

# **SAS/ETS<sup>®</sup> 14.3**

## **User's Guide**

### **The ESM Procedure**

This document is an individual chapter from *SAS/ETS® 14.3 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2017. *SAS/ETS® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

### **SAS/ETS® 14.3 User's Guide**

Copyright © 2017, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government License Rights; Restricted Rights:** The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

September 2017

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

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

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to <http://support.sas.com/thirdpartylicenses>.

# Chapter 14

## The ESM Procedure

### Contents

---

|  |            |
|--|------------|
| Overview: ESM Procedure . . . . .  | <b>840</b> |
| Getting Started: ESM Procedure . . . . .   | <b>840</b> |
| Syntax: ESM Procedure . . . . .  | <b>842</b> |
| Functional Summary . . . . .   | 842        |
| PROC ESM Statement . . . . .   | 844        |
| BY Statement . . . . .   | 847        |
| FORECAST Statement . . . . .   | 847        |
| ID Statement . . . . .   | 849        |
| Details: ESM Procedure . . . . .   | <b>852</b> |
| Accumulation . . . . .   | 852        |
| Missing Value Interpretation . . . . .   | 854        |
| Transformations . . . . .  | 854        |
| Parameter Estimation . . . . .   | 854        |
| Missing Value Modeling Issues . . . . .  | 855        |
| Forecasting . . . . .  | 855        |
| Inverse Transformations . . . . .  | 855        |
| Statistics of Fit . . . . .  | 855        |
| Forecast Summation . . . . .   | 855        |
| Data Set Output . . . . .  | 856        |
| Printed Output . . . . .   | 860        |
| ODS Table Names . . . . .  | 861        |
| ODS Graphics . . . . .   | 862        |
| Examples: ESM Procedure . . . . .  | <b>863</b> |
| Example 14.1: Forecasting of Time Series Data . . . . .                                | 863        |
| Example 14.2: Forecasting of Transactional Data . . . . .                              | 865        |
| Example 14.3: Specifying the Forecasting Model . . . . .                               | 867        |
| Example 14.4: Extending the Independent Variables for Multivariate Forecasts . . . . . | 867        |
| Example 14.5: Illustration of ODS Graphics . . . . .                                   | 869        |

---

---

## Overview: ESM Procedure

The ESM procedure generates forecasts by using exponential smoothing models with optimized smoothing weights for many time series or transactional data.

- 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

Graphics are available with the ESM procedure. For more information, see the section “[ODS Graphics](#)” on page 862.

The exponential smoothing models supported in PROC ESM differ from those supported in PROC FORECAST since all parameters associated with the forecasting model are optimized by PROC ESM based on the data.

The ESM 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 ESM procedure optionally produces printed output for these results by using the Output Delivery System (ODS).

The ESM procedure can forecast both time series data, whose observations are equally spaced by a specific time interval (for example, 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 are accumulated based on a specified time interval to form a time series prior to modeling and forecasting.

---

## Getting Started: ESM Procedure

The ESM procedure is simple to use and does not require in-depth knowledge of forecasting methods. It can provide results in output data sets or in other output formats by using the Output Delivery System (ODS). The following examples are more fully illustrated in “[Example 14.2: Forecasting of Transactional Data](#)” on page 865.

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

```
proc esm 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 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 ESM procedure could be used as follows:

```
proc esm data=sales out=nextyear;
  id date interval=month;
  forecast _numeric_;
run;
```

The preceding statements generate forecasts for every numeric variable in the input data set SALES for the next 12 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.

By default, PROC ESM generates no printed output. If you want to print the forecasts by using the Output Delivery System (ODS), then you need to add the PRINT=FORECASTS option to the PROC ESM statement, as shown in the following example:

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

Other PRINT= options can be specified to print the parameter estimates, statistics of fit, and summary data.

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

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

```
proc esm data=<input-data-set> out=<output-data-set>;
  id <time-ID-variable> interval=<frequency>
    accumulate=<accumulation>;
  forecast <time-series-variables> / model=<esm>;
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 website hit. The forecasts for the total daily values are to be recorded in the output data set NEXTWEEK. The ESM procedure could be used as follows:

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

The preceding statements accumulate the data into a daily time series, 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). Because the MODEL= option is not specified in the FORECAST statement, a simple exponential smoothing model is fit to each series.

---

## Syntax: ESM Procedure

The following statements are available in the ESM procedure:

```
PROC ESM options ;
  BY variables ;
  ID variable INTERVAL= interval < options > ;
  FORECAST variable-list / < options > ;
```

---

## Functional Summary

The statements and options that control the ESM procedure are summarized in Table 14.1.

**Table 14.1** Functional Summary

| Description  | Statement | Option         |
|--|-----------|----------------|
| <b>Statements</b>  |           |                |
| Specify data sets and options                              | PROC ESM  |                |
| Specify BY-group processing                                | BY        |                |
| Specify variables to forecast                              | FORECAST  |                |
| Specify the time ID variable                               | ID        |                |
| <b>Data Set Options</b>                                    |           |                |
| Specify the input data set                                 | PROC ESM  | DATA=          |
| Specify to output forecasts only                           | PROC ESM  | NOOUTALL       |
| Specify the output data set                                | PROC ESM  | OUT=           |
| Specify parameter output data set                          | PROC ESM  | OUTEST=        |
| Specify forecast output data set                           | PROC ESM  | OUTFOR=        |
| Specify the forecast procedure information output data set | PROC ESM  | OUTPROCINFO=   |
| Specify statistics output data set                         | PROC ESM  | OUTSTAT=       |
| Specify summary output data set                            | PROC ESM  | OUTSUM=        |
| Replace actual values held back                            | FORECAST  | REPLACEBACK    |
| Replace missing values                                     | FORECAST  | REPLACEMISSING |
| Use forecast value to append                               | FORECAST  | USE=           |
| <b>Accumulation and Seasonality Options</b>                |           |                |
| Specify accumulation frequency                             | ID        | INTERVAL=      |
| Specify length of seasonal cycle                           | PROC ESM  | SEASONALITY=   |
| Specify interval alignment                                 | ID        | ALIGN=         |

**Table 14.1** *continued*

| <b>Description</b>  | <b>Statement</b> | <b>Option</b> |
|---|------------------|---------------|
| Specify that 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=     |
| <b>Forecasting Horizon, Holdback Options</b>                  |                  |               |
| Specify data to hold back                                     | PROC ESM         | BACK=         |
| Specify forecast horizon or lead                              | PROC ESM         | LEAD=         |
| Specify horizon to start summation                            | PROC ESM         | STARTSUM=     |
| <b>Forecasting Model Options</b>                              |                  |               |
| Specify confidence limit width                                | FORECAST         | ALPHA=        |
| Specify forecast model  | FORECAST         | MODEL=        |
| Specify median forecasts                                      | FORECAST         | MEDIAN        |
| Specify backcast initialization                               | FORECAST         | NBACKCAST=    |
| Specify model transformation                                  | FORECAST         | TRANSFORM=    |
| <b>Printing and Plotting Control Options</b>                  |                  |               |
| Specify time ID format  | ID               | FORMAT=       |
| Specify graphical output                                      | PROC ESM         | PLOT=         |
| Specify printed output  | PROC ESM         | PRINT=        |
| Specify detailed printed output                               | PROC ESM         | PRINTDETAILS  |
| <b>Miscellaneous Options</b>                                  |                  |               |
| Specify that analysis variables are processed in sorted order | PROC ESM         | SORTNAMES     |
| Limit error and warning messages                              | PROC ESM         | MAXERROR=     |

The following sections describe the PROC ESM statement and then describe the other statements in alphabetical order.

---

## PROC ESM Statement

**PROC ESM** *options* ;

You can specify the following *options*:

**BACK**=*n*

specifies the number of observations before the end of the data where the multistep forecasts are to begin. By default, BACK=0.

**DATA**=*SAS-data-set*

names the SAS data set that contains 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). By default, LEAD=12.

The LEAD= value is relative to the BACK= option specification and to the last observation in the input data set or the accumulated series, 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 is 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. This option is particularly useful in BY-group processing where it can be used to suppress the recurring messages. By default, MAXERRORS=50.

**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; it does not include 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 and ending with 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 is also 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 USE= option in the FORECAST statement. The OUT= data set is particularly useful in extending the independent variables. The OUT= data set can be used as the input data set in a subsequent PROC step to forecast a dependent series by using a regression modeling procedure. If the OUT= option is not specified, a default output data set is created by using the DATA*n* convention. If you do not want the OUT= data set created, 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 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 useful for displaying the forecasts in tabular or graphical form.

**OUTPROCINFO=SAS-data-set**

names the output data set to contain information in the SAS log, specifically the number of notes, errors, and warnings and the number of series processed, forecasts requested, and forecasts failed.

**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 useful for evaluating how well the model fits the series.

**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= option is specified. The forecast summations are based on the LEAD=, STARTSUM=, and USE= options. The OUTSUM= data set is useful when forecasting large numbers of series and a summary of the results are needed.

**PLOT=option | ( options )**

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

|                       |   |
|-----------------------|---|
| <b>ACF</b>            | plots prediction error autocorrelation function graphics.   |
| <b>ALL</b>            | is the same as specifying all of the PLOT= options.   |
| <b>BASIC</b>          | equivalent to specifying PLOT=(CORR ERRORS MODELFORECASTS).   |
| <b>CORR</b>           | plots the prediction error series graphics panel containing the ACF, IACF, PACF, and white noise probability plots.   |
| <b>ERRORS</b>         | plots prediction error time series graphics.  |
| <b>FORECASTS</b>      | plots forecast graphics.  |
| <b>FORECASTSONLY</b>  | plots the forecast in the forecast horizon only.  |
| <b>IACF</b>           | plots prediction error inverse autocorrelation function graphics.   |
| <b>LEVELS</b>         | plots smoothed level component graphics.  |
| <b>MODELFORECASTS</b> | plots the one-step ahead model forecast and its confidence bands in the historical period; the forecast and its confidence bands over the forecast horizon. |
| <b>MODELS</b>         | plots model graphics.   |
| <b>PACF</b>           | plots prediction error partial autocorrelation function graphics.   |
| <b>PERIODOGRAM</b>    | plots prediction error periodogram.   |
| <b>SEASONS</b>        | plots smoothed seasonal component graphics.   |
| <b>SPECTRUM</b>       | plots periodogram and smoothed periodogram of the prediction error series in a single graph.  |

|               |  |
|---------------|--|
| <b>TRENDS</b> | plots smoothed trend (slope) component graphics. |
| <b>WN</b>     | plots white noise graphics.                      |

For example, PLOT=FORECASTS plots the forecasts for each series. The PLOT= option produces printed output for these results by using the Output Delivery System (ODS).

