

Chapter 1

Introduction

Chapter Contents

OVERVIEW OF SAS HPF SOFTWARE	5
Uses of SAS High-Performance Forecasting Software	7
Contents of SAS High-Performance Forecasting Software	7
ABOUT THIS BOOK	9
Chapter Organization	10
Typographical Conventions	11
Options Used in Examples	12
WHERE TO TURN FOR MORE INFORMATION	14
Accessing the SAS High-Performance Forecasting Sample Library	14
Online Help System	14
SAS Institute Technical Support Services	14
RELATED SAS SOFTWARE	14
Base SAS Software	15
SAS/GRAPH Software	17
SAS/STAT Software	17
SAS/IML Software	18
SAS/INSIGHT Software	19
SAS/OR Software	20
SAS/QC Software	20
Other Statistical Tools	21
REFERENCES	21

Chapter 1

Introduction

Overview of SAS High-Performance Forecasting Software

SAS High-Performance Forecasting software provides a large-scale automatic forecasting system. The software provides for the automatic selection of time series models for use in forecasting time-stamped data.

Given a time-stamped data set, the software provides the following automatic forecasting process:

- accumulates the time-stamped data to form a fixed-interval time series
- diagnoses the time series using time series analysis techniques
- creates a list of candidate model specifications based on the diagnostics
- fits each candidate model specification to the time series
- generates forecasts for each candidate fitted model
- selects the most appropriate model specification based on either in-sample or holdout-sample evaluation using a model selection criterion
- refits the selected model specification to the entire range of the time series
- creates a forecast score from the selected fitted model
- generate forecasts from the forecast score
- evaluates the forecast using in-sample analysis

The software also provides for out-of-sample forecast performance analysis.

For time series data without causal inputs (input variables or calendar events), the HPF procedure provides a single, relatively easy to use batch interface that supports the preceding automatic forecasting process. The HPF procedure uses exponential smoothing models (ESM) and intermittent demand models (IDM) in an automated way to extrapolate the time series. The HPF procedure is relatively simple to use and requires only one procedure call.

For time series data with or without causal inputs (input variables and/or calendar events), the software provides several procedures that provide a batch interface that supports the preceding automatic forecasting process with more complicated models. These procedures must be used in the proper sequence in order to get the desired results. Forecasting time series of this nature normally requires more than one procedure call.

Input variables are recorded in the time-stamped data set. These input variables may or may not be incorporated in time series models used to generate forecasts.

Calendar events are specified using the HPFEVENTS procedure. These event definitions are used to generate discrete-valued indicator variables or dummy variables. These event definitions are stored in a SAS data set. These indicator variables may or may not be incorporated in time series models used to generate forecasts.

Given the specified calendar events and input variables, the HPFDIAGNOSE procedure diagnoses the time series and decides which, if any, of the calendar events or input variables are determined to be useful in forecasting the time series. The HPFDIAGNOSE procedure automatically generates candidate model specifications and a model selection list using time series analysis techniques. These model specifications and model selection lists can then be used to automatically generate forecasts.

The user can specify model specifications using one of the following model specification procedures:

- PROC HPFARIMASPEC enables the user to specify one of the family of AutoRegressive Integrated Moving Average with eXogenous inputs (ARIMAX) models.
- PROC HPFESMSPEC enables the user to specify one of the family of exponential smoothing models (ESM).
- PROC HPFEXMSPEC allows the forecast to be generated by an external source.
- PROC HPFIDMSPEC enables the user to specify one of the family of intermittent demand models (IDM).
- PROC HPFSELECT enables the user to specify a model selection list. The model selection list references one or more candidate model specifications and specifies how to choose the appropriate model for a given time series.
- PROC HPFUCMSPEC enables the user to specify one of the family of unobserved component models (UCM).

Regardless of whether the model specifications or model selection lists are specified or automatically generated, the HPFENGINE procedure uses these files to automatically select an appropriate forecasting model, estimate the model parameters, and forecast the time series.

Most of the computational effort associated with automatic forecasting is time series analysis, diagnostics, model selection, and parameter estimation. Forecast scoring files summarize the time series model's parameter estimates and the final states (historical time series information). These files can be used to quickly generate the forecasts required for the iterative nature of scenario analysis, stochastic optimization, and goal seeking computations. The HPFSCSUB function can be used to score time series information.

