4. Forecasting Trends: Exponential Smoothing

What's past is prolog (Shakespeare, The Tempest)

Introduction

In chapter 3 we considered forecasts that were based upon subjective assessments. Such procedures are very valuable when we have no "track record" on which to base our forecasts, or when circumstances appear to have undergone fundamental changes. However, subjective forecasts are often time-consuming to generate and may be subject to a variety of conscious or unconscious biases, as we noted in that discussion. Often, we find that even simple analyses of available data can perform as well as judgmental procedures, and may be much quicker and less expensive to produce. The effective possible choices are judgment only, quantitative method only and quantitative method with results adjusted by user judgment. All three options have their place in the forecasting lexicon, depending upon costs, available data and the importance of the task in hand. Careful subjective adjustment of quantitative forecasts may often the best combination, but we first need to develop an effective arsenal of quantitative methods. Accordingly, we will now focus upon quantitative methods and return to combining issues in Chapter 12.

We begin in section 4.1 by drawing a distinction between *methods* and *models*, a distinction that is often ignored but which has important implications for the way in which we approach a forecasting task. Section 4.2 then provides a general overview of extrapolation methods before we move on in section 4.3 to the use of different weighted means as forecasts. Our aim is to balance the potentially conflicting directives:

- Use all the data, and
- Pay more attention to the recent past.

Or, to restate "All data are important, but recent data are even more important." These ideas lead directly to the use of time-dependent averages, notably simple moving averages and exponentially weighted moving averages.

When there are clear trends in the data, a simple averaging procedure cannot capture the trend and therefore will not work. As a consequence we must extend our methods to incorporate such systematic movements, the subject of sections 4.4 and 4.5. Sections 4.6 and 4.7 consider more specialized topics, including the use of transformations and damped smoothing methods; section 4.8 briefly examines other approaches to trend forecasting. Section 4.9 provides prediction intervals for one-step-ahead forecasts for these various methods, relying implicitly upon models we will develop in Chapter 6. Finally, in section 4.10 we explore some of the principles that underlie the use of simple extrapolation methods.

Software

The level of support for exponential smoothing methods varies considerably across different software providers. For example, Excel provides only simple exponential smoothing and, even then, the user must specify the smoothing constant. Further, different programs use different starting values and a variety of methods for estimating the smoothing parameters. In order to provide some consistency in the numerical results for these methods, we provide two Excel macros to carry out the estimation procedures. These macros are available on the book's web site and are described in detail in Appendices 4A and 4B at the end of this chapter.

4.1 Method or Model?

The development of various quantitative approaches to forecasting will occupy the next eight chapters, as we move steadily from heuristic forecasting methods to procedures that rely upon careful modeling of the process being studied. One of the pleasant discoveries we make along the way is that the heuristic methods often match up to specific models, even though this was not known at the time they were first proposed. As in other areas of research, we often proceed by discovering what works and then we try to figure out why. This chapter concentrates on what works and we emphasize the distinction in our terminology even though the difference between the two terms is sometimes ignored in discussions about forecasting.

> A *forecasting method* is a (numerical) procedure for generating a forecast. When such methods are not based upon an underlying statistical model, they are termed *heuristic*.
>
> A *statistical* (*forecasting) model* is a statistical description of the data generating process from which a forecasting method may be derived. Forecasts are made by using a forecast function that is derived from the model.

Example 4.1: Methods and Models

A forecasting method

Suppose that we decide to forecast sales as a linear function of time. That is, the forecasting method involves the specification

$$F_t = b_0 + b_1 t \qquad (4.1)$$

Here, F_t denotes a forecast for time period t, b_0 is the *intercept* that represent the value at time zero and b_1 is the *slope*, which represent the increase in forecast values from one period to the next.

Once we choose the parameters, that is the intercept, b_0 , and the slope, b_1, the forecasting method is primed to go. The advantages of such an approach are that it is simple to set up and to operate. The drawbacks are that we lack a basis for choosing values for the parameters, although various ad-hoc procedures are available. Finally, there is no way to assess the uncertainty inherent in the forecasts.

A forecasting model

In a similar fashion we may formulate a forecasting model as:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon \qquad (4.2)$$

Here Y denotes the time series being studied, β_0 and β_1 are the level and slope parameters, and ε denotes a random error term corresponding to that part of the series that cannot be fitted by the trend line. Once we make appropriate assumptions about the nature of the error term, we can estimate the unknown parameters, β_0 and β_1. These estimates are typically written as b_0 and b_1. Thus the forecasting *model* gives rise to a forecast *function*, which may be written as:

$$F_t = b_0 + b_1 t \qquad (4.3)$$

But is not equation (4.3) the same as equation (4.1)? Precisely! But the difference is that the underlying model, equation (4.2), enables us to make statements about the uncertainty in the forecast, something that the heuristic method did not provide. Of course, if you choose a poor model, you will get poor forecasts and poor assessments of uncertainty; maybe not always, but on average.

In this chapter we focus primarily on methods. These methods will be underpinned by the statistical models we will introduce in Chapter 6. Based on these models, we are able to consider prediction intervals, which we do in section 4.9.

4.2 Extrapolation Methods

The fundamental concept underlying forecasting with a single series is that we examine past data and then map out the likely future path of the series based upon the patterns we see in the historical record. The overall structure is shown in Figure 4.1.

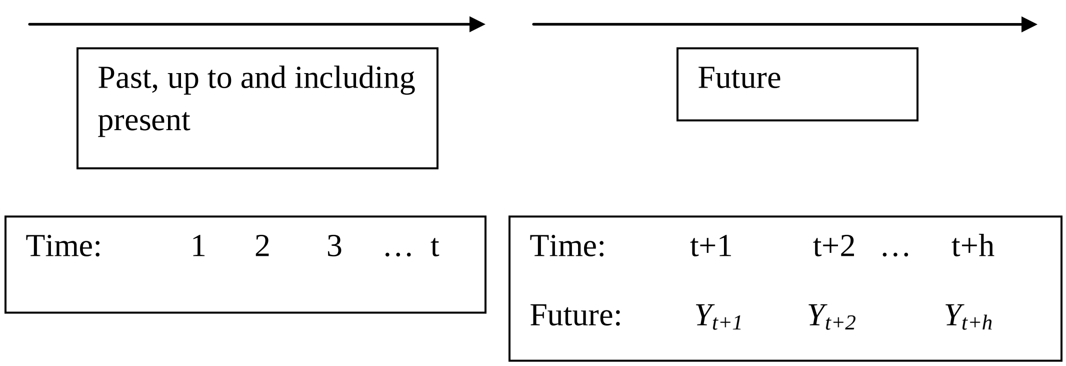


Figure 4.1: General framework for forecasting with a single series

Figure 4.1 is to be interpreted as follows. We denote the particular series of interest (e.g. weekly sales) by Y. We have recorded the observations $Y_1, Y_2, \ldots, Y_t$ over *t* time periods, which represent all the data currently available. Our interest lies in forecasting sales over the next *h* weeks, known as the *forecasting horizon*; that is, we are interested in providing forecasts for future sales, denoted by $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+h}$. Note that although we use the same notation to describe past, present and future, there is a key difference. The past and present values are already observed, whereas the future *Y*'s represent random variables; that is, we cannot write down their values but we can describe them in terms of a set of possible values and the associated probabilities. This concept is illustrated in Figure 4.2, which shows a time series observed for periods 1 to 12, but to be forecast for periods 13 to 20, with increasing uncertainty in the forecast as the horizon increases.