**PRINT=option | ( options )**

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

|                           |  |
|---------------------------|--|
| <b>ESTIMATES</b>          | prints the results of parameter estimation.                    |
| <b>FORECASTS</b>          | prints the forecasts.  |
| <b>PERFORMANCE</b>        | prints the performance statistics for each forecast.           |
| <b>PERFORMANCESUMMARY</b> | prints the performance summary for each BY group.              |
| <b>PERFORMANCEOVERALL</b> | prints the performance summary for all of the BY groups.       |
| <b>STATISTICS</b>         | prints the statistics of fit.                                  |
| <b>STATES</b>             | prints the backcast, initial, and final states.                |
| <b>SUMMARY</b>            | prints the summary statistics for the accumulated time series. |
| <b>ALL</b>                | Same as PRINT=(ESTIMATES FORECASTS STATISTICS SUMMARY).        |

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

**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 12.

**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. By default, STARTSUM=1; that is, the sum from the one-step ahead forecast (which is the first forecast in the forecast horizon) to the multistep forecast specified by the LEAD= option.

The prediction standard errors of the summation of forecasts take into account the correlation between the multistep forecasts. For more information about the STARTSUM= option, see the section “[Forecast Summation](#)” on page 855.

---

## BY Statement

**BY** *variables* ;

A BY statement can be used with PROC ESM to obtain separate dummy variable definitions for groups of observations defined by the BY variables.

When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the option NOTSORTED or DESCENDING in the BY statement for the ESM procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure.

For more information about the BY statement, see *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *SAS Visual Data Management and Utility Procedures Guide*.

---

## 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.

A data set variable can be specified in only one FORECAST statement. Any number of FORECAST statements can be used. You can specify the following *options*:

### **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 in the ID statement. Use the ACCUMULATE= option with multiple FORECAST statements when you want different accumulation specifications for different variables. For more information, see the [ACCUMULATE=](#) option in the ID statement.

### **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. By default, ALPHA=0.05, which produces 95% confidence intervals.

**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 time series by 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 time series. You can specify the following forecasting *model-names*:

|                               |  |
|-------------------------------|--|
| <b>NONE</b>                   | produces no forecast, but the time series is appended with missing values in the OUT= data set. This option is useful when the results stored in the OUT= data set are used in a subsequent analysis where forecasts of the independent variables are needed to forecast the dependent variable. |
| <b>SIMPLE</b>                 | performs simple (single) exponential smoothing.  |
| <b>DOUBLE</b>                 | performs double (Brown) exponential smoothing.   |
| <b>LINEAR</b>                 | performs linear (Holt) exponential smoothing.  |
| <b>DAMP TREND</b>             | performs damped trend exponential smoothing.   |
| <b>ADDSEASONAL   SEASONAL</b> | performs additive seasonal exponential smoothing.  |
| <b>MULTSEASONAL</b>           | performs multiplicative seasonal exponential smoothing.  |
| <b>WINTERS</b>                | uses the Winters multiplicative method.  |
| <b>ADDWINTERS</b>             | uses the Winters additive method.  |

By default, MODEL=SIMPLE.

**NBACKCAST=***n*

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

**REPLACEBACK**

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

**REPLACEMISSING**

replaces embedded missing values with one-step-ahead forecasts in the OUT= data set.

**SETMISSING=***option | number*

specifies how missing values (either input 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. For more information, see the SETMISSING= option in the ID statement.

**TRANSFORM=***option*

specifies the time series transformation to be applied to the input or accumulated time series. The following transformations are provided:

|             |                            |
|-------------|----------------------------|
| <b>NONE</b> | no transformation.         |
| <b>LOG</b>  | logarithmic transformation |

|                            |   |
|----------------------------|---|
| <b>SQRT</b>                | square-root transformation  |
| <b>LOGISTIC</b>            | logistic transformation   |
| <b>BOXCOX</b> ( <i>n</i> ) | Box-Cox transformation with parameter number where number is between -5 and 5 |

By default, TRANSFORM=NONE.

When the TRANSFORM= option is specified, the time series must be strictly positive. After the time series is transformed, the model parameters are estimated by 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. For more information, see the sections “Transformations” on page 854 and “Inverse Transformations” on page 855.

#### **USE=***option*

specifies which forecast values are appended to the actual values in the OUT= and OUTSUM= data sets. You can specify the following *options*:

|                |   |
|----------------|---|
| <b>PREDICT</b> | appends the predicted values to the actual values.              |
| <b>LOWER</b>   | appends the lower confidence limit values to the actual values. |
| <b>UPPER</b>   | appends the upper confidence limit values to the actual values. |

By default, USE=PREDICT.

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

#### **ZEROMISS=NONE | LEFT | RIGHT | BOTH**

specifies how beginning or ending zero values (either input 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, beginning or ending zero values are set to missing values based on the ZEROMISS= option in the ID statement. For more information, see the ZEROMISS= option in the ID statement.

---

## 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 or datetime values. In addition, the ID statement specifies the (desired) frequency associated with the time series. The ID statement options also specify how the observations are accumulated and how the time ID values are aligned to form the time series to be forecast. The information specified affects all variables specified in subsequent FORECAST statements. If the ID statement is specified, the INTERVAL= option must be specified. If an ID statement is not specified, the observation number, with respect to the BY group, is used as the time ID. You can specify the following *options*.

**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 time series, which is used in subsequent model fitting and forecasting.

This option is particularly useful when there are gaps in the input data or when there are multiple input observations that coincide with a particular time period (for example, 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:

|                      |   |
|----------------------|---|
| <b>NONE</b>          | No accumulation occurs; the <code>ID</code> variable values must be equally spaced with respect to the frequency. |
| <b>TOTAL</b>         | accumulates observations based on the total sum of their values.  |
| <b>AVERAGE   AVG</b> | accumulates observations based on the average of their values.  |
| <b>MINIMUM   MIN</b> | accumulates observations based on the minimum of their values.  |
| <b>MEDIAN   MED</b>  | accumulates observations based on the median of their values.   |
| <b>MAXIMUM   MAX</b> | accumulates observations based on the maximum of their values.  |
| <b>N</b>             | accumulates observations based on the number of nonmissing observations.  |
| <b>NMISS</b>         | accumulates observations based on the number of missing observations.   |
| <b>NOBS</b>          | accumulates observations based on the number of observations.   |
| <b>FIRST</b>         | accumulates observations based on the first of their values.  |
| <b>LAST</b>          | accumulates observations based on the last of their values.   |
| <b>STDDEV   STD</b>  | accumulates observations based on the standard deviation of their values.   |
| <b>CSS</b>           | accumulates observations based on the corrected sum of squares of their values.                                   |
| <b>USS</b>           | accumulates observations based on the uncorrected sum of squares of their values.                                 |

By default, `ACCUMULATE=NONE`.

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. For more information about accumulation, see the section “[Accumulation](#)” on page 852.

**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=date | datetime**

specifies a SAS date or datetime literal 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= `1jan2008` D** specifies that data for time periods after the first of January 2008 not be used. The option **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 contain the same number of observations.

**FORMAT=***format*

specifies the SAS format for the time ID values. If the FORMAT= option is not specified, the default format is implied from the INTERVAL= option.

**INTERVAL=***interval*

specifies the frequency of the input time series or for the time series to be accumulated from the input data. 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 by 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. For more information about the intervals that can be specified, see Chapter 4, “Date Intervals, Formats, and Functions.”

**NOTSORTED**

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

**SETMISSING=***option | number*

specifies how missing values (either input or accumulated) are assigned in the accumulated time series. If a number is specified, missing values are set to that number. If a missing value in the input data set indicates an unknown value, the SETMISSING= option should not be used. If a missing value indicates no value, SETMISSING=0 should be used. You 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:

|                        |   |
|------------------------|---|
| <b>MISSING</b>         | sets missing values to missing. The missing observations are replaced with predicted values that are computed from the exponential smoothing model. |
| <b>AVERAGE   AVG</b>   | sets missing values to the accumulated average value.   |
| <b>MINIMUM   MIN</b>   | sets missing values to the accumulated minimum value.   |
| <b>MEDIAN   MED</b>    | sets missing values to the accumulated median value.  |
| <b>MAXIMUM   MAX</b>   | sets missing values to the accumulated maximum value.   |
| <b>FIRST</b>           | sets missing values to the accumulated first nonmissing value.  |
| <b>LAST</b>            | sets missing values to the accumulated last nonmissing value.   |
| <b>PREVIOUS   PREV</b> | sets missing values to the previous accumulated nonmissing value. Missing values at the beginning of the accumulated series remain missing.         |
| <b>NEXT</b>            | sets missing values to the next accumulated nonmissing value. Missing values at the end of the accumulated series remain missing.                   |

By default, SETMISSING=MISSING.

**START=***date* | *datetime*

specifies a SAS date or datetime literal value that represents the beginning of the data. If the first time ID variable value is greater than the START= value, the series is prefixed 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 contain the same number of observations.