Uses of SAS High-Performance Forecasting Software

HPF software provides tools for a wide variety of applications in business, government, and academia. Major uses of HPF procedures include: forecasting, forecast scoring, market response modeling, and time series data mining.

Contents of SAS High-Performance Forecasting Software

Procedures

HPF software includes the following SAS procedures:

HPFARIMASPEC The HPFARIMASPEC procedure is used to create an Autoregressive Integrated Moving Average (ARIMA) model specification file. The output of the procedure is an XML file that stores the intended ARIMA model specification. This XML specification file can be used to populate the model repository used by the HPFENGINE procedure. (Likewise, the XML files generated by the other model specification procedures in this section can also be used to populate the model repository used by PROC HPFENGINE.)

HPFDIAGNOSE The HPFDIAGNOSE procedure is an automatic modeling procedure to find the best model among ARIMA Models, Exponential Smoothing Models, and Unobserved Component Models.

The HPFDIAGNOSE procedure has the following functionality:

- intermittency test
- functional transformation test
- simple differencing and seasonal differencing test
- tentative simple ARMA order identification
- tentative seasonal ARMA order identification
- outlier detection
- significance test of events
- transfer functions identification
- intermittent demand model
- exponential smoothing model
- unobserved component model

HPFENGINE The HPFENGINE procedure provides large-scale automatic forecasting of transactional or time series data. The HPFENGINE procedure extends the foundation built by PROC HPF, enabling the user to determine the list of models over which automatic selection is performed.

The use of many forecast model families is supported when HPFENGINE is used in conjunction with new experimental procedures that generate generic model specifications. Among these models are

- ARIMA
- Unobserved Component Models (UCM)
- Exponential Smoothing Models (ESM)
- Intermittent Demand Models (IDM)
- External Models (EXM)

Furthermore, users may completely customize the operation by defining their own code to generate forecasts.

For models with inputs, the STOCHASTIC statement is especially helpful for automatically forecasting those inputs that have no future values.

Also supported is the generation of a portable forecast score. The output of the SCORE statement is a file or catalog entry which, when used with the new function HPFSCSUB, can be used to efficiently generate forecasts outside of the HPFENGINE procedure.

The new HPFDIAGNOSE procedure produces output that is compatible with HPFENGINE. As a result, the task of candidate model specification can be entirely automated.

HPFESMSPEC

The HPFESMSPEC procedure is used to create an Exponential Smoothing Model (ESM) specification file. The output of the procedure is an XML file that stores the intended ESM model specification.

HPFEVENTS

The HPFEVENTS procedure provides a way to create and manage events associated with time series. The procedure can create events, read events from an events data set, write events to an events data set, and create dummies based on those events, if date information is provided.

A SAS event is used to model any incident that disrupts the normal flow of the process that generated the time series. Examples of commonly used events include natural disasters, retail promotions, strikes, advertising campaigns, policy changes, and data recording errors.

An event has a reference name, a date or dates associated with the event, and a set of qualifiers. The event exists separately from any time series; however, the event may be applied to one or more time series. When the event is applied to a time series, a dummy variable is generated that may be used to analyze the impact of the event on the time series.

HPFEXMSPEC	The HPFEXMSPEC procedure is used to create an External Model (EXM) specification file. The output of the procedure is an XML file that stores the intended EXM model specification.
HPFIDMSPEC	The HPFIDMSPEC procedure is used to create an Intermittent Demand Model (IDM) specification file. The output of the procedure is an XML file that stores the intended IDM model specification.
HPFRECONCILE	The HPFRECONCILE procedure reconciles forecasts of time series data at two different levels of aggregation. Optionally, the HPFRECONCILE procedure can disaggregate forecasts from upper level forecasts or aggregate forecasts from lower level forecasts. Additionally, the procedure enables the user to specify the direction and the method of reconciliation, equality constraints and bounds on the reconciled values at each point in time.
HPFSELECT	The HPFSELECT procedure is used to create model selections lists. A model selection list contains references to candidate model specifications stored in the model repository. The output of the procedure is an XML file that stores the intended model selection list.
HPFUCMSPEC	The HPFUCMSPEC procedure is used to create an Unobserved Component Model (UCM) specification file. The output of the procedure is an XML file that stores the intended UCM model specification.
HPFSCSIG	The HPFSCSIG function generates a sample signature for subsequent use by the HPFSCSUB function.
HPFSCSUB	The HPFSCSUB function uses score files to produce forecasts outside of the HPFENGINE procedure. Being a function, it is particularly well suited for use within other SAS programming contexts, such as the DATA step, or procedures that permit the specification of functions, such as the NLP procedure. The only input required is a reference to the score function, the horizon, and future values of any inputs.

