

Chapter 465

Exponential Smoothing – Horizontal

Introduction

Simple exponential smoothing forecasts horizontal series: those without trends or seasonal patterns. It is appropriate for short-term forecasts of series using a weighted average of the most recent observations.

The forecasting algorithm makes use of the following formulas:

$$F_t = \alpha X_t + (1 - \alpha)F_{t-1}$$

Here α is the smoothing constant which is between zero and one.

The forecast at time T for the value at time $T+k$ is F_T .

Another form of the above equation which shows how this procedure received its name is

$$F_t = \alpha X_t + \alpha(1 - \alpha)X_{t-1} + \alpha(1 - \alpha)^2 X_{t-2} + \alpha(1 - \alpha)^3 X_{t-3} + \dots$$

From this equation we see that the method constructs a weighted average of the observations. The weight of each observation decreases exponentially as we move back in time. Hence, since the weights decrease exponentially and averaging is a form of smoothing, the technique was named exponential smoothing.

Smoothing Constants

Notice that the *smoothing constant*, α , determines how fast the weights of the series decay. The value may be chosen either subjectively or objectively. Values near one put almost all weight on the most recent observations. Values of the smoothing constant near zero allow the distant past observations to have a large influence.

When selecting the smoothing constant *subjectively*, you use your own experience with this, and similar, series. Also, specifying the smoothing constant yourself lets you tune the forecast to your own beliefs about the future of the series. If you believe that the mechanism generating the series has recently gone through some fundamental changes, use a smoothing constant value of 0.9 which will cause distant observations to be ignored. If, however, you think the series is fairly stable and only going through random fluctuations, use a value of 0.1.

To select the value of the smoothing constant *objectively*, you search for a value that is best in some sense. Our program searches for that value that minimizes the size of the combined forecast errors of the currently available series. Three methods of summarizing the amount of error in the forecasts are available: the mean square error (MSE), the mean absolute error (MAE), and the mean absolute percent error (MAPE). The forecast error is the difference between the forecast of the current period made at the last period and the value of the series at the current period. This is written as

$$e_t = X_t - F_{t-1}$$

Exponential Smoothing – Horizontal

Using this formulation, we can define the three error-size criterion as follows:

$$MSE = \frac{1}{n} \sum e_t^2$$

$$MAE = \frac{1}{n} \sum |e_t|$$

$$MAPE = \frac{100}{n} \sum \left| \frac{e_t}{X_t} \right|$$

To find the value of the smoothing constants objectively, we select one of these criterion and search for that value of α that minimize this function. The program conducts a search for the appropriate values using an efficient grid-searching algorithm.

Initial Values

Exponential smoothing requires special initialization since the forecast for period one requires the forecast at period zero, which we do not, by definition, have. Several methods have been proposed for generating starting values. We have adopted the backcasting method which is currently considered to be one of the best. Backcasting is simply reversing the series so that we forecast into the past instead of into the future. This produces the required starting value. Once we have done this, we can then switch the series back and apply the algorithm in the regular manor.

Relationship to ARIMA Method

It can be shown that both exponential smoothing is equivalent to the ARIMA(0,1,1) model (see Kendall and Ord (1990) page 130). This is why backcasting is recommended for initial values.

Assumptions and Limitations

This algorithm is useful for short-term forecasting of nonseasonal time series with no apparent upward or downward. The series is assumed to have a changing (or evolving) mean that is not fixed over all time. We assume that future values of this average are unpredictable, so that the current level (current average or mean) of the series is the best forecast of future values.

Data Structure

The data are entered in a single variable.

Missing Values

When missing values are found in the series, they are either replaced or omitted. The replacement value is the average of the nearest observation in the future and in the past or the nearest non-missing value in the past.

If you do not feel that this is a valid estimate of the missing value, you should manually enter a more reasonable estimate before using the algorithm. These missing value replacement methods are particularly poor for seasonal data. We recommend that you replace missing values manually before using the algorithm.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

Specify the variable(s) on which to run the analysis.

Time Series Variables

Time Series Variable(s)

Specify the variable(s) on which to run the analysis. A separate analysis will be conducted for each variable listed.

Use Logarithms

Specifies that the log (base 10) transformation should be applied to the values of the variable. The forecasts are converted back to the original metric before display.

Missing Values

Choose how missing (blank) values are processed.

The algorithm used in this procedure cannot tolerate missing values since each row is assumed to represent the next point in a time sequence. Hence, when missing values are found, they must be removed either by imputation (filling in with a reasonable value) or by skipping the row and pretending it does not exist.

Whenever possible, we recommend that you replace missing values manually.

Here are the available options.

Average the Adjacent Values

Replace the missing value with the average of the nearest values in the future (below) and in the past (above).

Carry the Previous Value Forward

Replace the missing value with the first non-missing value immediately above (previous) this value.

Omit Row from Calculations

Ignore the row in all calculations. Analyze the data as if the row was not on the database.