**ZEROMISS=**NONE | LEFT | RIGHT | BOTH

specifies how beginning and ending zero values (either input or accumulated) are interpreted in the accumulated time series. You can specify the following values:

- NONE** Beginning and ending zeros are unchanged.
- LEFT** Beginning zeros are set to missing.
- RIGHT** Ending zeros are set to missing.
- BOTH** Both beginning and ending zeros are set to missing.

By default, ZEROMISS=NONE.

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

---

## Details: ESM Procedure

The ESM procedure can be used to forecast time series data as well as transactional data. If the data are transactional, then the procedure must first accumulate the data into a time series before it can be forecast. The procedure uses the sequential steps in [Table 14.2](#) to produce forecasts, with the options that control the step listed to the right.

**Table 14.2** ESM Processing Steps and Control Options

| Step | Operation                    | Option            | Statements         |
|------|------------------------------|-------------------|--------------------|
| 1    | Accumulation                 | ACCUMULATE=       | ID                 |
| 2    | Missing value interpretation | SETMISSING=       | ID, FORECAST       |
| 3    | Transformations              | TRANSFORM=        | FORECAST           |
| 4    | Parameter estimation         | MODEL=            | FORECAST           |
| 5    | Forecasting                  | MODEL=, LEAD=     | FORECAST, PROC ESM |
| 6    | Inverse transformation       | TRANSFORM, MEDIAN | FORECAST           |
| 7    | Summation of forecasts       | LEAD=, STARTSUM=  | PROC ESM           |

Each of the steps shown in [Table 14.2](#) is described in the following sections.

---

### Accumulation

If the ACCUMULATE= option is specified in the ID statement, data set observations are accumulated within each time period. The frequency (width of each time interval) is specified by the INTERVAL= option,

and 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 time series that is used in subsequent analyses by the ESM procedure.

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

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

If the INTERVAL=MONTH option is specified in the ID statement, all of the preceding observations fall within three time periods: 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, the resulting time series is

```
O1MAR1999    40
O1APR1999    .
O1MAY1999    90
```

If the ACCUMULATE=AVERAGE option is specified, the resulting time series is

```
O1MAR1999    20
O1APR1999    .
O1MAY1999    30
```

If the ACCUMULATE=MINIMUM option is specified, the resulting time series is

```
O1MAR1999    10
O1APR1999    .
O1MAY1999    20
```

If the ACCUMULATE=MEDIAN option is specified, the resulting time series is

```
O1MAR1999    20
O1APR1999    .
O1MAY1999    20
```

If the ACCUMULATE=MAXIMUM option is specified, the resulting time series is

```
O1MAR1999    30
O1APR1999    .
O1MAY1999    50
```

If the ACCUMULATE=FIRST option is specified, the resulting time series is

```
O1MAR1999    10
O1APR1999    .
O1MAY1999    50
```

If the ACCUMULATE=LAST option is specified, the resulting time series is

```
O1MAR1999    30
O1APR1999    .
O1MAY1999    20
```

If the ACCUMULATE=STDDEV option is specified, the resulting time series is

```
O1MAR1999    14.14
O1APR1999    .
O1MAY1999    17.32
```

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

---

## Missing Value Interpretation

Sometimes missing values should be interpreted as truly unknown values and retained as missing values in the data set. The forecasting models used by the ESM procedure can effectively handle missing values (see the section “[Missing Value Modeling Issues](#)” on page 855). However, sometimes missing values are known, such as when missing values are created from accumulation and represent no observed values for the variable. In this case, the value for the period should be interpreted as zero (no values), and the SETMISSING=0 option should be used to cause PROC ESM to recode missing values as zero. In other cases, missing values should be interpreted as global values, such as minimum or maximum values of the accumulated series. The accumulated and missing-value-recoded time series is used in subsequent analyses in PROC ESM.

---

## Transformations

If the TRANSFORM= option is specified in the FORECAST statement, 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. (For more information about forecasting transformed time series, see Chapter 65, “[Forecasting Process Details](#).”)

---

## Parameter Estimation

All the parameters (smoothing weights) associated with the exponential smoothing model used to forecast the time series (as specified by the MODEL= option) 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.

The techniques used in the ESM procedure are identical to those used for exponential smoothing models in the Time Series Forecasting System of SAS/ETS software. For more information, see Chapter 57, “[Overview of the Time Series Forecasting System](#).”

---

## Missing Value Modeling Issues

The treatment of missing values varies with the forecasting model. Missing values after the start of the series are replaced with one-step-ahead predicted values, and the predicted values are used in the smoothing equations.

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

**NOTE:** Even if all of the observed data are nonmissing, the `ACCUMULATE=` option can create missing values in the accumulated series (when the data contain no observations for some of the time periods specified by the `INTERVAL=` option).

---

## Forecasting

Once the model parameters are estimated, one-step-ahead forecasts are generated for the full range of the accumulated and optionally transformed time series data, and multistep forecasts are generated from the end of the time series to the future time period 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 in the `FORECAST` statement, the forecasts of the transformed time series are inverse transformed. By default, forecasts of the mean (expected value) are generated. If the `MEDIAN` option is specified, median forecasts are generated. (For more information about forecasting transformed time series, see Chapter 65, “Forecasting Process Details.”)

---

## Statistics of Fit

The statistics of fit are computed by comparing the time series data (after accumulation and missing value recoding, if specified) with the generated forecasts. If the `TRANSFORM=` option is specified, the statistics of fit are based on the inverse transformed forecasts. (For more information about statistics of fit for forecasting models, see Chapter 65, “Forecasting Process Details.”)

---

## Forecast Summation

The multistep forecasts generated by the preceding steps can optionally be summed from the `STARTSUM=` value to the `LEAD=` value. For example, if the options `STARTSUM=4` and `LEAD=6` are specified in the `PROC` `ESM` statement, the four-step-ahead through six-step-ahead forecasts are summed.

The forecasts are simply summed; however, the prediction error variance of this sum is computed by taking into account the correlation between the individual predictions. (These variance-related computations are performed only when no transformation is specified; that is, when `TRANSFORM=NONE`.) 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 but 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.

---

## Data Set Output

The ESM procedure can create the OUT=, OUTEST=, OUTFOR=, OUTSTAT=, and OUTSUM= data sets. These data sets contain the variables listed in the BY statement and statistics related to the variables listing in the FORECAST statement. In general, if a forecasting step related to an output data set 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. For more information about how the variables in the output data sets are computed, see “[Smoothing Models](#)” on page 4156).

### 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, embedded missing values are replaced by the one-step-ahead predicted values.

These FORECAST variables are then extrapolated based on the forecasts from the fitted models, 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 contain either mean or median forecasts depending on whether or not the MEDIAN option is specified.

If any of the forecasting steps fail for a particular variable, the variable is 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 in the OUTEST= data set. For variables listed in FORECAST statements where the option MODEL=NONE is not specified, the following variables in the OUTEST= data set contain observations related to the parameter estimation step:

|                          |                   |
|--------------------------|-------------------|
| <code>_NAME_</code>      | variable name     |
| <code>_MODEL_</code>     | forecasting model |
| <code>_TRANSFORM_</code> | transformation    |
| <code>_PARM_</code>      | parameter name    |

|                       |                    |
|-----------------------|--------------------|
| <code>_EST_</code>    | parameter estimate |
| <code>_STDERR_</code> | standard errors    |
| <code>_TVALUE_</code> | <i>t</i> values    |
| <code>_PVALUE_</code> | probability values |

If the parameter estimation step fails for a particular variable, no observations are output to the `OUTEST=` data set for that variable.

### **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 in the `OUTFOR=` data set for these variables. For variables listed in `FORECAST` statements where the option `MODEL=NONE` is not specified, the following variables in the `OUTFOR=` data set contain observations related to the forecasting step:

|                       |                                    |
|-----------------------|------------------------------------|
| <code>_NAME_</code>   | variable name                      |
| <code>_TIMEID_</code> | time ID values                     |
| <code>ACTUAL</code>   | actual values                      |
| <code>PREDICT</code>  | predicted values                   |
| <code>STD</code>      | prediction standard errors         |
| <code>LOWER</code>    | prediction lower confidence limits |
| <code>UPPER</code>    | prediction upper confidence limits |
| <code>ERROR</code>    | prediction errors                  |

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

### **OUTPROCINFO= Data Set**

The `OUTPROCINFO=` data set contains information about the run of the ESM procedure. The following variables are present:

|                       |   |
|-----------------------|---|
| <code>_SOURCE_</code> | set to the name of the procedure, in this case ESM  |
| <code>_NAME_</code>   | name of an item being reported; can be the number of errors, notes, or warnings, number of forecasts requested, and so on |
| <code>_LABEL_</code>  | descriptive label for the item in <code>_NAME_</code>   |
| <code>_STAGE_</code>  | set to the current stage of the procedure; for <code>PROC ESM</code> this is set to <code>ALL</code>                      |
| <code>_VALUE_</code>  | value of the item specified in <code>_NAME_</code>  |

**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 in the OUTSTAT= data set. For variables listed in FORECAST statements where the option MODEL=NONE is not specified, the following variables in the OUTSTAT= data set contain observations related to the statistics of fit:

|                       |  |
|-----------------------|--|
| <code>_NAME_</code>   | variable name  |
| <code>_REGION_</code> | the region in which the statistics are calculated. Statistics calculated in the fit region are indicated by FIT. Statistics calculated in the forecast region, which happens only if the BACK= option is greater than zero, are indicated by FORECAST. |
| DFE                   | degrees of freedom error   |
| N                     | number of observations   |
| NOBS                  | number of observations used  |
| NMISSA                | number of missing actuals  |
| NMISSP                | number of missing predicted values   |
| NPARMS                | number of parameters   |
| TSS                   | total sum of squares   |
| SST                   | corrected total sum of squares   |
| 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  |
| MASE                  | mean absolute scaled error   |
| RSQUARE               | R-square   |
| ADJRSQ                | adjusted R-square  |
| AADJRSQ               | Amemiya's adjusted R-square  |
| RWRSQ                 | random walk R-square   |
| AIC                   | Akaike's information criterion   |
| AICC                  | finite sample corrected AIC  |
| SBC                   | Schwarz Bayesian information criterion   |
| APC                   | Amemiya's prediction criterion   |
| MAXERR                | maximum error  |
| MINERR                | minimum error  |

|         |   |
|---------|---|
| MINPE   | minimum percent error                                       |
| MAXPE   | maximum percent error                                       |
| ME      | mean error  |
| MPE     | mean percent error  |
| MDAPE   | median absolute percent error                               |
| GMAPE   | geometric mean absolute 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 absolute predictive percent error                    |
| GMAPPE  | geometric mean absolute 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 absolute symmetric percent error                     |
| GMASPE  | geometric mean absolute 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                      |
| MINAPES | minimum absolute error percent of standard deviation        |
| MAXAPES | maximum absolute error percent of standard deviation        |
| MAPES   | mean absolute error percent of standard deviation           |
| MDAPES  | median absolute error percent of standard deviation         |
| GMAPES  | geometric mean absolute error percent of standard deviation |

If the statistics of fit cannot be computed for a particular variable, no observations are recorded in the OUTSTAT= data set for that variable. If the TRANSFORM= option is specified, the values in the preceding variables 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.