About This Book

This book is a user's guide to HPF software. Since HPF software is a part of the SAS System, this book assumes that you are familiar with Base SAS software and have the books *SAS Language: Reference* and *SAS Procedures Guide* available for reference. It also assumes that you are familiar with SAS data sets, the SAS DATA step, and with basic SAS procedures such as PROC PRINT and PROC SORT.

Chapter Organization

The new features added to HPF software since the publication of *SAS High-Performance Forecasting Software: Changes and Enhancements for Release 8.2* are summarized in “What’s New in SAS 9, 9.1, 9.1.2, and 9.1.3 SAS High-Performance Forecasting.” If you have used this software in the past, you may want to skim this chapter to see what is new.

Following the brief “What’s New” section, this book is divided into three major parts.

- Part One contains general information to aid you in working with HPF software.

The current chapter provides an overview of this software and summarizes related SAS Institute publications, products, and services.

This chapter is followed by an “Examples” chapter.

- Part Two, the “Procedure Reference,” contains the chapters that explain the SAS procedures that make up HPF software. The chapters documenting each of the procedures appear in alphabetical order by procedure name and are organized as follows:
 1. Each chapter begins with an “Overview” section that gives a brief description of the procedure.
 2. The “Getting Started” section provides a tutorial introduction on how to use the procedure.
 3. The “Syntax” section is a reference to the SAS statements and options that control the procedure.
 4. The “Details” section discusses various technical details.
 5. The “Examples” section contains examples of the use of the procedure.
 6. The “References” section contains technical references on methodology.
- Part Three provides a summary of and computational details on the SAS High-Performance Forecasting System, an interactive forecasting menu system. Two of the chapters in Part Three document the details of the forecasting process.

Typographical Conventions

This book uses several type styles for presenting information. The following list explains the meaning of the typographical conventions used in this book:

roman	is the standard type style used for most text.
UPPERCASE ROMAN	is used for SAS statements, options, and other SAS language elements when they appear in the text. However, you can enter these elements in your own SAS programs in lowercase, uppercase, or a mixture of the two.
UPPERCASE BOLD	is used in the “Syntax” sections’ initial lists of SAS statements and options.
<i>oblique</i>	is used for user-supplied values for options in the syntax definitions. In the text, these values are written in <i>italic</i> .
helvetica	is used for the names of variables and data sets when they appear in the text.
bold	is used to refer to matrices and vectors, and to refer to commands (e.g., end or cd .)

<i>italic</i>	is used for terms that are defined in the text, for emphasis, and for references to publications.
monospace	is used for example code. In most cases, this book uses lowercase type for SAS code.

Options Used in Examples

Output of Examples

For each example, the procedure output is numbered consecutively starting with 1, and each output is given a title. Each page of output produced by a procedure is enclosed in a box.

Most of the output shown in this book is produced with the following SAS System options:

```
options linesize=80 pagesize=200 nonumber nodate;
```

The template STATDOC.TPL is used to create the HTML output that appears in the online (CD) version. A style template controls stylistic HTML elements such as colors, fonts, and presentation attributes. The style template is specified in the ODS HTML statement as follows:

```
ODS HTML style=statdoc;
```

If you run the examples, you may get slightly different output. This is a function of the SAS System options used and the precision used by your computer for floating-point calculations.

Graphics Options

The examples that contain graphical output are created with a specific set of options and symbol statements. The code you see in the examples creates the color graphics that appear in the online (CD) version of this book. A slightly different set of options and statements is used to create the black-and-white graphics that appear in the printed version of the book.

If you run the examples, you may get slightly different results. This may occur because not all graphic options for color devices translate directly to black-and-white output formats. For complete information on SAS/GRAPH software and graphics options, refer to *SAS/GRAPH Software: Reference*.