Forecasting Options

Number of Forecasts

This option specifies the number of forecasts to be generated.

Smoothing Constant Search Options

Search Method

This option specifies whether a search is conducted for the best value of the smoothing constant and what the criterion for the search will be.

- **Specified Value**
No search is conducted. Use the value of the smoothing constant that is set in Alpha box.
- **Search on MSE**
A search is conducted to find the value of the smoothing constant that minimizes MSE.

Exponential Smoothing – Horizontal

- **Search on MAE**

A search is conducted to find the value of the smoothing constant that minimizes MAE.

- **Search on MAPE**

A search is conducted to find the value of the smoothing constant that minimizes MAPE.

Alpha Smoothing Constant

When the Search Method is set to Specified Value, this option specifies the value of the smoothing constant to be used. The limits of this value are zero and one. Usually, a value between 0.1 and 0.3 are used. As the value gets closer to one, more and more weight is given to recent observations.

Reports Tab

The following options control which reports are displayed.

Select Reports

Summary Report

This option specifies whether the indicated report is displayed.

Forecast Report

This option specifies which parts of the series are listed on the numeric reports: the original data and forecasts, just the forecasts, or neither.

Report Options

Precision

Specify the precision of numbers in the report. Single precision will display seven-place accuracy, while the double precision will display thirteen-place accuracy. Note that all reports are formatted for single precision only.

Variable Names

Specify whether to use variable names or (the longer) variable labels in report headings.

Page Title

Specify a title to be shown at the top of the reports.

Plots Tab

A plot of the data and forecast over time may be displayed, as well as the residual plot. Click the plot format button to change the plot settings.

Select Plots

Forecast Plot - Residual Plot

Each of these options specifies whether the indicated plot is displayed.

Exponential Smoothing – Horizontal

Horizontal Axis Variable if there are Missing or Filtered Values

Horizontal Variable

This option controls the spacing on the horizontal axis when missing or filtered values occur.

Your choices are

Actual Row Number

Use the actual row number of each row from the dataset along the horizontal axis.

Sequence Number

Use the sequence (relative row) number formed by ignoring any missing or filtered values.

Storage Tab

The forecasts and residuals may be stored on the current dataset for further analysis. These options let you designate which statistics (if any) should be stored by designating which columns should receive the statistics.

Note that existing data is replaced. Be careful that you do not specify columns that contain important data.

Data Storage Columns

Forecasts

The forecasts are stored in this column.

Residuals

The residuals are stored in this column.

Example 1 – Horizontal Exponential Smoothing

This section presents an example of how to generate a forecast of a horizontal series. The data in the Intel dataset gives price and volume data for Intel stock during August, 1995. We will forecast values for daily volumes. These values are contained in the variable Intel_Volume.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the Exponential Smoothing – Horizontal window.

1 Open the Intel dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Intel.NCSS**.
- Click **Open**.

2 Open the Exponential Smoothing – Horizontal window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Exponential Smoothing – Horizontal** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- On the Exponential Smoothing – Horizontal window, select the **Variables tab**.
- Double-click in the **Time Series Variable** box. This will bring up the variable selection window.
- Select **Intel_Volume** from the list of variables and then click **Ok**.

4 Specify the reports.

- On the Exponential Smoothing – Horizontal window, select the **Reports tab**.
- Select **Data and Forecasts** in the **Forecast Report** list box.

5 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Forecast Summary Section

Forecast Summary Section	
Variable	Intel_Volume
Number of Rows	20
Missing Values	None
Mean	10974.54
Pseudo R-Squared	0.000000
Mean Square Error	1.632774E+07
Mean Error	2876.168
Mean Percent Error	25.98573
Alpha Search	Mean Square Error
Alpha	0.3769885
Forecast	13100.84

This report summarizes the forecast equation.

Variable

The name of the variable for which the forecasts are generated.

Number of Rows

The number of rows that were in the series. This is provided to allow you to double-check that the correct series was used.

Exponential Smoothing – Horizontal

Missing Values

If missing values were found, this option lists the method used to estimate them.

Mean

The mean of the variable across all time periods.

Pseudo R-Squared

This value generates a statistic that acts like the R-Squared value in multiple regression. A value near zero indicates a poorly fitting model, while a value near one indicates a well-fitting model. The statistic is calculated as follows:

$$R^2 = 100 \left(1 - \frac{SSE}{SST} \right)$$

where *SSE* is the sum of square residuals and *SST* is the total sum of squares after correcting for the mean.

Mean Square Error

The average squared residual (MSE) is a measure of how closely the forecasts track the actual data. The statistic is popular because it shows up in analysis of variance tables. However, because of the squaring, it tends to exaggerate the influence of outliers (points that do not follow the regular pattern).

Mean |Error|

The average absolute residual (MAE) is a measure of how closely the forecasts track the actual data without the squaring.