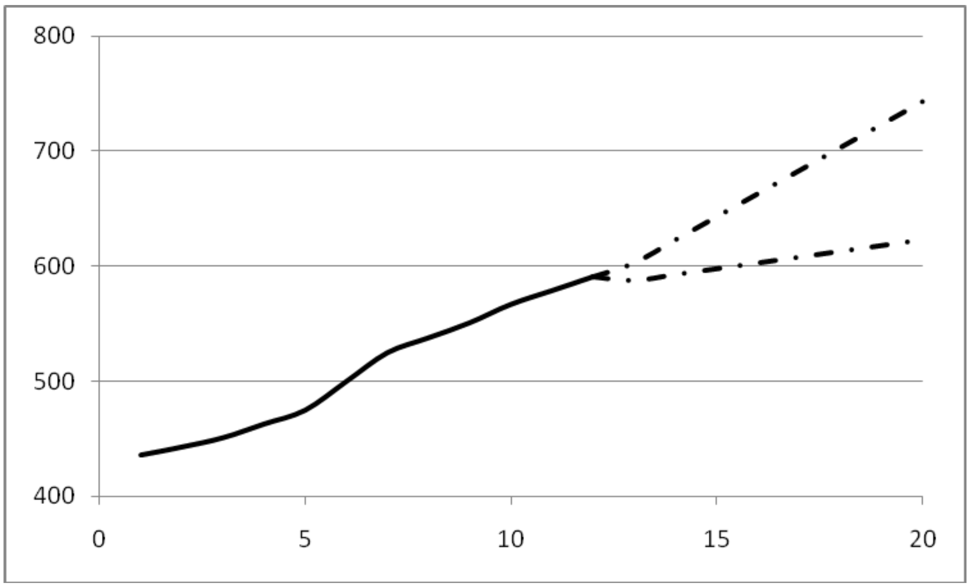


Figure 4.2: Actual sales for periods 1-12 and a range of possible values for periods 13-20

The (point) forecasts for future sales are all made at time *t*, known as the *forecast origin*, so the first forecast (for period 13) will be made one step ahead, the second (for period 14) two steps ahead, and so on. When the results for period 13 are known, we can compute a new forecast for period 14, which will now be only one step ahead. We need to distinguish these different forecasts, as we now illustrate in Table 4.1. Suppose that the initial forecast origin is week 12 and we wish to make forecasts in week 12 for weeks 13, 14 and 15. Then, in week 13, we would make a new set of forecasts for weeks 14 and 15; finally in week 14 we would make a forecast for week 15. A quick look at the table indicates that we are considering six forecasts, all of which are either based upon different information or relate to different time periods. Thus the notation $F_{14|12}$ refers to the forecast for Y_{14} made two weeks earlier in week 12. The subscripts always tell us which time period is being forecast and when the forecast was made. When no ambiguity arises, we will use F_{t+1} to represent the one-step-ahead forecast $F_{t+1|t}$ so that $F_{13} = F_{13|12}$ and so on.

Forecast origin	12	13	14
Forecast for period 13	$F_{13\|12}$	--	
Forecast for period 14	$F_{14\|12}$	$F_{14\|13}$	
Forecast for period 15	$F_{15\|12}$	$F_{15\|13}$	$F_{15\|14}$

Table 4.1: Notation for forecasts made at different forecast origins and for varying steps ahead

This notation may seem a bit elaborate, but it is important to know both the forecast origin and for how many periods ahead the forecast is being made. That is, the term "forecast for period 15" is ambiguous until we know when the forecast was made and it would be impossible to evaluate the forecast's accuracy.

4.2.1 Extrapolation of the mean value

Figure 4.3 [previously shown as Figure 2.1] shows the weekly sales figures for 62 weeks of a product line for a major US manufacturer, which we continue to refer to as WFJ Sales. The series starts at the beginning of the year, picks up after 12 weeks or so, and then stabilizes until a surge in the last few weeks of the year, before dropping back at the beginning of the next year (but at a higher level than a year earlier).

Suppose we were back at mid-year (week 26) and wished to forecast sales for the next few weeks. A naïve approach would be to take the average of the 26 weeks to date, which we write as:

$$\bar{Y}(26) = \sum_{i=1}^{26} Y_i / 26 = 30102\,. \tag{4.4}$$

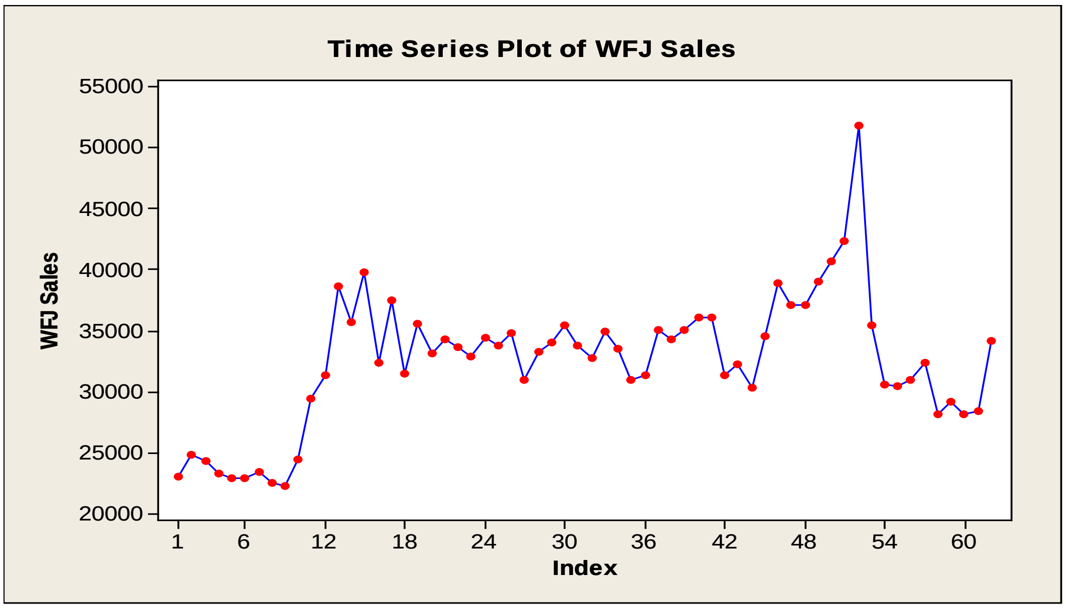


Figure 4.3: Weekly sales for WFJ [WFJ Sales.xlsx] compared to the average

The bar denotes the operation of taking the mean over the 26 observations. Inspection of Figure 4.4, which shows the first 26 weeks of the series and the mean level, suggests that such a value would be too low a forecast. Somehow, we need to give less weight to the first part of the series and focus upon the more recent values.