For more information about the calculation of forecasting statistics of fit, see Chapter 65, “Forecasting Process Details.”

## 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 those variables in the OUTSUM= data set. For variables listed in FORECAST statements where the option MODEL=NONE is not specified, the forecast values are set based on the USE= option.

The following variables related to summary statistics are based on the ACCUMULATE= and SETMISSING= options:

|                       |   |
|-----------------------|---|
| <code>_NAME_</code>   | variable name   |
| <code>_STATUS_</code> | forecasting status. Nonzero values imply that no forecast was generated for the series. |
| <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>MEAN</code>     | mean value  |
| <code>STDDEV</code>   | standard deviation  |

The following variables related to forecast summation are based on the LEAD= and STARTSUM= options:

|                      |   |
|----------------------|---|
| <code>PREDICT</code> | forecast summation predicted values           |
| <code>STD</code>     | forecast summation prediction standard errors |
| <code>LOWER</code>   | forecast summation lower confidence limits    |
| <code>UPPER</code>   | forecast summation upper confidence limits    |

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

The following variables related to multistep forecast are based on the LEAD= and USE= options:

|                                   |   |
|-----------------------------------|---|
| <code>_LEAD<math>n</math>_</code> | multistep forecast ( $n$ ranges from one to the value of the LEAD= option). If USE=LOWER, this variable contains the lower confidence limits; if USE=UPPER, this variable contains the upper confidence limits; otherwise, this variable contains the predicted values. |
|-----------------------------------|---|

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

---

## Printed Output

The ESM procedure optionally produces printed output by using the Output Delivery System (ODS). By default, the procedure produces no printed output. All output is controlled by the PRINT= and PRINTDE-  
TAILS options in the PROC ESM statement. In general, if a forecasting step that is related to printed output

fails, the values of this step are not printed and appropriate error or warning messages are recorded in the log. The printed output is similar to the output data sets.

The printed output produced by the PRINT= option values is described as follows:

|                    |   |
|--------------------|---|
| SUMMARY            | prints the summary statistics and forecast summaries similar to the OUTSUM= data set. |
| ESTIMATES          | prints the parameter estimates similar to the OUTEST= data set.                       |
| FORECASTS          | prints the forecasts similar to the OUTFOR= data set.                                 |
| PERFORMANCE        | prints the performance statistics.  |
| PERFORMANCESUMMARY | prints the performance summary for each BY group.                                     |
| PERFORMANCEOVERALL | prints the performance summary for all BY groups.                                     |
| STATES             | prints the backcast, initial, and final smoothed states.                              |
| STATISTICS         | prints the statistics of fit similar to the OUTSTAT= data set.                        |

The PRINTDETAILS option is the opposite of the NOOUTALL option. Specifically, if PRINT=FORECASTS and the PRINTDETAILS options are specified in the PROC ESM statement, the one-step-ahead forecasts through the range of the data are printed in addition to 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

Table 14.3 relates the PRINT= options to ODS tables.

**Table 14.3** ODS Tables Produced in PROC ESM

| ODS Table Name        | Description                                   | PRINT= Option      |
|-----------------------|---|--------------------|
| DescStats             | Descriptive statistics                        | SUMMARY            |
| ForecastSummary       | Forecast summary                              | SUMMARY            |
| ForecastSummation     | Forecast summation                            | SUMMARY            |
| ParameterEstimates    | Parameter estimates                           | ESTIMATES          |
| Forecasts             | Forecasts                                     | FORECASTS          |
| Performance           | Performance statistics                        | PERFORMANCE        |
| PerformanceSummary    | Performance summary                           | PERFORMANCESUMMARY |
| PerformanceOverall    | Performance overall                           | PERFORMANCEOVERALL |
| SmoothedStates        | Smoothed states                               | STATES             |
| FitStatistics         | Evaluation statistics of fit                  | STATISTICS         |
| PerformanceStatistics | Performance (out-of-sample) statistics of fit | 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

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “Statistical Graphics Using ODS” (*SAS/STAT User’s Guide*).

Before you create graphs, ODS Graphics must be enabled (for example, with the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” in that chapter.

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “A Primer on ODS Statistical Graphics” in that chapter.

This section describes the use of ODS for creating graphics with the ESM procedure. To request these graphs you must specify the PLOT= option in the PROC ESM statement.

## ODS Graph Names

PROC ESM 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 14.4.

**Table 14.4** ODS Graphics Produced by the PLOT= Option in PROC ESM

| ODS Graph Name             | Plot Description  | PLOT= Option |
|----------------------------|---|--------------|
| ErrorACFNORMPlot           | Standardized autocorrelation of prediction errors         | ACF          |
| ErrorACFPlot               | Autocorrelation of prediction errors                      | ACF          |
| ErrorHistogram             | Prediction error histogram                                | ERRORS       |
| ErrorCorrelationPlots      | Prediction error plot panel                               | CORR         |
| ErrorIACFNORMPlot          | Standardized inverse autocorrelation of prediction errors | IACF         |
| ErrorIACFPlot              | Inverse autocorrelation of prediction errors              | IACF         |
| ErrorPACFNORMPlot          | Standardized partial autocorrelation of prediction errors | PACF         |
| ErrorPACFPlot              | Partial autocorrelation of prediction errors              | PACF         |
| ErrorPeriodogramPlot       | Periodogram of prediction errors                          | PERIODOGRAM  |
| ErrorPlot                  | Plot of prediction errors                                 | ERRORS       |
| ErrorSpectralDensityPlot   | Combined periodogram and spectral density estimate plot   | SPECTRUM     |
| ErrorWhiteNoiseLogProbPlot | White noise log probability plot of prediction errors     | WN           |
| ErrorWhiteNoiseProbPlot    | White noise probability plot of prediction errors         | WN           |

Table 14.4 *continued*

| ODS Graph Name     | Plot Description           | PLOT= Option   |
|--------------------|----------------------------|----------------|
| ForecastsOnlyPlot  | Forecasts only plot        | FORECASTSONLY  |
| ForecastsPlot      | Forecasts plot             | FORECASTS      |
| LevelStatePlot     | Smoothed level state plot  | LEVELS         |
| ModelForecastsPlot | Model and forecasts plot   | MODELFORECASTS |
| ModelPlot          | Model plot                 | MODELS         |
| SeasonStatePlot    | Smoothed season state plot | SEASONS        |
| TrendStatePlot     | Smoothed trend state plot  | TRENDS         |

## Examples: ESM Procedure

### Example 14.1: Forecasting of Time Series Data

This example uses retail sales data to illustrate how the ESM procedure can be used to forecast time series data.

The following DATA step creates a data set from data recorded monthly at numerous points of sale. The data set, SALES, contains 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;
01JAN1994 3557 3718 6368.80 575 987 10.8200 15.0000 102.600 12410 15013
... more lines ...
```

The following ESM procedure statements forecast each of the monthly time series:

```
proc esm data=sales out=nextyear;
  id date interval=month;
  forecast _numeric_;
run;
```

The preceding statements generate forecasts for every numeric variable in the input data set SALES for the next 12 months and store these forecasts in the output data set NEXTYEAR.

The following statements plot the forecasts:

```
title1 "Shoe Department Sales";
proc sgplot data=nextyear;
  series x=date y=shoes / markers
        markerattrs=(symbol=circlefilled color=red)
```

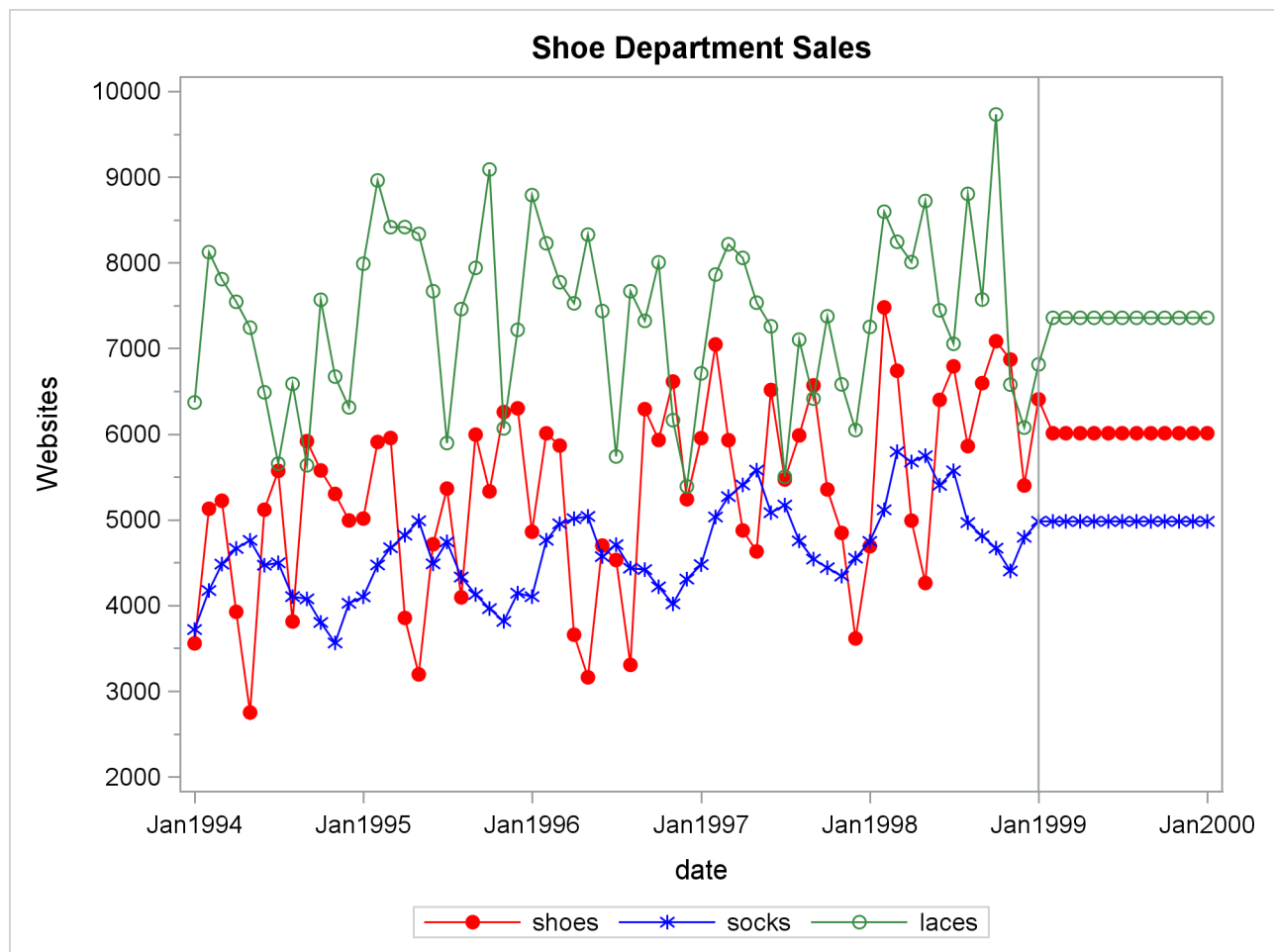
```

                                lineattrs=(color=red);
series x=date y=socks / markers
                                markerattrs=(symbol=asterisk color=blue)
                                lineattrs=(color=blue);
series x=date y=laces / markers
                                markerattrs=(symbol=circle color=styg)
                                lineattrs=(color=styg);
refline '01JAN1999'd / axis=x;
xaxis values=('01JAN1994'd to '01DEC2000'd by year);
yaxis values=(2000 to 10000 by 1000) minor label='Websites';
run;

```

The plots are shown in [Output 14.1.1](#). The historical data are shown to the left of the reference line, and the forecasts for the next 12 monthly periods are shown to the right.

**Output 14.1.1** Retail Sales Forecast Plots



The default simple exponential smoothing model is used because the `MODEL=` option is omitted from the `FORECAST` statement. Note that for simple exponential smoothing the forecasts are constant.

The following ESM procedure statements are identical to the preceding statements except that the `PRINT=FORECASTS` option is specified:

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

In addition to forecasting each of the monthly time series, the preceding statements print the forecasts by using the Output Delivery System (ODS); the forecasts are partially shown in [Output 14.1.2](#). This output shows the predictions, prediction standard errors, and upper and lower confidence limits for the next 12 monthly periods.

### Output 14.1.2 Forecast Tables

#### Shoe Department Sales

##### The ESM Procedure

---

Forecasts for Variable shoes

| Obs | Time    | Forecasts | Standard Error | 95% Confidence Limits |           |
|-----|---------|-----------|----------------|-----------------------|-----------|
| 62  | FEB1999 | 6009.1986 | 1069.4059      | 3913.2016             | 8105.1956 |
| 63  | MAR1999 | 6009.1986 | 1075.7846      | 3900.6996             | 8117.6976 |
| 64  | APR1999 | 6009.1986 | 1082.1257      | 3888.2713             | 8130.1259 |
| 65  | MAY1999 | 6009.1986 | 1088.4298      | 3875.9154             | 8142.4818 |
| 66  | JUN1999 | 6009.1986 | 1094.6976      | 3863.6306             | 8154.7666 |
| 67  | JUL1999 | 6009.1986 | 1100.9298      | 3851.4158             | 8166.9814 |
| 68  | AUG1999 | 6009.1986 | 1107.1269      | 3839.2698             | 8179.1274 |
| 69  | SEP1999 | 6009.1986 | 1113.2895      | 3827.1914             | 8191.2058 |
| 70  | OCT1999 | 6009.1986 | 1119.4181      | 3815.1794             | 8203.2178 |
| 71  | NOV1999 | 6009.1986 | 1125.5134      | 3803.2329             | 8215.1643 |
| 72  | DEC1999 | 6009.1986 | 1131.5758      | 3791.3507             | 8227.0465 |
| 73  | JAN2000 | 6009.1986 | 1137.6060      | 3779.5318             | 8238.8654 |

---

## Example 14.2: Forecasting of Transactional Data

This example illustrates how the ESM procedure can be used to forecast transactional data.

The following DATA step creates a data set from data recorded at several Internet websites. The data set WEBSITES contains a variable, TIME, that represents time and the variables ENGINE, BOATS, CARS, and PLANES that represent Internet website 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.

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

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

The preceding statements accumulate the data into a daily time series, generate forecasts for the BOATS, CARS, and PLANES variables in the input data set WEBSITES for the next week, and the forecasts are stored in the OUT= data set NEXTWEEK.

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

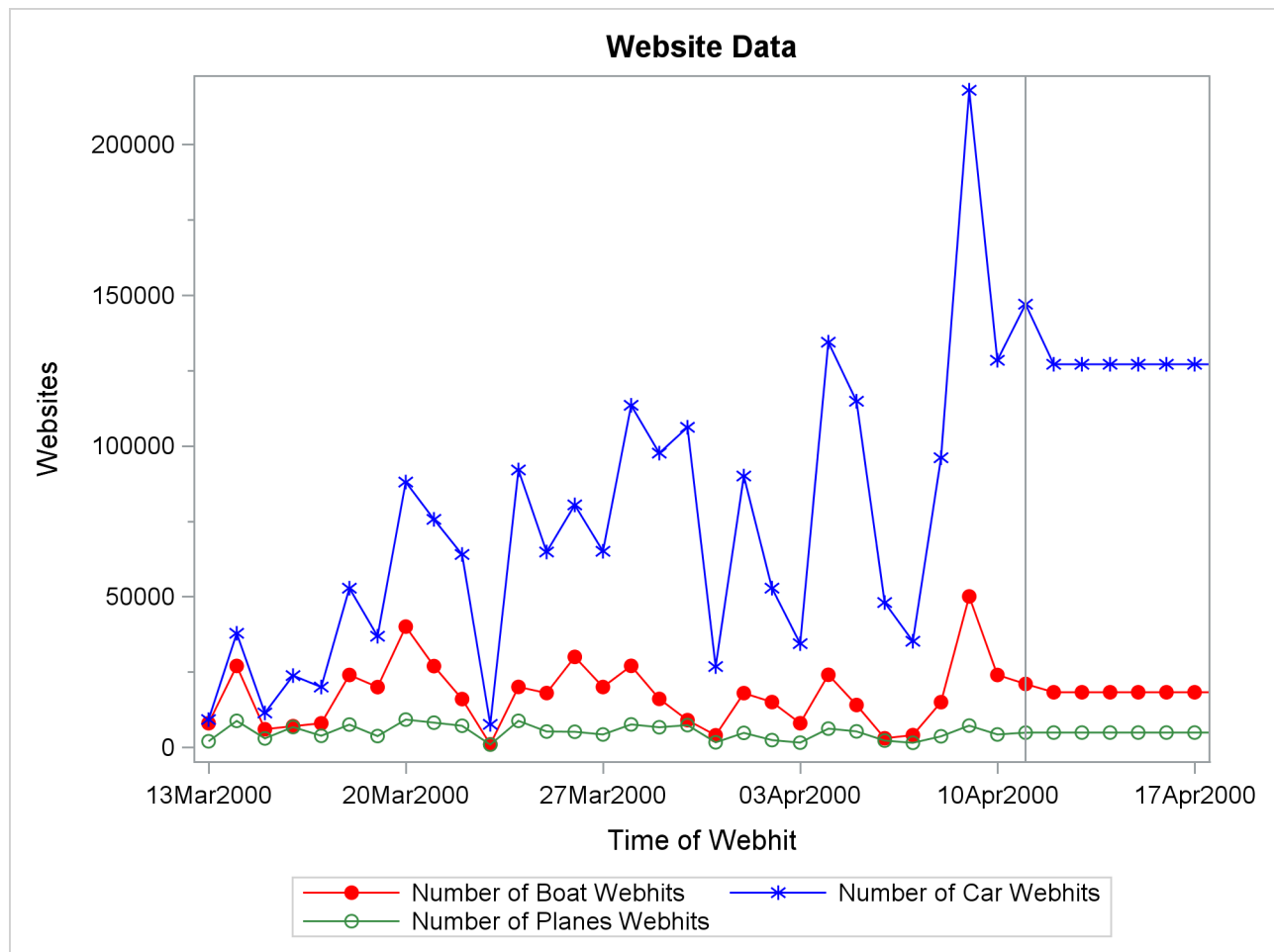
```

title1 "Website Data";
proc sgplot data=nextweek;
  series x=time y=boats / markers
        markerattrs=(symbol=circlefilled color=red)
        lineattrs=(color=red);
  series x=time y=cars / markers
        markerattrs=(symbol=asterisk color=blue)
        lineattrs=(color=blue);
  series x=time y=planes / markers
        markerattrs=(symbol=circle color=styg)
        lineattrs=(color=styg);
  refline '11APR2000:00:00:00'dt / axis=x;
  xaxis values=('13MAR2000:00:00:00'dt to '18APR2000:00:00:00'dt by dtweek);
  yaxis label='Websites' minor;
run;

```

The plots are shown in [Output 14.2.1](#). The historical data are shown to the left of the reference line, and the forecasts for the next seven days are shown to the right.

**Output 14.2.1** Internet Data Forecast Plots



---

### Example 14.3: Specifying the Forecasting Model

This example illustrates how the ESM procedure can be used to specify different models for different series. Internet data from the previous example are used for this illustration.

This example forecasts the BOATS variable by using the seasonal exponential smoothing model (SEASONAL), the CARS variable by using the Winters (multiplicative) model (MULTWINTERS), and the PLANES variable by using the Log Winters (additive) model. The following ESM procedure statements forecast each of the transactional data series based on these requirements:

