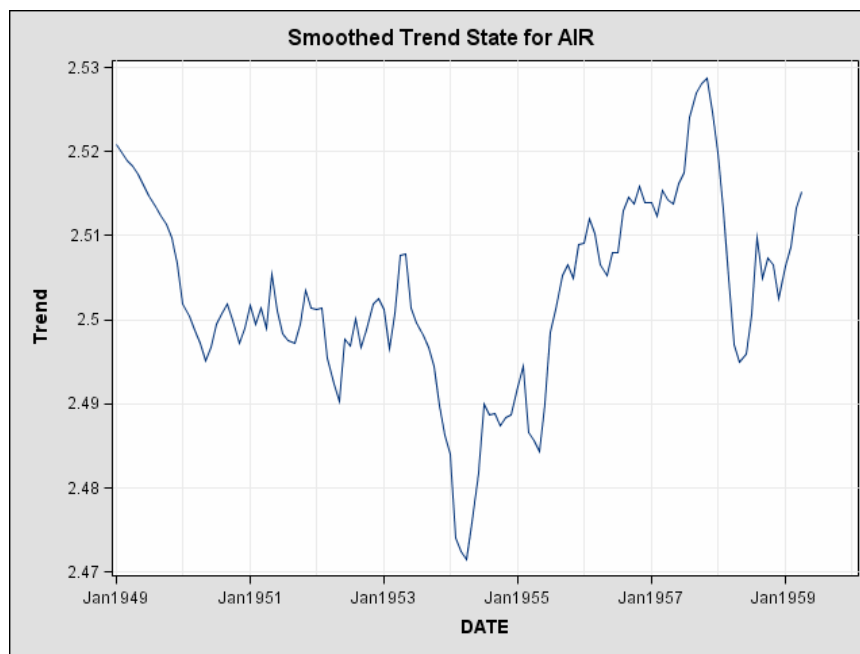
The following statements utilize the SASHELP.AIR data set to automatically forecast the time series of international airline travel.

The graphical displays are requested by specifying the experimental ODS GRAPHICS statement and the experimental PLOT= option in the PROC HPF statement. In this case, all plots are requested. Output 1.6.1 through Output 1.6.4 show a selection of the plots created.

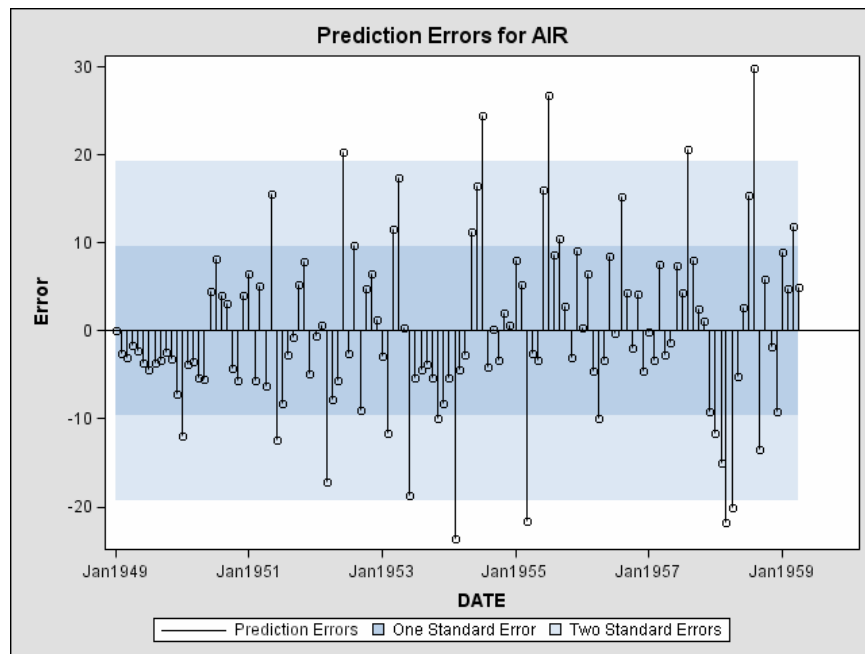
For general information about ODS graphics, refer to Chapter 9, “Statistical Graphics Using ODS” (SAS/ETS User’s Guide). For specific information about the graphics available in the HPF procedure, see the “ODS Graphics” section on page 39.

```
ods html;  
ods graphics on;  
  
proc hpf data=sashelp.air out=_null_ lead=20 back=20 print=all plot=all;  
  id date interval=month;  
  forecast air / model=best transform=auto select=mape;  
run;  
  
ods graphics off;  
ods html close;
```