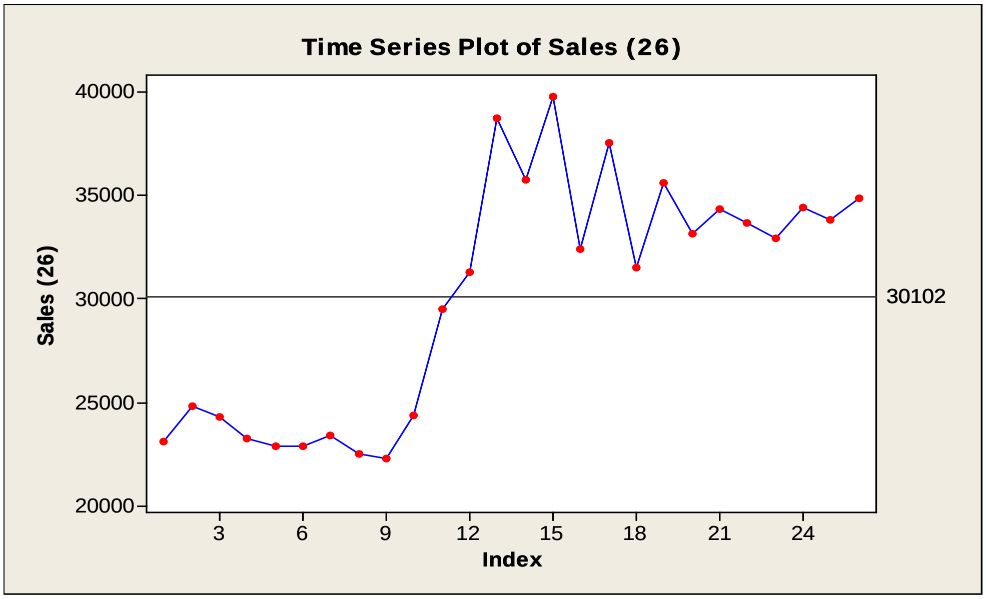


Figure 4.4: Weekly sales for WFJ for the first 26 weeks

4.2.2 Use of moving averages

One way to proceed is to use an average of the last few values in the series. In general, we may use the last K terms of the series and update each time to include only those K values. We refer to the result as a moving average, which contains K terms, the most recent of which is observation *n*:

> *The Moving Average of order K evaluated at time t is denoted by MA(n|K):*
>
> $$MA(t\,|\,K) = \frac{Y_t + Y_{t-1} + \ldots + Y_{t-K+1}}{K} \qquad (4.5)$$

Following from our previous discussion, suppose we decide to use a 3-week moving average, or K = 3. The first such average would be $MA(3,3) = [23056\ +\ 24817 + 24300]/3\ =\ 24058$, as shown in Table 4.2. When a new observation becomes available, the new moving average is

$MA(4,3) = \left[24817 + 24300 + 23242\right]/3 \; = \; 24120$, and so on. Each time, we drop the oldest observation and enter the new one, so that a slightly quicker way of doing the calculation is to write

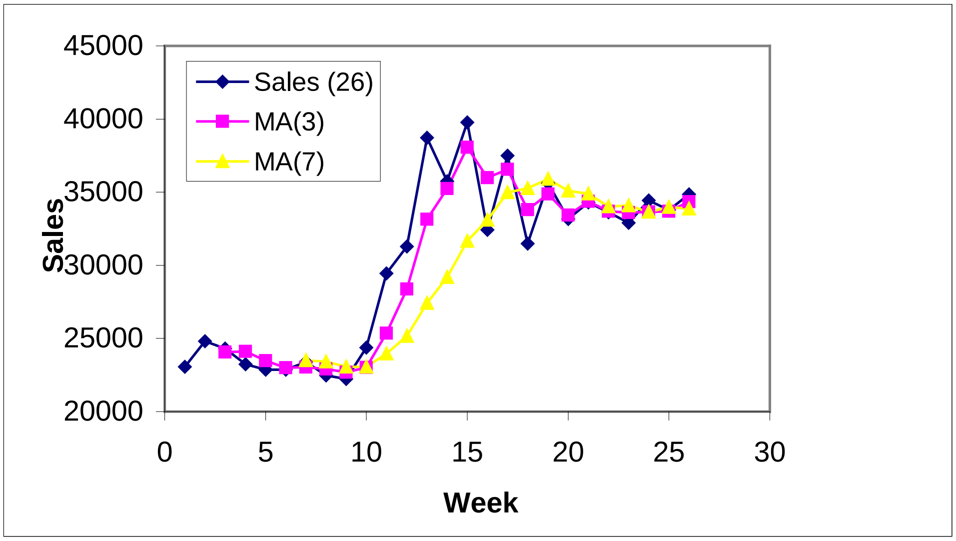
$$\begin{aligned} New\ average = MA(t+1 \mid K) &= old\ \ average + \frac{[new\ value - oldest\ \ value]}{K} \\ &= MA(t \mid K) + \frac{Y_{t+1} - Y_{t+1-K}}{K} \end{aligned} \tag{4.6}$$

A variation of this updating mechanism will feature prominently in later developments. How should we choose K? A sample comparison provides an insight. Table 4.2 summarizes the calculations for K = 3 and for K = 7, which we refer to as MA(3) and MA(7) and the two averages and the original series are plotted in Figure 4.5. The first average that we can calculate is for week 4 which uses the first three observations. This corresponds to the one-step-ahead forecast for week 4 made in week 3.

From Figure 4.5, we see that the 3-term moving average adapts more quickly to movements in the series, but the 7-term average produces a greater degree of smoothing. In order to decide which method is preferable, we need to evaluate suitable measures of forecast performance, as discussed in Chapter 2. However, before we do so, we introduce another form of adaptive average.

Week	WFJ Sales	MA(3)	MA(7)
1	23,056		
2	24,817		
3	24,300		
4	23,242	24058	
5	22,862	24120	
6	22,863	23468	
7	23,391	22989	
8	22,469	23039	23504
9	22,241	22908	23421
10	24,367	22701	23053
11	29,457	23026	23062
12	31,294	25355	23950
13	38,713	28372	25154
14	35,749	33155	27419
15	39,768	35252	29184
16	32,419	38077	31656
17	37,503	35979	33110
18	31,474	36563	34986
19	35,625	33799	35274
20	33,159	34867	35893
21	34,306	33419	35100
22	33,631	34363	34893
23	32,900	33698	34017
24	34,426	33612	34085
25	33,777	33652	33646
26	34,849	33701	33975

Table 4.2: Calculation of moving averages for WFJ Sales [WFJ Sales_MA]



.

Figure 4.5: WFJ Sales for first 26 weeks, with moving averages of lengths 3 and 7

4.3 Simple Exponential Smoothing

The simple moving average introduced in the previous section suffers from two drawbacks. First, the averaging process seems rather capricious in that an observation is given full weight one period, and none the next, when it reaches the K^{th} or "oldest" position. Second, if we use a large number of terms, we have to keep them all around until they are finally removed from the average. This second objection is now of minor importance in practice, although it used to be critical when computer storage was much more limited, or when forecasts had to be updated by hand. Ultimately, the method will be judged by its performance in forecasting, although it would be nice to have a technique that adjusted more smoothly over time. Exactly such a method was introduced by Robert G. Brown (often referred to as the father of exponential smoothing), whose 1959 and 1963 books on forecasting are justly recognized as classics.

To motivate the discussion, we may note that if $\bar{Y}(n)$ denotes the arithmetic mean of the past n observations, the updated mean for *(n+1)* observations, $\bar{Y}(n+1)$ may be computed as:

$$\begin{aligned}\bar{Y}(n+1) &= \frac{Y_1 + Y_2 + \ldots + Y_n + Y_{n+1}}{n+1} \\ &= \bar{Y}(n) + \frac{Y_{n+1} - \bar{Y}(n)}{n+1}\end{aligned} \qquad (4.7)$$

A verbal description of expression (4.7) is:

$$\text{New mean} = \text{old mean} + \frac{\text{[difference between new observation and old mean]}}{\text{n+1}}.$$

A feature of (4.7) is that we need only record the previous mean and the latest value, a useful feature when updating anything from sales reports to batting averages. However, as the series length increases the forecasts generated by the latest mean, become increasingly unresponsive to fluctuations in recent values, since each observation has weight *1/(sample size)*. The update of the simple MA, given in (4.6) avoided this problem and maintained a constant coefficient (1/K), but at the cost of dumping the oldest observation completely. These two conflicting elements may be resolved by using an updating relationship of the form:

$$\bar{Y}(n+1,\alpha) = \bar{Y}(n,\alpha) + \alpha[Y_{n+1} - \bar{Y}(n,\alpha)] \qquad (4.8)$$

That is:

New mean = old mean + α*[difference between new observation and old mean].