The following GOPTIONS statement is used to create the online (color) version of the graphic output.

```
filename GSASFILE '<file-specification>';

goptions reset=all
          gaccess=GSASFILE      gsfmode=replace
          fileonly
          transparency          dev = gif
          ftext = swiss          lfactor = 1
          htext = 4.0pct        htitle = 4.5pct
          hsize = 5.5in         vsize = 3.5in
          noborder              cback = white
          horigin = 0in         vorigin = 0in ;
```

The following GOPTIONS statement is used to create the black-and-white version of the graphic output, which appears in the printed version of the manual.

```
filename GSASFILE '<file-specification>';

goptions reset=all
          gaccess=GSASFILE      gsfmode=replace
          fileonly
          dev = pslepsf
          ftext = swiss          lfactor = 1
          htext = 3.0pct        htitle = 3.5pct
          hsize = 5.5in         vsize = 3.5in
          border                 cback = white
          horigin = 0in         vorigin = 0in;
```

In most of the online examples, the plot symbols are specified as follows:

```
symbol11 value=dot color=white height=3.5pct;
```

The SYMBOL n statements used in online examples order the symbol colors as follows: white, yellow, cyan, green, orange, blue, and black.

In the examples appearing in the printed manual, symbol statements specify COLOR=BLACK and order the plot symbols as follows: dot, square, triangle, circle, plus, x, diamond, and star.

Where to Turn for More Information

This section describes other sources of information about HPF software.

Accessing the SAS High-Performance Forecasting Sample Library

The HPF Sample Library includes many examples that illustrate the use of this software, including the examples used in this documentation. To access these sample programs, select **Help** from the menu and select **SAS Help and Documentation**. From the Contents list, choose **Learning to Use SAS** and then **Sample SAS Programs**.

Online Help System

You can access online help information about HPF software in two ways, depending on whether you are using the SAS windowing environment in the command line mode or the pull-down menu mode.

If you are using a command line, you can access the help menus by typing **help** on the SAS windowing environment command line. Or you can issue the command **help ARIMA** (or another procedure name) to bring up the help for that particular procedure.

If you are using the SAS windowing environment pull-down menus, you can pull-down the **Help** menu and make the following selections:

- SAS Help and Documentation
- SAS Products (on the Contents tab)
- SAS High-Performance Forecasting

The content of the Online Help System follows closely the one of this book.

SAS Institute Technical Support Services

As with all SAS Institute products, the SAS Institute Technical Support staff is available to respond to problems and answer technical questions regarding the use of HPF software.

Related SAS Software

Many features not found in HPF software are available in other parts of the SAS System. If you do not find something you need in this software, you may find it in one of the following SAS software products.

Base SAS Software

The features provided by HPF software are extensions to the features provided by Base SAS software. Many data management and reporting capabilities you will need are part of Base SAS software. Refer to *SAS Language: Reference* and the *SAS Procedures Guide* for documentation of Base SAS software.

The following sections summarize Base SAS software features of interest to users of HPF software. See Chapter 2 (*SAS/ETS User's Guide*) for further discussion of some of these topics as they relate to time series data and HPF software.

SAS DATA Step

The DATA step is your primary tool for reading and processing data in the SAS System. The DATA step provides a powerful general purpose programming language that enables you to perform all kinds of data processing tasks. The DATA step is documented in *SAS Language: Reference*.

Base SAS Procedures

Base SAS software includes many useful SAS procedures. Base SAS procedures are documented in the *SAS Procedures Guide*. The following is a list of Base SAS procedures you may find useful:

CATALOG	for managing SAS catalogs
CHART	for printing charts and histograms
COMPARE	for comparing SAS data sets
CONTENTS	for displaying the contents of SAS data sets
COPY	for copying SAS data sets
CORR	for computing correlations
CPORT	for moving SAS data libraries between computer systems
DATASETS	for deleting or renaming SAS data sets
FREQ	for computing frequency crosstabulations
MEANS	for computing descriptive statistics and summarizing or collapsing data over cross sections
PLOT	for printing scatter plots
PRINT	for printing SAS data sets
RANK	for computing rankings or order statistics
SORT	for sorting SAS data sets
SQL	for processing SAS data sets with Structured Query Language
STANDARD	for standardizing variables to a fixed mean and variance
SYLK	for translating spreadsheets to batch SAS programs. The SYLK procedure is experimental. The documentation can be found at http://support.sas.com/documentation/onlinedoc by selecting "Base SAS" from the Product-Specific Documentation list