```
proc esm data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast boats / model=seasonal;
  forecast cars / model=multwinters;
  forecast planes / model=addwinters transform=log;
run;
```

---

### Example 14.4: Extending the Independent Variables for Multivariate Forecasts

In the previous example, the ESM procedure was used to forecast several transactional series variables by using univariate models. This example illustrates how the ESM procedure can be used to extend the independent variables that are associated with a multiple regression forecasting problem.

This example accumulates and forecasts the BOATS, CARS, and PLANES variables that were illustrated in the previous example. In addition, this example accumulates 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 esm data=websites out=nextweek lead=7;
  id time interval=dtday accumulate=total;
  forecast engines / model=none;
  forecast boats / model=seasonal;
  forecast cars / model=multwinters;
  forecast planes / model=addwinters 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 NEXTWEEK data set created by PROC ESM is used as an input data set to PROC AUTOREG. The output data set from PROC AUTOREG contains the forecast of the variable ENGINES based on the regression model with the variables BOATS, CARS, and PLANES as regressors. For more information about autoregression models, see Chapter 8, “[The AUTOREG Procedure](#).”

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

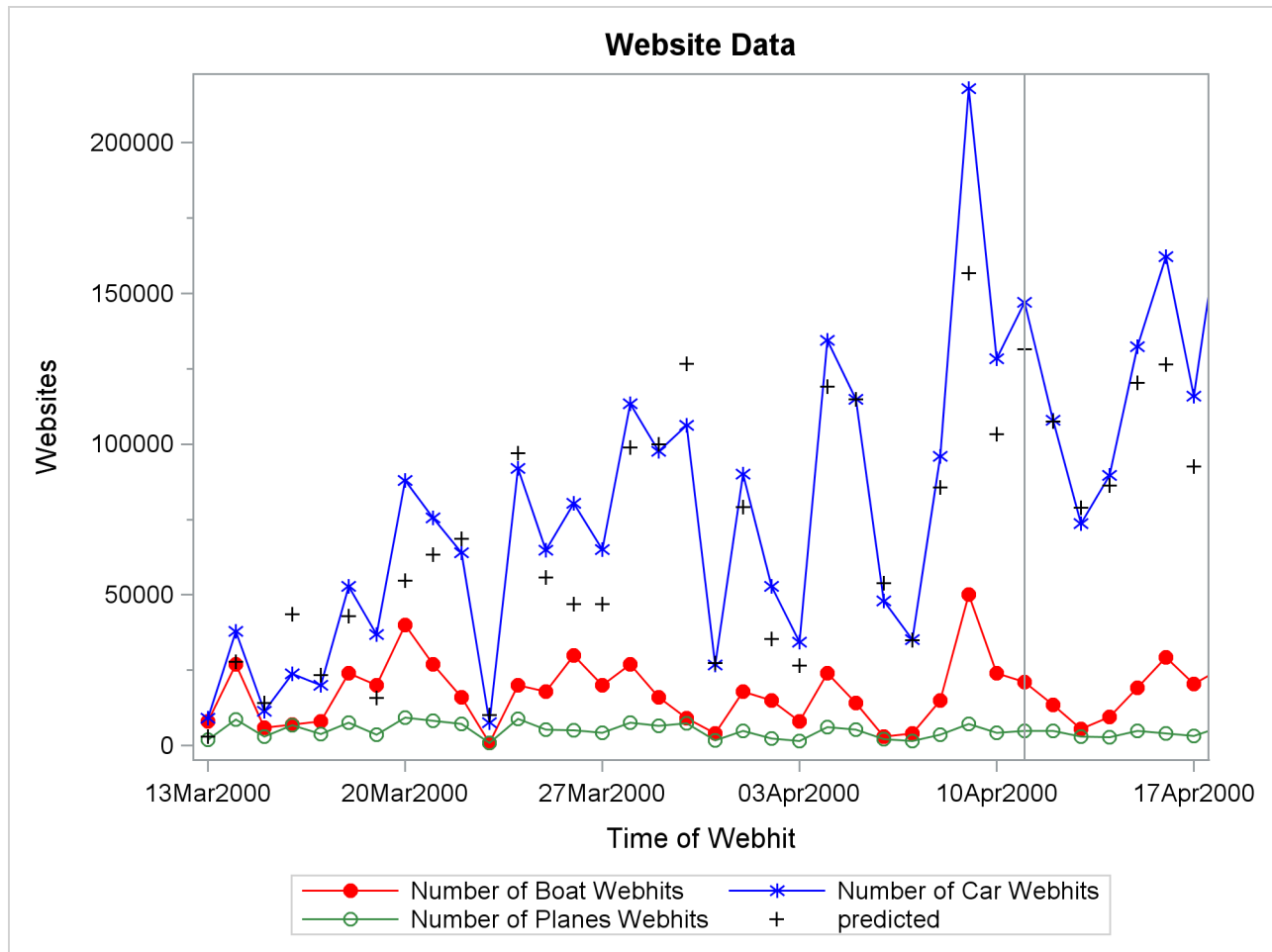
```