> The Basic Equation of Exponential Smoothing:
>
> $$\bar{Y}(n+1,\alpha) = \bar{Y}(n,\alpha) + \alpha[Y_{n+1} - \bar{Y}(n,\alpha)]$$

The process involves comparing the latest observation with the previous weighted average and making a proportional adjustment, governed by the coefficient α, known as the *smoothing constant*. By convention we constrain the coefficient to the range $0 < \alpha < 1$, so that only a part of the difference between the old mean and the new observation is used in the updating. Inspection of (4.8) indicates that in contrast to (4.7), this form of average provides constant weight (α) to the latest observation. Further, the updates require only the latest observation and the previous mean. The average in (4.8) is known as an *exponentially weighted moving average [EWMA]* and we now explain the origin of that name. If we start with the expression for time *(n+1)* and substitute into it the comparable expression for time *n*, we obtain:

$$\begin{aligned}\bar{Y}(n+1,\alpha) &= (1-\alpha)\bar{Y}(n,\alpha)+\alpha Y_{n+1}\\ &= (1-\alpha)[(1-\alpha)\bar{Y}(n-1,\alpha)+\alpha Y_n]+\alpha Y_{n+1}\\ &= (1-\alpha)^2\bar{Y}(n-1,\alpha)+\alpha[(1-\alpha)Y_n+Y_{n+1}]\end{aligned}$$

Continuing to substitute the earlier means, we eventually arrive back at the start of the series with the expression:

$$\bar{Y}(n+1,\alpha)=(1-\alpha)^{n+1}\bar{Y}(0,\alpha)+\alpha[Y_{n+1}+(1-\alpha)Y_n+(1-\alpha)^2Y_{n-1}+\ldots+(1-\alpha)^nY_1] \quad (4.9)$$

The right hand side contains a weighted average of the observations and the weights: $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \ldots$ decay steadily over time. If the weights are plotted against time and a smooth curve is drawn through the values, that curve is exponential; hence the name exponentially weighted moving average. The decay is slower for small values of α, so we can control the rate of decay by choosing α appropriately. Figure 4.6 illustrates the decay rates for α = 0.5 and α = 0.2. At α = 0.5, over 99% of the weight falls on the first seven observations whereas the comparable figure for α = 0.2 is only 79%. As α gets smaller, so does this percentage. Thus, just as we could choose K to control the rate of adjustment of the simple moving average, we may select α to achieve similar adjustments for the EWMA.

Before we proceed further, a word of caution is in order. Equation (4.9) depends on a starting value $\bar{Y}(0,\alpha)$. When n and α are both small, the weight attached to the starting value may be high, as seen in the table below. In these circumstances, it is recommended that more sophisticated software is used that enables an optimal choice of starting level.

α	n	Weight on start value
0.50	10	0.001
	30	0.000
0.20	10	0.107
	30	0.001
0.05	10	0.599
	30	0.215

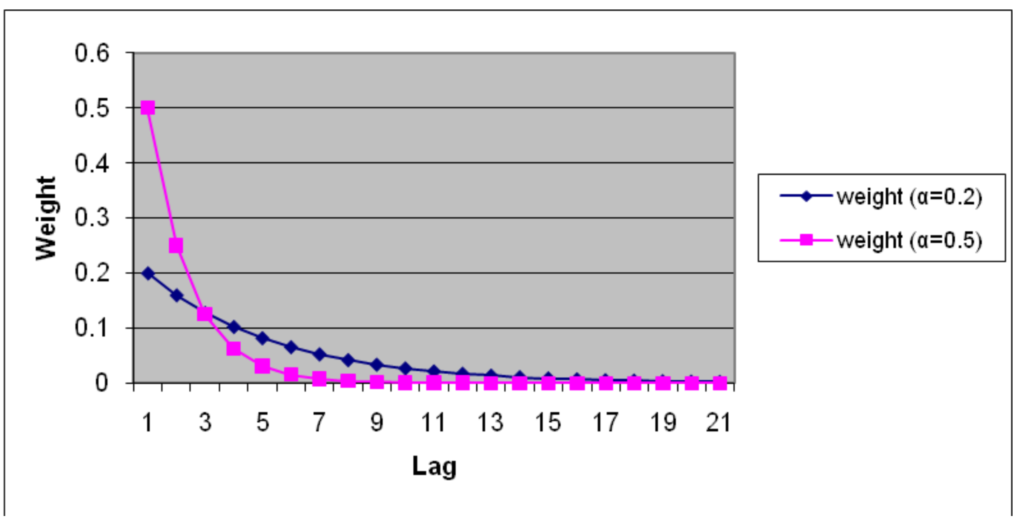


Figure 4.6: Weights on different lags in an exponentially weighted moving average

4.3.1 Forecasting using the EWMA, or Simple Exponential Smoothing

The next step we must take is to convert these averages into forecasts. When we use the EWMA for forecasting, we refer to the method as *Simple (or Single) Exponential Smoothing (SES)*[1]. The underlying logic of this process is that although we believe the process will fluctuate, we have no evidence to suggest that it is likely to go up, down or stay the same. In such circumstances, the average level for future observations is best estimated by our current weighted average. Thus, forecasts made at time *t* for all future time periods will be the same; that is:

$$F_{t+1|t} = F_{t+2|t} = \ldots = F_{t+h|t} = \bar{Y}(t,\alpha) \qquad (4.10)$$

We must recognize that as the lead time increases the forecasts will usually be progressively less accurate. We note that forecasts based upon the mean or the simple moving average also have the same property of being equal for all future periods. Of course, once the next observation is made, the common forecast value changes. Expression (4.8) may be written in the terms of the forecasts as:

$$F_{t+1|t} = F_{t|t-1} + \alpha(Y_t - F_{t|t-1}) \qquad (4.11)$$

That is, the new one-step-ahead forecast is the previous forecast, partially adjusted by the amount that forecast was in error. Since this expression considers only the one-step-ahead forecasts, it may also be written as:

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \qquad (4.12)$$

To set up these calculations, we need to specify the value for α and a starting value, F_1. We have already discussed the effects of different α values. Earlier literature recommended a choice in

[1] The name SES is perhaps unfortunate as the term “smoothing” tends to be overused in time series analysis, but it is in common usage so we will abide by convention.

the range 0.1 < α < 0.3 to allow the EWMA to change relatively slowly and such values often work well for series such as sales figures. However, do not rely on an arbitrary pre-set smoothing parameter. Most computer programs now provide efficient estimates of the smoothing constant, based upon minimizing the mean squared error for the one-step-ahead forecasts:

$$MSE = \frac{\sum_{i=1}^{n}(Y_i - F_i)^2}{n} \tag{4.13}$$

When forecasting with exponential smoothing, use ‘optimum’ Smoothing parameters, not pre-set values.

Example 4.2: Basic SES calculations

A short series of sales data (hypothetical) is given in Table 4.3. The first observation is used as the forecast for period 2 and the smoothing constant is set at $\alpha = 0.3$. Thus, from (4.11) we have for t=2:

$$\begin{aligned} F_{t+1|t} &= F_{t|t-1} + \alpha(Y_t - F_{t|t-1}) \\ &= 5.00 + 0.3(6 - 5.00) = 5.30 \end{aligned}$$

The forecasts are computed successively in the same way. The forecast errors, their squares, absolute values and absolute values as a percentage of the observed series are shown in succeeding columns. The last row of the table gives the mean error, MSE, MAE and MAPE. The calculations may be checked using *SES.xlsm* as described in Appendix 4A.