General Information ♦ Introduction

TABULATE	for printing descriptive statistics in tabular format
TIMEPLOT	for plotting variables over time
TRANSPOSE	for transposing SAS data sets
UNIVARIATE	for computing descriptive statistics

Global Statements

Global statements can be specified anywhere in your SAS program, and they remain in effect until changed. Global statements are documented in *SAS Language: Reference*. You may find the following SAS global statements useful:

FILENAME	for accessing data files
FOOTNOTE	for printing footnote lines at the bottom of each page
%INCLUDE	for including files of SAS statements
LIBNAME	for accessing SAS data libraries
OPTIONS	for setting various SAS system options
RUN	for executing the preceding SAS statements
TITLE	for printing title lines at the top of each page
X	for issuing host operating system commands from within your SAS session

Some Base SAS statements can be used with any SAS procedure, including HPF procedures. These statements are not global, and they only affect the SAS procedure they are used with. These statements are documented in *SAS Language: Reference*.

The following Base SAS statements are useful with HPF procedures:

BY	for computing separate analyses for groups of observations
FORMAT	for assigning formats to variables
LABEL	for assigning descriptive labels to variables
WHERE	for subsetting data to restrict the range of data processed or to select or exclude observations from the analysis

SAS Functions

SAS functions can be used in DATA step programs and in the COMPUTAB and MODEL procedures. The following kinds of functions are available:

- character functions for manipulating character strings
- date and time functions, for performing date and calendar calculations
- financial functions, for performing financial calculations such as depreciation, net present value, periodic savings, and internal rate of return

- lagging and differencing functions, for computing lags and differences
- mathematical functions, for computing data transformations and other mathematical calculations
- probability functions, for computing quantiles of statistical distributions and the significance of test statistics
- random number functions, for simulation experiments
- sample statistics functions, for computing means, standard deviations, kurtosis, and so forth

Formats, Informats, and Time Intervals

Base SAS software provides formats to control the printing of data values, informats to read data values, and time intervals to define the frequency of time series.

SAS/GRAPH Software

SAS/GRAPH software includes procedures that create two- and three-dimensional high-resolution color graphics plots and charts. You can generate output that graphs the relationship of data values to one another, enhance existing graphs, or simply create graphics output that is not tied to data. SAS/GRAPH software can produce

- charts
- plots
- maps
- text
- three-dimensional graphs

With SAS/GRAPH software you can produce high-resolution color graphics plots of time series data.

SAS/STAT Software

SAS/STAT software is of interest to users of HPF software because many econometric and other statistical methods not included in HPF software are provided in SAS/STAT software.

SAS/STAT software includes procedures for a wide range of statistical methodologies, including

- logistic regression
- censored regression
- principal component analysis
- structural equation models using covariance structure analysis
- factor analysis

- survival analysis
- discriminant analysis
- cluster analysis
- categorical data analysis; log-linear and conditional logistic models
- general linear models
- mixed linear and nonlinear models
- generalized linear models
- response surface analysis
- kernel density estimation
- LOESS regression
- spline regression
- two-dimensional kriging
- multiple imputation for missing values

SAS/IML Software

SAS/IML software gives you access to a powerful and flexible programming language (Interactive Matrix Language) in a dynamic, interactive environment. The fundamental object of the language is a data matrix. You can use SAS/IML software interactively (at the statement level) to see results immediately, or you can store statements in a module and execute them later. The programming is dynamic because necessary activities such as memory allocation and dimensioning of matrices are done automatically.

You can access built-in operators and call routines to perform complex tasks such as matrix inversion or eigenvector generation. You can define your own functions and subroutines using SAS/IML modules. You can perform operations on an entire data matrix. You have access to a wide choice of data management commands. You can read, create, and update SAS data sets from inside SAS/IML software without ever using the DATA step.

SAS/IML software is of interest to users of HPF software because it enables you to program your own econometric and time series methods in the SAS System. It contains subroutines for time series operators and for general function optimization. If you need to perform a statistical calculation not provided as an automated feature by HPF or other SAS software, you can use SAS/IML software to program the matrix equations for the calculation.