title1 "Website Data";
proc sgplot data=enginehits;
  series x=time y=boats / markers
          markerattrs=(symbol=circlefilled color=red)
          lineattrs=(color=red);
  series x=time y=cars / markers
          markerattrs=(symbol=asterisk color=blue)
          lineattrs=(color=blue);
  series x=time y=planes / markers
          markerattrs=(symbol=circle color=styg)
          lineattrs=(color=styg);
  scatter x=time y=predicted / markerattrs=(symbol=plus color=black);
  refline '11APR2000:00:00:00'dt / axis=x;
  xaxis values=('13MAR2000:00:00:00'dt to '18APR2000:00:00:00'dt by dtweek);
  yaxis label='Websites' minor;
run;

```

The plots are shown in [Output 14.4.1](#). The historical data are shown to the left of the reference line, and the forecasts for the next seven daily periods are shown to the right.

**Output 14.4.1** Internet Data Forecast Plots



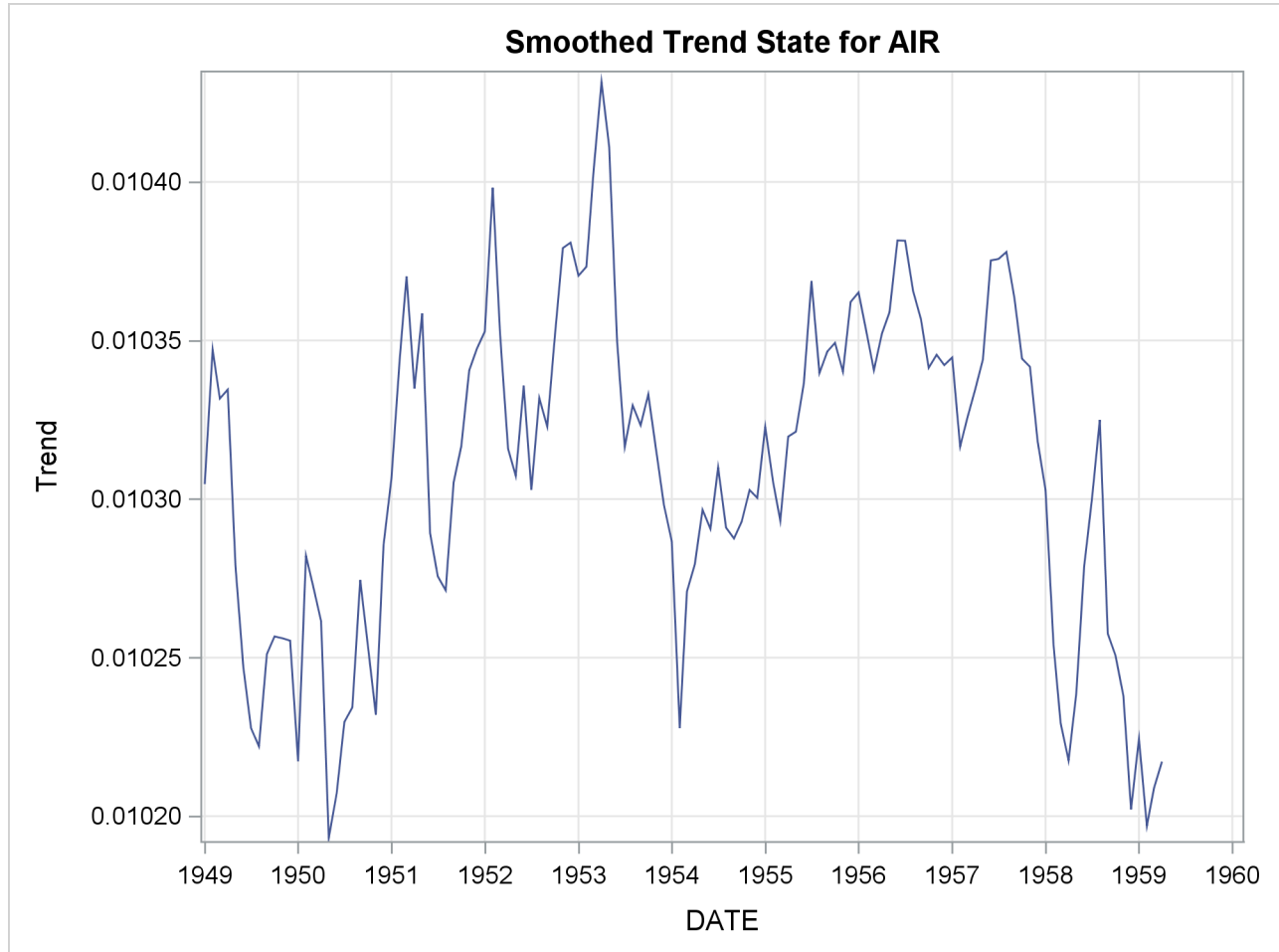
### Example 14.5: Illustration of ODS Graphics

This example illustrates the use of ODS graphics in the ESM procedure and uses the SASHELP.AIR data set to forecast the time series of international airline travel.

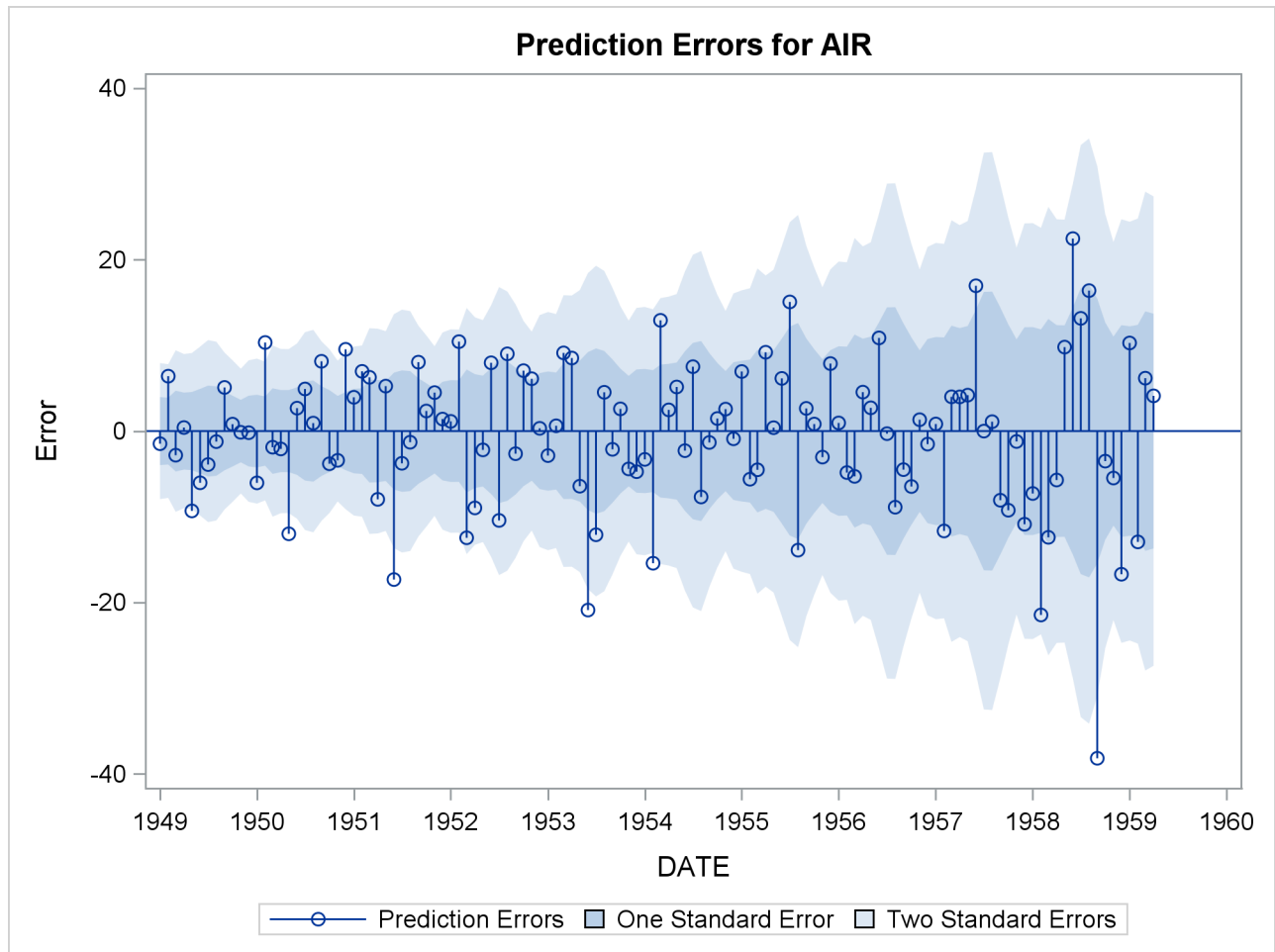
The graphical displays are requested by specifying the `PLOT=` option in the `PROC ESM` statement. In this case, all plots are requested. [Output 14.5.1](#) through [Output 14.5.5](#) show a selection of the plots created.

For information about the graphics available in the ESM procedure, see the section “[ODS Graphics](#)” on page 862.

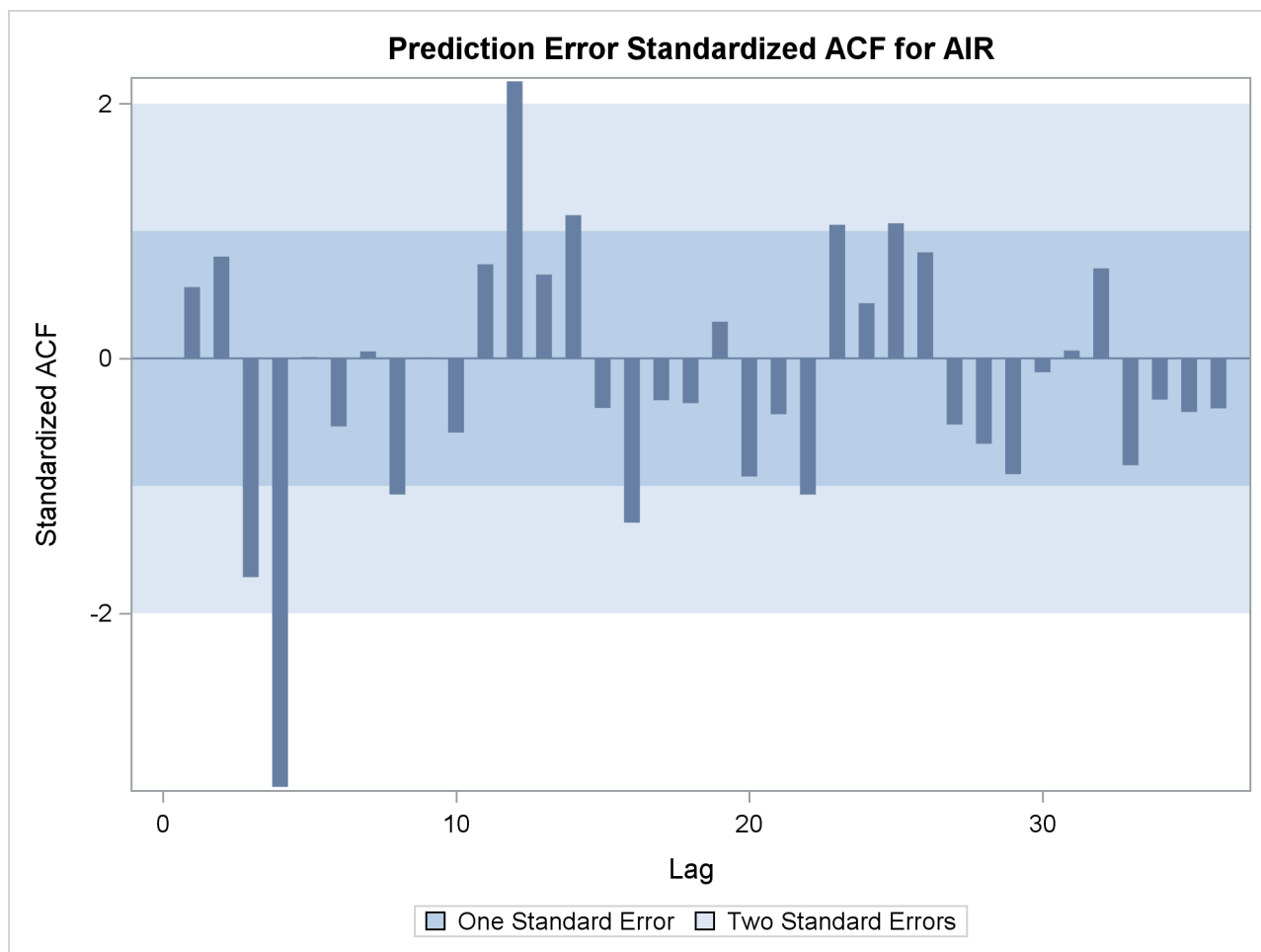
```
proc esm data=sashelp.air out=_null_  
  lead=20  
  back=20  
  print=all  
  plot=all;  
  id date interval=month;  
  forecast air / model=addwinters transform=log;  
run;
```

**Output 14.5.1** Smoothed Trend Plot

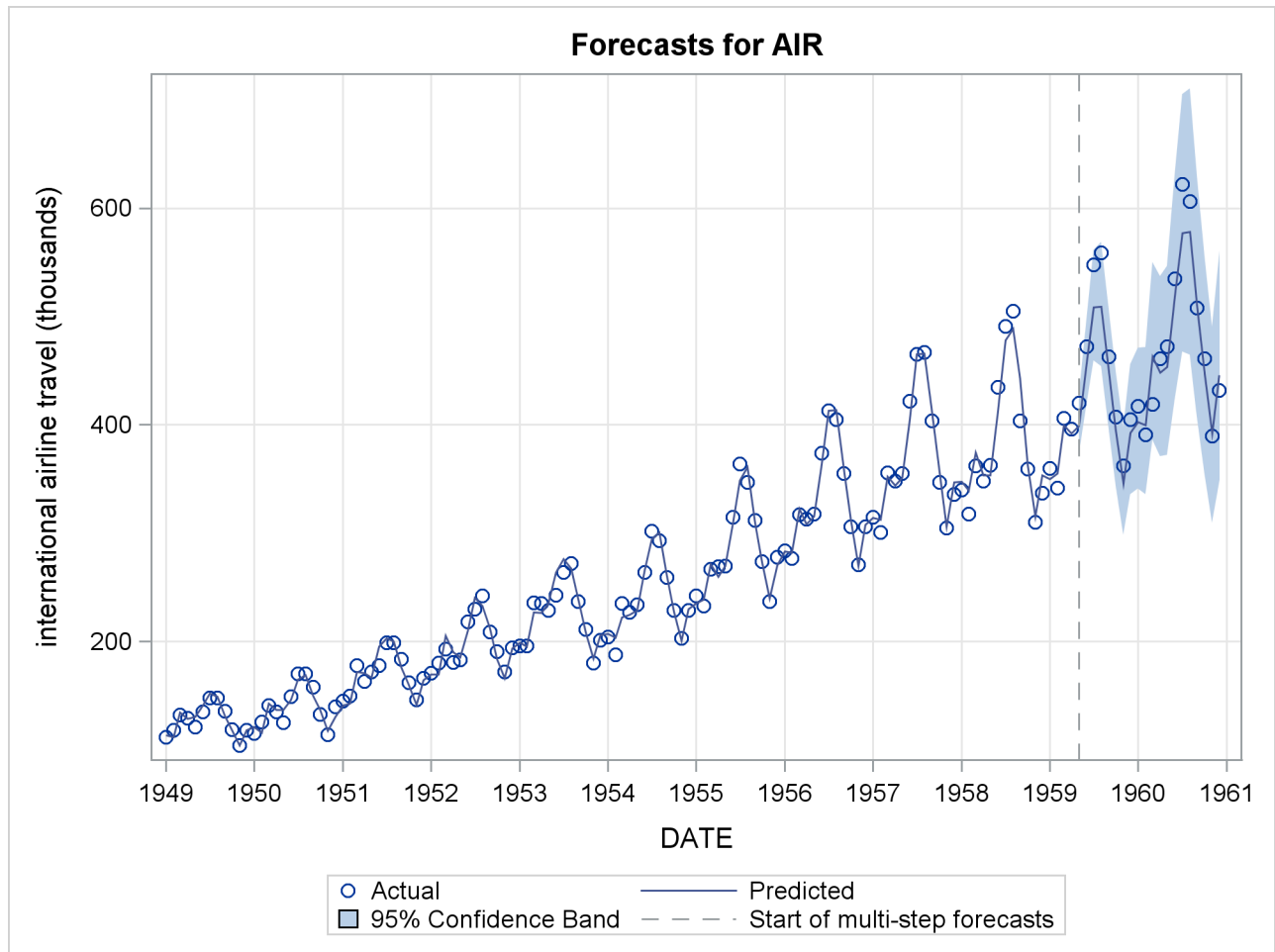
Output 14.5.2 Prediction Error Plot



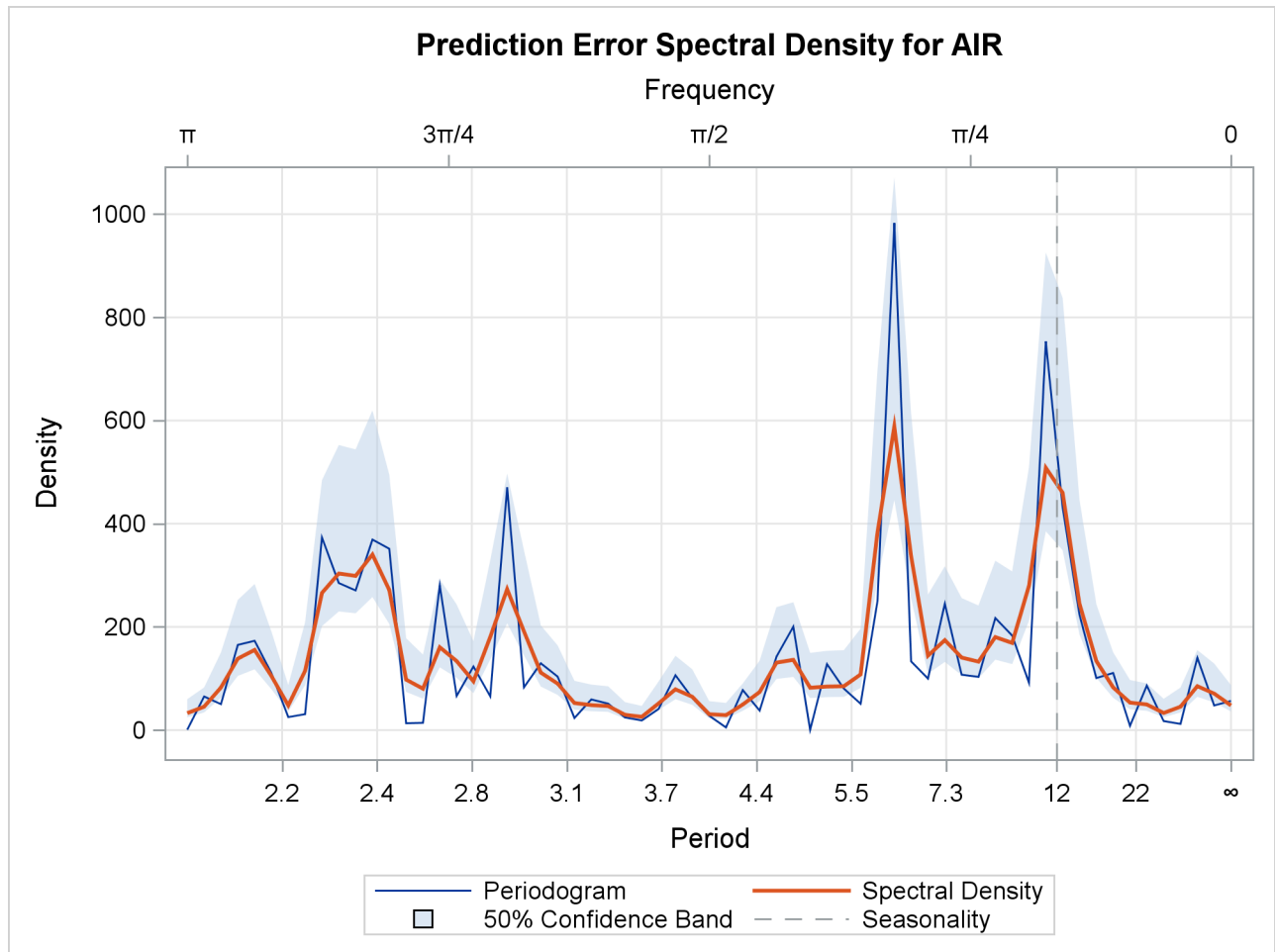
**Output 14.5.3** Prediction Error Standardized ACF Plot



Output 14.5.4 Forecast Plot



**Output 14.5.5** Prediction Error Spectral Density



# Subject Index

BY groups