Time	Sales	Forecast	Error	E^2	Abs E	APE
1	5					
2	6	5.00	1.00	1.00	1.00	16.67
3	7	5.30	1.70	2.89	1.70	24.29
4	8	5.81	2.19	4.80	2.19	27.38
5	7	6.47	0.53	0.28	0.53	7.61
6	6	6.63	-0.63	0.39	0.63	10.45
7	5	6.44	-1.44	2.07	1.44	28.78
8	6	6.01	-0.01	0.00	0.01	0.12
9	7	6.01	0.99	0.99	0.99	14.21
10	8	6.30	1.70	2.88	1.70	21.21
11	7	6.81	0.19	0.04	0.19	2.68
12	6	6.87	-0.87	0.75	0.87	14.48
13		6.61				
		Means	0.44	1.51	1.02	15.12

Table 4.3: Illustration of spreadsheet calculations for SES

4.3.2 An Excel macro for SES

Appendix 4A describes an Excel macro using Solver to select the best value for α. The macro allows the user to partition the data series into estimation and hold-out samples and computes summary statistics for each sub-sample. The mean absolute error (MAE) or the mean absolute percentage error (MAPE) could be used in place of the MSE. The estimated value of α is typically similar in all three cases, unless we are dealing with a very short series or one that contains some severe outliers.

We must still resolve the choice of starting values. Two principal options are commonly used: to use the first observation as F_1, or to use an average of a number of observations; recommendations vary from the first three or four up to six or twelve or even the mean of the whole sample. Gardner (1985) provides a good review of the options. With such variations, it is evident that different software programs will produce different forecasts; further, some packages are very coy about the method they employ. The macro described in the appendix at the end of the chapter allows the choice of the first observation only, or an average of the first four observations. When either n or α is large, the choice of starting value is relatively unimportant and the different approaches yield very similar results. When both n and α are small, more sophisticated methods should be used[2].

We now explore the effects of changing α for forecasts of WFJ Sales. The macro was used in each case to generate the results.

Example 4.3: SES forecasts for WFJ Sales [WFJ Sales.xlsx]

Table 4.4 shows the first ten one-step-ahead forecasts for WFJ sales, using SES with various α = 0.2, 0.5 and finally 0.728, the value that minimizes the one-step-ahead MSE over observations 6 through 26. Thus, the first out-of-sample forecast is available at period 27 and then the one-step-ahead forecasts are listed for weeks 27 – 36. The starting value was based upon an average of the first four observations in each case. The three sets of forecasts correspond to:

- SES(0.2) – Simple exponential smoothing with α = 0.2
- SES(0.5) – Simple exponential smoothing with α = 0.5
- SES(opt) – Simple exponential smoothing with α = 0.728

Period	WFJ Sales	SES (0.2)	SES (0.5)	SES (opt)
27	30986	33884	33842	33858
28	33321	33304	34346	34580
29	34003	33308	32666	31963
30	35417	33447	32993	32952

[2] In section 6.x we discuss fitting a complete model to estimate the both the smoothing parameter and the starting value.

31	33822	33841	33498	33717
32	32723	33837	34458	34955
33	34925	33614	34140	34130
34	33460	33876	33432	33105
35	30999	33793	34178	34431
36	31286	33234	33819	33723

Table 4.4: Actual and forecast WFJ sales, using SES.xlsm [see WFJ Sales_SES.xlsx]

To illustrate the nature of the calculations, consider time period 28 for α = 0.2. From (4.11) we have:

For t=28

$$F_{t+1|t} = F_{t|t-1} + \alpha(Y_t - F_{t|t-1})$$

or

$$F_{28|27} = 33884 + 0.2(30986 - 33884)$$

$$= 33304$$

Similarly for period 29, $F_{29|28} = 33308$

Figure 4.7 shows the one-step-ahead forecasts for periods 27 – 36.

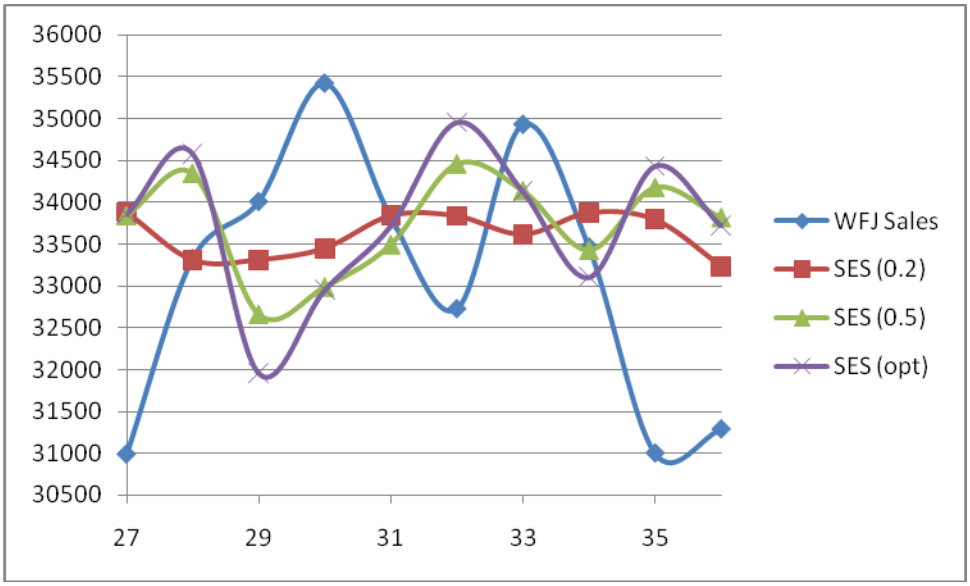


Figure 4.7: SES forecasts for WFJ sales

4.3.3 The use of hold-out samples

The reader will observe that in Example 4.3 we partitioned the original series into two parts. We now explain the motivation for that decision, which provides the basis for an examination of the relative performance of these methods. For this purpose, we use the accuracy measures developed in section 2.7.

When we split the series into two parts we refer to the first part as the *estimation* sample, which is used to estimate the starting values and the smoothing parameters. This sample often contains 75 – 80 percent of the observations, although the length of the series clearly affects this choice. The parameters are commonly estimated by minimizing the mean squared error (MSE), although the mean absolute error (MAE) or mean absolute percentage error (MAPE) are also used.

The *hold-out* sample represents the remaining 20 -25 percent of the observations and is used to check forecasting performance. No matter how many parameters are estimated using the estimation sample, each method under consideration can be evaluated using the "new" observations contained in the hold-out sample. Thus, the hold-out sample provides a level playing field for such comparisons.

Some programs allow repeated estimation and forecast error evaluation by advancing the estimation sample one observation at a time and repeating the error calculations. For example, if the hold-out sample comprised 12 observations, we would have an initial set of forecast errors at lags 1, 2, through 12. We then advance one period and generate forecasts 1, 2, through 11 steps ahead using the last 11 values. This operations is repeated using the last 10, 9, …, 1 observations to produce 12 one-step-ahead forecasts, 11 two-step-ahead forecasts and so on, thereby providing a more complete assessment of forecast performance. This use of a rolling horizon provides a more reliable assessment of performance (Fildes, 1992).