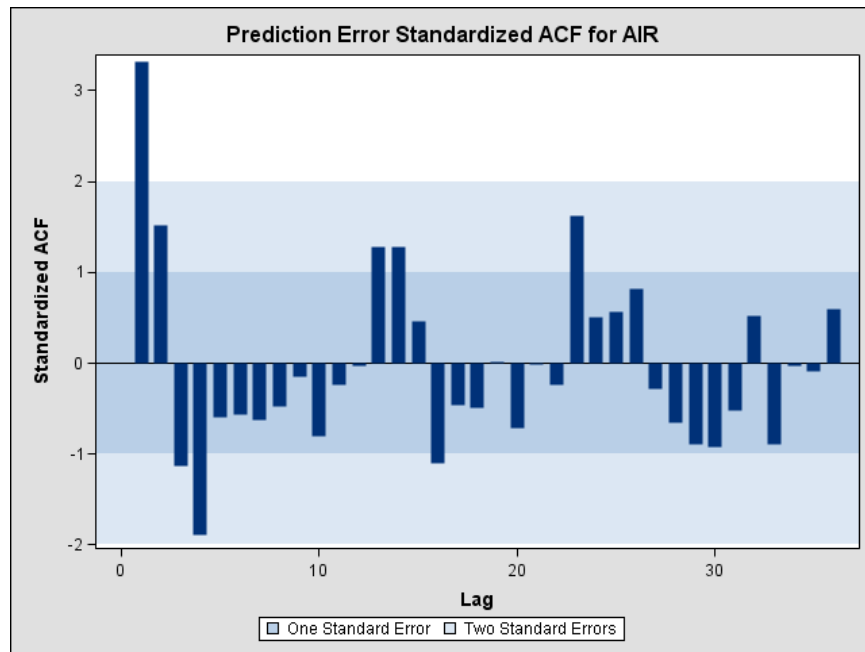
Output 1.6.1. Smoothed Trend Plot (Experimental)



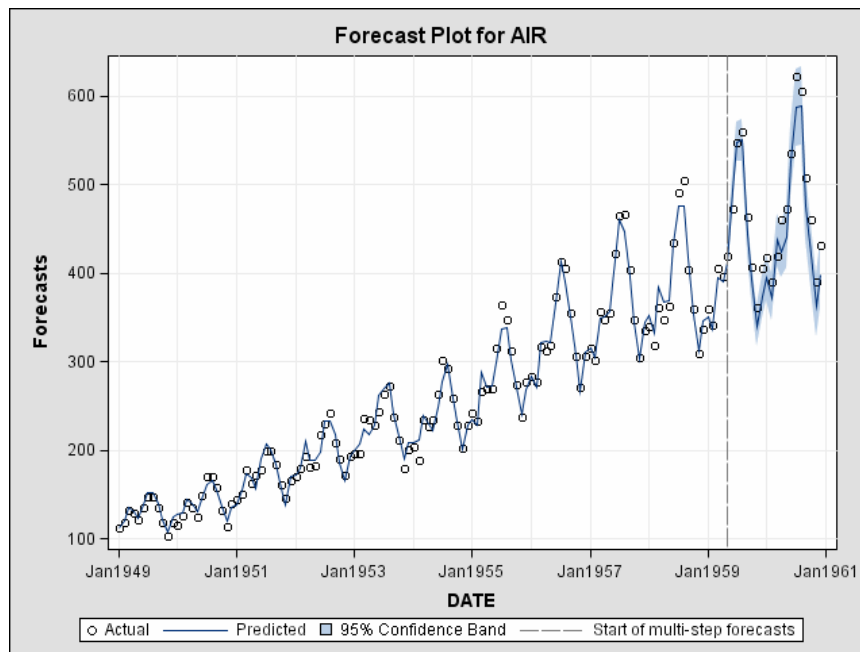
Output 1.6.2. Prediction Error Plot (Experimental)



Output 1.6.3. Prediction Error Standardized ACF Plot (Experimental)



Output 1.6.4. Forecast Plot (Experimental)



References

Pyle, D. (1999), *Data Preparation for Data Mining*, San Francisco: Morgan Kaufman Publishers, Inc.