ESM procedure, 847

ESM procedure

BY groups, 847

ODS graph names, 862

ODS graph names

ESM procedure, 862



# Syntax Index

- ACCUMULATE= option
  - FORECAST statement (ESM), 847
  - ID statement (ESM), 850
- ALIGN= option
  - ID statement (ESM), 850
- ALPHA= option
  - FORECAST statement (ESM), 847
- BACK= option
  - PROC ESM statement, 844
- BY statement
  - ESM procedure, 847
- DATA= option
  - PROC ESM statement, 844
- END= option
  - ID statement (ESM), 850
- ESM, 839
- ESM procedure, 842
  - syntax, 842
- FORECAST statement
  - ESM procedure, 847
- FORMAT= option
  - ID statement (ESM), 851
- ID statement
  - ESM procedure, 849
- INTERVAL= option
  - ID statement (ESM), 851
- LEAD= option
  - PROC ESM statement, 844
- MAXERROR= option
  - PROC ESM statement, 844
- MEDIAN option
  - FORECAST statement (ESM), 848
- MODEL= option
  - FORECAST statement (ESM), 848
- NBACKCAST= option
  - FORECAST statement (ESM), 848
- NOOUTALL option
  - PROC ESM statement, 844
- NOTSORTED option
  - ID statement (ESM), 851
- OUT= option
  - PROC ESM statement, 844
- OUTEST= option
  - PROC ESM statement, 844
- OUTFOR= option
  - PROC ESM statement, 845
- OUTPROCINFO= option
  - PROC ESM statement, 845
- OUTSTAT= option
  - PROC ESM statement, 845
- OUTSUM= option
  - PROC ESM statement, 845
- PLOT= option
  - PROC ESM statement, 845
- PRINT=option
  - PROC ESM statement, 846
- PRINTDETAILS option
  - PROC ESM statement, 846
- PROC ESM statement, 844
- REPLACEBACK option
  - FORECAST statement (ESM), 848
- REPLACEMISSING option
  - FORECAST statement (ESM), 848
- SEASONALITY= option
  - PROC ESM statement, 846
- SETMISSING= option
  - FORECAST statement (ESM), 848
  - ID statement (ESM), 851
- SORTNAMES option
  - PROC ESM statement, 846
- START= option
  - ID statement (ESM), 852
- STARTSUM= option
  - PROC ESM statement, 846
- TRANSFORM= option
  - FORECAST statement (ESM), 848
- USE= option
  - FORECAST statement (ESM), 849
- ZEROMISS= option
  - FORECAST statement (ESM), 849
- ZEROMISSING= option
  - ID statement (PROC ESM), 852