Example 4.4: Comparison of one-step-ahead forecasts

In order to examine the forecasting performance of the various methods we have discussed to date, we carried out the following experiment using the series WFJ Sales. Following recommended practice, we use an out-of-sample evaluation. That is, we fit the data using the first part of the series and then examine the performance of the forecasting method by seeing how well it works on later observations. We compared five methods:

- MA(3) – a moving average of three terms
- MA(8) – a moving average of eight terms
- SES(0.2) – simple exponential smoothing with $\alpha = 0.2$
- SES(0.5) – simple exponential smoothing with $\alpha = 0.5$
- SES(opt) – simple exponential smoothing with $\alpha = 0.729$

For each of the SES sets of forecasts, we used the first four terms to generate the starting value, as in the previous example. We then generated the one-step-ahead forecasts for 36 time periods (i.e. periods 27-62) starting at forecast origin t = 26. The summary results appear in the following table. The best performance on each criterion is shaded. In this case, the best fitting SES scheme performs best on all counts. However, it may well happen that the different criteria lead to different conclusions.

Method	MAE	RMSE	MAPE
MA(3)	3067	4320	8.9
MA(8)	3749	4865	11.0
SES(0.2)	3389	4342	9.9
SES(0.5)	2832	3980	8.2
SES(opt)	2561	3915	7.3

Table 4.5: Summary error measures for WFJ sales data

4.3.4 Some general comments

On average, SES tends to outperform MA, as observed in an empirical comparison of their performance in the M3 forecasting competition (as reported by Makridakis and Hibon (2000) and in various other studies). In addition, as we shall see in Chapter 6, SES corresponds to an intuitively appealing underlying model, whereas MA does not. Given the inferior properties and overall performance of moving average-based procedures, we do not recommend their direct use for forecasting. However, moving averages have other uses as we shall see in sections 5.5 – 5.7.

One final question refers to the choice of fitting procedure. We could minimize the MSE (or equivalently the RMSE) or instead choose to minimize the MAPE or the MAE. In order to examine the effects of such a choice we again examine the WFJ series. The data through period 26 were used for estimation of the smoothing constant by each method in turn. Different options for the starting values (first observation and first four observations) for the level produced essentially identical forecasts for periods 27-62 and only the results for the start value based upon the average of the first four observations are given in Table 4.6.

In this example, the choice of fitting method produces only marginal differences in the results[3], which tends to happen provided that adequate data are available for estimation. Solver produced identical values for the smoothing constant using MAE and MAPE; these two criteria produce similar results whenever the typical percentage error is small. Finally, we note that the out-of-sample error measures tend to be somewhat higher than those calculated for the estimation sample. This should not come as a surprise, since the smoothing constant was chosen to minimize the appropriate criterion within the estimation sample.

Estimation Criterion	RMSE		MAE		MAPE	
Sample	Estimation	Hold-out	Estimation	Hold-out	Estimation	Hold-out
Error Measure						
RMSE	3200	3915	3217	3921	3217	3921
MAE	2288	2561	2247	2624	2247	2624
MAPE	7.0	7.3	6.9	7.5	6.9	7.5

[3] The calculations were made with Excel Solver, using the spreadsheet described in Appendix 4A.

Value of α	0.729	0.660	0.660

Table 4.6: Effects of fitting by minimization of RMSE, MAE or MAPE

4.4 Linear Exponential Smoothing

Table 4.7 shows the quarterly revenues (sales) for Netflix. For such rapidly increasing series it is natural to think about incorporating a trend line into our forecasting method. Figure 4.8 shows two such possibilities: plot (a) shows a linear trend and plot (b) shows a quadratic trend.[4]

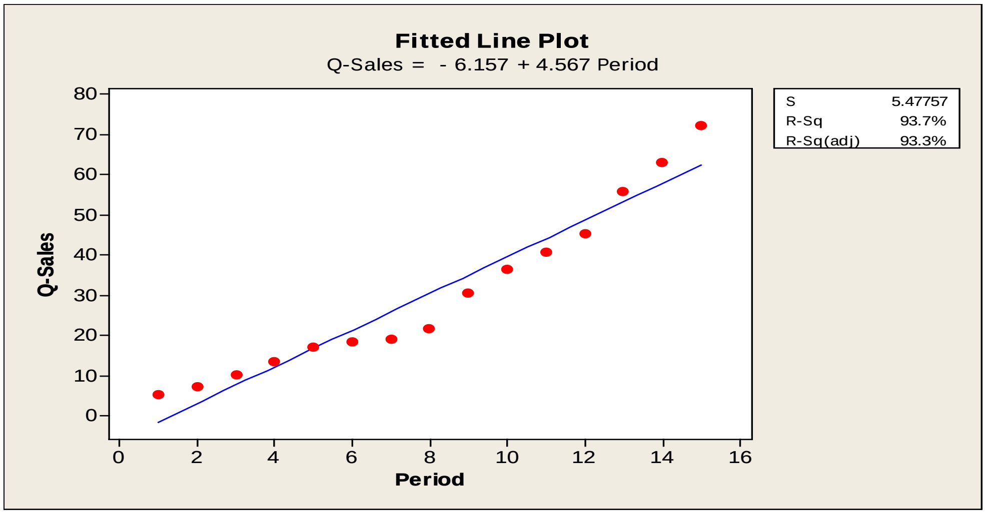
Year	Quarter	Quarterly Sales	Year	Quarter	Quarterly Sales
2000	1	5.17	2002	1	30.53
2000	2	7.15	2002	2	36.36
2000	3	10.18	2002	3	40.73
2000	4	13.39	2002	4	45.19
2001	1	17.06	2003	1	55.67
2001	2	18.36	2003	2	63.19
2001	3	18.88	2003	3	72.20
2001	4	21.62	2003	4	81.19

Table 4.7: *Quarterly revenue (Sales) for Netflix,2001:Q1 to 2003Q4 [measured in $million]*

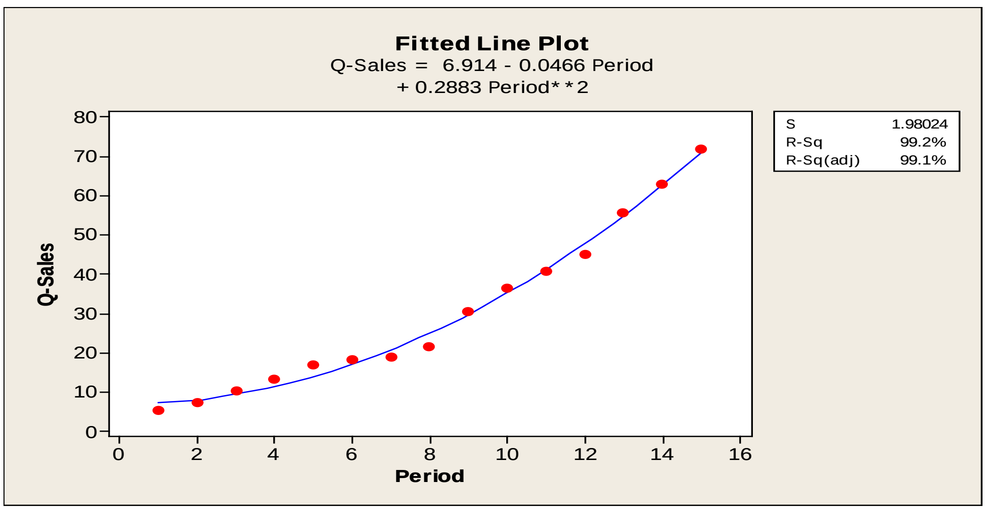
Figure 4.8: Linear and quadratic trend lines for Netflix sales

(a) Linear trend

[4] The calculations involved in creating these trend lines do not concern us here. A detailed explanation is given in section 6.x.



(b) Quadratic trend



The quadratic seems to provide an excellent fit, but we must be very wary of such "global" models; that is, models that assume a never-changing trend. Indeed, the road to Chapter 11 bankruptcy proceedings is littered with the remains of companies that believed such optimistic growth patterns would persist. A successful start-up company will often show dramatic growth in the early years, but this will moderate as the company and its markets mature; see the end-of-chapter Mini-case for details. The Netflix record is very impressive, but investors should not get too carried away!