Mean |Percent Error|

The average percent absolute residual (MAPE) is a measure of how closely the forecasts track the actual data put on a percentage basis.

Alpha Search

If a search was made to find the best value of the smoothing constant, this row gives the criterion used during the search.

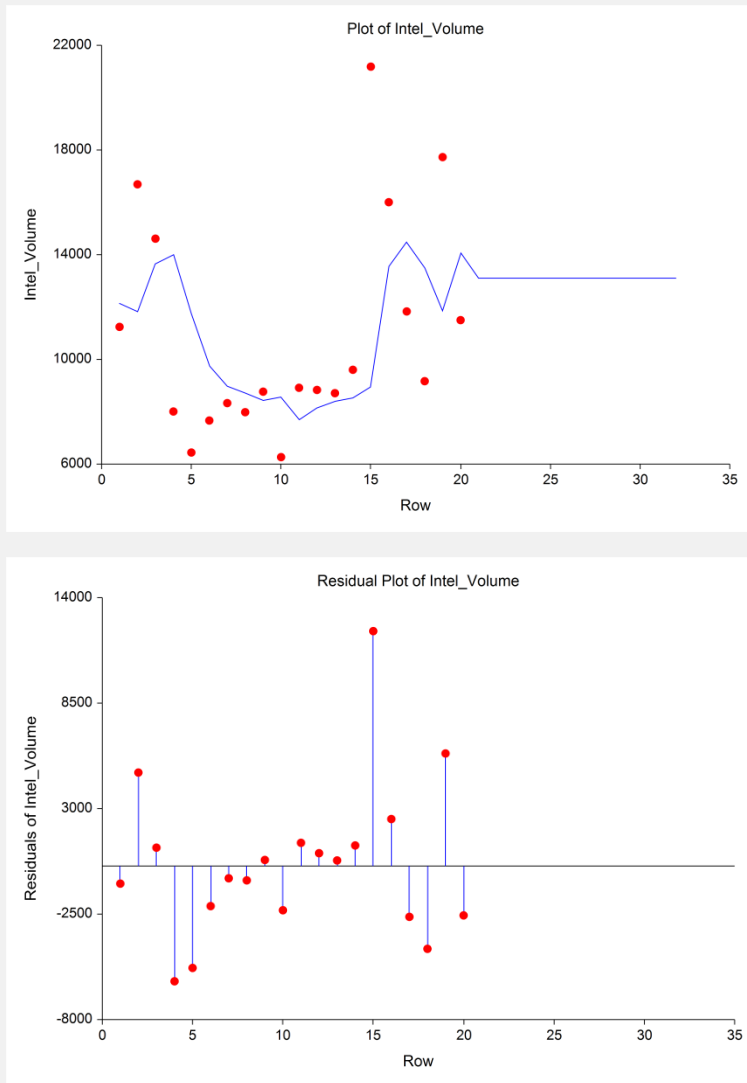
Alpha

The value of the smoothing constant that was used to generate the forecasts.

Forecast

The value of the forecast. This is the value that used to forecast future values from this point on. Remember that this method does not adjust for trend or seasonality, so only the current average is used for forecasting.

Forecast and Residuals Plots



Forecast Plot

The forecast plot lets you analyze how closely the forecasts track the data. The plot also shows the forecasts at the end of the data series.

Residual Plot

This plot lets you analyze the residuals themselves. You are looking for patterns, outliers, or any other information that may help you improve the forecasting model. The first thing to compare is the scale of the Residual Plot versus the scale of the Forecast Plot. If your forecasting is working well, the vertical scale of the Residual Plot will be much less than the scale of the Forecast Plot.

Forecasts Section

Forecasts Section

Row No.	Forecast Intel_Volume	Actual Intel_Volume	Residuals
1	12153.88	11242.2	-911.6825
2	11810.19	16689.9	4879.711
3	13649.78	14613.3	963.5162
4	14013.02	8009	-6004.018
5	11749.57	6441.8	-5307.772
6	9748.604	7664.5	-2084.103
7	8962.92	8330.3	-632.6202
8	8724.43	7983	-741.4297
9	8444.919	8767.1	322.1808
10	8566.378	6266.4	-2299.978
11	7699.313	8915.3	1215.987
12	8157.726	8833	675.2742
13	8412.297	8709.7	297.4036
14	8524.414	9603	1078.586
15	8931.028	21185.2	12254.17
16	13550.71	16006.5	2455.79
17	14476.51	11832.4	-2644.115
18	13479.71	9168.1	-4311.614
19	11854.29	17729.3	5875.015
20	14069.1	11500.7	-2568.398
21	13100.84		
22	13100.84		
23	13100.84		
24	13100.84		
25	13100.84		
26	13100.84		
27	13100.84		
28	13100.84		
29	13100.84		
30	13100.84		
31	13100.84		
32	13100.84		

This section shows the values of the forecasts, the actual values, and the residuals.