Kalman Filtering and Time Series Analysis in SAS/IML

SAS/IML software includes a library for Kalman filtering and time series analysis which provides the following functions:

- generating univariate, multivariate, and fractional time series
- computing likelihood function of ARMA, VARMA, and ARFIMA models

- computing an autocovariance function of ARMA, VARMA, and ARFIMA models
- checking the stationarity of ARMA and VARMA models
- filtering and smoothing of time series models using Kalman method
- fitting AR, periodic AR, time-varying coefficient AR, VAR, and ARFIMA models
- handling Bayesian seasonal adjustment model

SAS/INSIGHT Software

SAS/INSIGHT software is a highly interactive tool for data analysis. You can explore data through a variety of interactive graphs including bar charts, scatter plots, box plots, and three-dimensional rotating plots. You can examine distributions and perform parametric and nonparametric regression, analyze general linear models and generalized linear models, examine correlation matrixes, and perform principal component analyses. Any changes you make to your data show immediately in all graphs and analyses. You can also configure SAS/INSIGHT software to produce graphs and analyses tailored to the way you work.

SAS/INSIGHT software is an integral part of the SAS System. You can use it to examine output from a SAS procedure, and you can use any SAS procedure to analyze results from SAS/INSIGHT software.

SAS/INSIGHT software includes features for both displaying and analyzing data interactively. A data window displays a SAS data set as a table with columns of the table displaying variables and rows displaying observations. Data windows provide data management features for editing, transforming, subsetting, and sorting data. A graph window displays different types of graphs: bar charts, scatter plots, box plots, and rotating plots. Graph windows provide interactive exploratory techniques such as data brushing and highlighting. Analysis windows display statistical analyses in the form of graphs and tables. Analysis window features include

- univariate statistics
- robust estimates
- density estimates
- cumulative distribution functions
- theoretical quantile-quantile plots
- multiple regression analysis with numerous diagnostic capabilities
- general linear models
- generalized linear models
- smoothing spline estimates
- kernel density estimates

- correlations
- principal components

SAS/INSIGHT software may be of interest to users of HPF software for interactive graphical viewing of data, editing data, exploratory data analysis, and checking distributional assumptions.

SAS/OR Software

SAS/OR software provides SAS procedures for operations research and project planning and includes a menu-driven system for project management. SAS/OR software has features for

- solving transportation problems
- linear, integer, and mixed-integer programming
- nonlinear programming and optimization
- scheduling projects
- plotting Gantt charts
- drawing network diagrams
- solving optimal assignment problems
- network flow programming

SAS/OR software may be of interest to users of HPF software for its mathematical programming features. In particular, the NLP procedure in SAS/OR software solves nonlinear programming problems and can be used for constrained and unconstrained maximization of user-defined likelihood functions.

SAS/QC Software

SAS/QC software provides a variety of procedures for statistical quality control and quality improvement. SAS/QC software includes procedures for

- Shewhart control charts
- cumulative sum control charts
- moving average control charts
- process capability analysis
- Ishikawa diagrams
- Pareto charts
- experimental design

SAS/QC software also includes the SQC menu system for interactive application of statistical quality control methods and the ADX Interface for experimental design.

Other Statistical Tools

Many other statistical tools are available in Base SAS, SAS/STAT, SAS/OR, SAS/QC, SAS/INSIGHT, and SAS/IML software. If you do not find something you need in HPF software, you may find it in SAS/STAT software and in Base SAS software. If you still do not find it, look in other SAS software products or contact the SAS Institute Technical Support staff.

References

- Leonard, Michael (2002), "Large-Scale Automatic Forecasting: Millions of Forecasts," Cary, North Carolina: SAS Institute, Inc.
- Leonard, Michael (2004), "Large-Scale Automatic Forecasting with Inputs and Calendar Events," Cary, North Carolina: SAS Institute, Inc.
- Leonard, Michael (2003), "Mining Transactional and Time Series Data," Cary, North Carolina: SAS Institute, Inc.
- Leonard, Michael (2000), "Promotional Analysis and Forecasting for Demand Planning: A Practical Time Series Approach," Cary, North Carolina: SAS Institute, Inc.