4.4.1 Basic structure for LES

Global trend models assume the trend remains constant over the whole time series. Since we are reluctant to rely upon the continuation of global patterns, we now develop tools that project trends more locally. To understand the approach, we begin with the equation for a straight line, and ignore any statistical variation. Let:

Y_t = value of series at time t

L_t = level of series at time t

B_t = slope of series at time t.

If we start out with a constant slope, $B_t = B$ we can write the straight-line equation as $Y_t = L_0 + Bt$. That is, the value of the series starts out at the value L_0 at time zero and increases by an amount B in each time period. Another way of writing the right-hand side of this expression is to state directly that the new level, L_t is obtained from the previous level by adding one unit of slope, B, or:

$$L_t = L_0 + Bt = L_{t-1} + B.$$

That is, we define the level at time t, L_t in terms of its value in the previous time period and the previous slope. We may then define the variable Y_t in terms of the level and the slope as $Y_t = L_{t-1} + B$. Further, we can see that if we go h periods ahead, we can define the variable at time $(t+h)$ in terms of the level at time $t-1$ and the appropriate number of slope increments:

$$Y_{t+h} = L_0 + B(t+h) = L_t + Bh\,.$$

There is a lot of redundancy in these expressions, since we are considering the error-free case. When we turn back to the real problem with random errors and changes over time in the level and the slope, these equations suggest that we consider forecasts of the form:

$$\begin{aligned} F_{t+h|t} &= \text{[level at time t] + h*[slope at time t]} \\ &= L_t + hB_t \end{aligned} \qquad (4.14)$$

Thus, the one-step-ahead forecast made at time *(t-1)* is:

$$F_{t|t-1} = F_t = L_{t-1} + B_{t-1}$$

Recall that for SES we just used the F-notation. This was convenient because we implicitly assumed a zero slope. When the slope is strictly zero we can see that $F_t = L_{t-1}$ so the L-notation was not needed at that time.

4.5.2 Updating relationships

We can now consider updating the level and the slope using equations like those we used for SES in section 4.2. We define the observed error (e_t) as the difference between the newly observed value of the series and its previous one-step-ahead forecast:

$$e_t = Y_t - F_t = Y_t - (L_{t-1} + B_{t-1})\,.$$

Given the latest observation, Y_t we update the expressions for the level and the slope by making partial adjustments that depend upon the error:

$$\begin{aligned} L_t &= L_{t-1} + B_{t-1} + \alpha e_t \\ B_t &= B_{t-1} + \alpha\beta e_t \end{aligned} \qquad (4.15)$$

These expressions may be explained as follows:

- The new level is the old level (adjusted for the increase produced by the slope) plus a partial adjustment (weight α) for the error.
- The new slope is the old slope plus a partial adjustment (weight αβ) for the error.

The expressions in equation (4.15) are known as the *error correction form* of the updating equations. As may be checked by substitution, the slope update can also be expressed as:

$$B_t = B_{t-1} + \beta(L_t - L_{t-1} - B_{t-1}) \tag{4.16}$$

The term, L_t-L_{t-1} represents the latest estimate of the slope and therefore $(L_t - L_{t-1} - B_{t-1})$ is the latest error made if the smoothed slope is used as an estimate instead. In this form, it is apparent that a second round of smoothing is applied to estimate the slope which has led some authors to describe the method as *double exponential smoothing.* Since the forecast function (4.14) defines a straight line, we prefer the name *linear exponential smoothing [LES]*. It is also known as Holt's method after one of its originators. If we set the slope equal to zero at all times, we are back at simple exponential smoothing; the second equation in (4.15) would disappear, and the first one reduces to (4.11) on identifying L_t as F_{t+1}.

4.4.3 Starting values

To set this forecasting procedure in motion, we need starting values for the level and slope, and values for the two smoothing constants, α and β. The smoothing constants may be specified by the user, and conventional wisdom decrees using $0.05 < \alpha < 0.3$ and $0.05 < \beta < 0.15$. These guidelines are not always appropriate and it is better to view them as suggesting initial values in a procedure to select optimal coefficients by minimizing the MSE over some initial sample, as we did for SES. As before, the performance of the resulting forecasting equations should be checked out-of-sample as we did in Example 4.4.

As for SES, different programs use a variety of procedures to set starting values (Gardner, 1985).The macro *LES.xlsm* described in Appendix 4B[5] uses the starting values:

$$B_3 = \frac{(Y_3 - Y_1)}{2} \text{ for the slope and } L_3 = \frac{(Y_1 + Y_2 + Y_3)}{3} + \frac{(Y_3 - Y_1)}{2} \text{ for the level}$$

These values correspond to fitting a straight line to the first three observations. Once the initial values are set, equations (4.15) are used to update the level and slope as each new observation becomes available.

Example 4.5: Spreadsheet for LES calculations

A short series of sales data (hypothetical) is given in Table 4.8. The smoothing constants are set at $\alpha = 0.3, \beta = 0.1$ and the calculations were performed using the macro. Thus, the first three observations were used to set initial values for the level and slope and the forecast error measures were then computed using observations 5 – 12; these values are genuine forecasts since no estimation is involved. The entries corresponding to time period 13 represent the one-step-ahead forecast and its component parts. The forecast errors, their squares, absolute values and absolute percentage errors (APE) are shown in succeeding columns. The last three rows of the table gives the RMSE, MAE and MAPE, calculated over periods 5 through 12.

[5] It should be noted that different starting values for the parameter estimation may lead to different final values. Try several values to ensure that a global minimum has been reached.

Time	sales	Level	Slope	Forecast	Error	Error sq	|Error|	APE
1	5							
2	7							
3	9							
4	10	9.00	2.00	11.00	-1.00	1.00	1.00	10.00
5	11	10.70	1.97	12.67	-1.67	2.79	1.67	15.18
6	12	12.17	1.92	14.09	-2.09	4.36	2.09	17.41
7	16	13.46	1.86	15.32	0.68	0.46	0.68	4.25
8	17	15.52	1.88	17.40	-0.40	0.16	0.40	2.36
9	20	17.28	1.87	19.15	0.85	0.73	0.85	4.27
10	17	19.40	1.89	21.29	-4.29	18.44	4.29	25.26
11	21	20.01	1.76	21.77	-0.77	0.59	0.77	3.66
12	22	21.54	1.74	23.28	-1.28	1.63	1.28	5.80
13		22.89	1.70	24.59				
						MSE	MAE	MAPE
						3.35	1.45	9.80

Table 4.8: LES calculations for hypothetical sales data [Hyp sales.xlsx]

Example 4.6: Linear exponential smoothing for WFJ sales.

We now consider the use of LES for the WFJ Sales data (given in Table 4.4). We consider the three fitting criteria in turn: MSE, MAE and MAPE, again using *LES.xlsm*. Table 4.9 shows the results for each criterion. As may be expected, the values of the error measures in the estimation sample are systematically less than those for the hold-out forecasts.

Error Measure	Fitting Criterion					
	MSE		MAE		MAPE	
	Estimation Sample	Hold-Out	Estimation Sample	Hold-Out	Estimation Sample	Hold-Out
RMSE	3129	4033	3147	4079	3129	4034

MAE	2333	2694	2325	2752	2333	2695
MAPE	7.36	7.91	7.38	8.13	7.36	7.92
Value of α	0.70		0.62		0.70	
Value of β	0.00		0.00		0.00	

Table 4.9: Estimation sample (ES) and Hold-out sample (HO) results for WFJ Sales

This table shows a number of interesting features:

- The MSE and MAPE criteria produce almost identical results in this case and the overall differences among all three criteria are slight
- The hold-out sample produces somewhat higher values for the error measures, as may be expected.
- The estimates of β are always zero, suggesting a constant slope.

Too much should not be made of a single example, but three important conclusions can be drawn for general application:

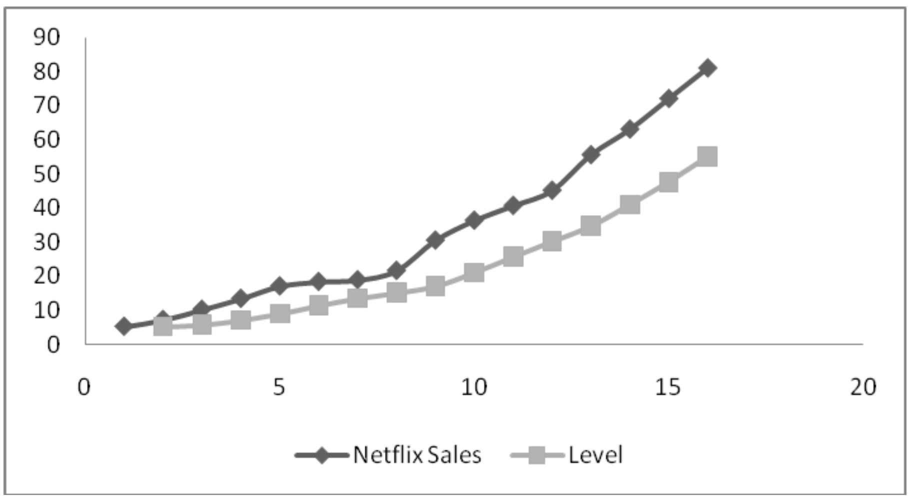
1. Retain an adequate number of observations for out-of-sample testing. If you have sufficient data, retain at least 10-12 observations for this purpose.
2. Different fitting criteria tend to produce similar results.

Finally, comparison of Tables 4.7 and 4.9 indicates that there is no benefit to using LES in this case; SES does just as well, confirming our initial impression that the series did not show any marked tendency to change over time after the initial period when sales were lower.

Example 4.7: Forecasting Netflix sales using LES

The sales figures for Netflix, shown in Table 4.7, show a very strong trend and we would expect LES to perform much better than SES in this case. The series was fitted using the macro described in Appendix 4B to observations 1 – 16. The error measures are calculated using the fitted values for observations 5 – 16 in each case; a hold-out sample was not used as we are interested in the quality of the fit in this example, rather than forecast performance. From Figure 4.9A we see that the SES forecast always undershoots the next value of the series because it fails to allow for the upward trend.

(A) Simple Exponential Smoothing (SES)



(B) Linear Exponential Smoothing (LES)

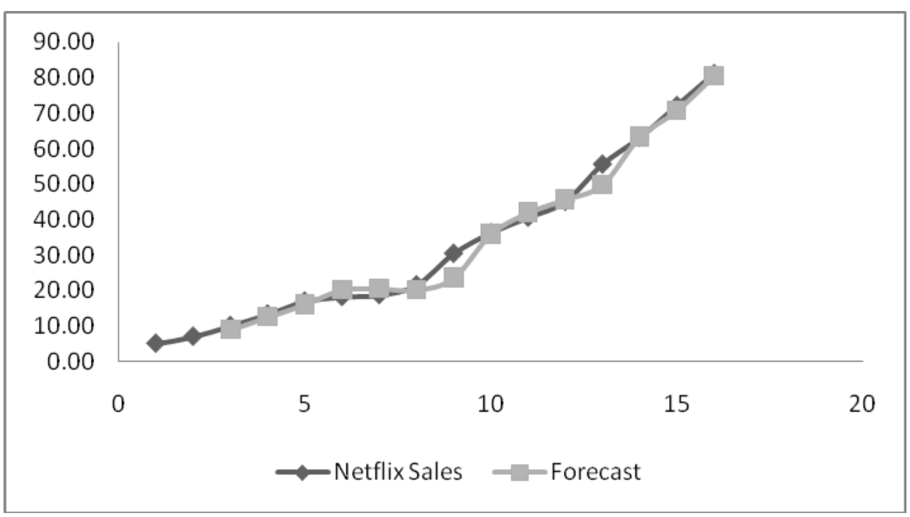


Figure 4.9: Netflix sales and one-step-ahead forecasts from Single Exponential Smoothing and Linear Exponential Smoothing (for observations 1 – 16)

Method (fitting criterion)	SES (All)	LES (RMSE(	LES (MAE/MAPE)
RMSE	6.29	2.80	2.80
MAE	5.46	1.90	1.90
MAPE	14.47	5.98	5.97
alpha	1.00	1.00	1.00
beta		0.49	0.55

Table 4.10: Summary measures for SES and LES for Netflix Sales.

The results are summarized in Table 4.10. LES is clearly much superior to SES, as we would expect. The choice of fitting method makes very little difference in this case. However, we observe that α is set at its upper level of 1.0, indicating that even LES may have problems. Again, if we step back from the technical details and look at the plot of the data in Figure 4.8 we see that the growth is exponential rather than linear. . In the next section we examine the use of transformations as a way of dealing with this question.

4.5 The Use of Transformations

The LES method requires that the series is locally linear. Intuitively, if the trend for the last few time periods in the series appears to be close to a straight line, the method should work well. However, in many cases this assumption is not realistic. For example, suppose we are trying to forecast the GDP of a country, which has experienced growth of around 5 percent per year in recent years. Ignoring statistical variation for the moment, such growth could be described by the expression:

$$Y_t = (1.05)^t Y_0 \text{ or } \mathrm{Y}_t = 1.05Y_{t-1}$$

When plotted, this function is an exponential curve, which increases at an increasing rate. Any linear approximation will undershoot the true function sooner or later (an observation first made hundred of years ago by Malthus regarding linear growth in food supply and exponential population growth leading to starvation). For very short-term forecasts this bias may not matter, but it may become serious as the forecasting horizon increases.

4.5.1 The log transform

Referring back to section 2.6, we identified two approaches to the problem. The first is to use a logarithmic transformation. As noted earlier, we usually use natural logs to the base e = 2.71828…; any software package will have a function such as *ln* or *loge* to compute the logarithm and an inverse function (or reverse transformation) called *exp* or *expo* to convert the logarithm back to the original values. Thus, for exponential growth, the log transform yields:

$$\ln Y_t = \ln(1.05) + \ln Y_{t-1}$$

This expression may also be written as:

$$\ln(Y_t / Y_{t-1}) = \ln(1.05)$$

If we now write $Z_t = \ln Y_t$ the reverse transformation is:

$$\begin{aligned} Y_t &= \exp(Z_t) = \exp[\ln(1.05) + \ln Y_{t-1}] \\ &= 1.05 Y_{t-1} \end{aligned}$$

Thus, the log-transform produces a linear trend to which we can apply LES. We must then transform back to the original series to obtain the forecasts of interest.

The net effect of the process would typically improve forecasting performance.

Example 4.8: Forecasting Netflix using the log- transform

We applied LES to the transformed series, using the same procedures for setting starting values and estimating the parameters as before. The summary statistics are given in Table 4.11. The table gives the results of estimating the parameters by minimizing the MSE of (a) the transformed values and (b) the original values. The summary values are computed after transforming back to the original units, otherwise comparisons between methods are not feasible.

The comparative results in Table 4.11 indicate that the log transform does somewhat worse in this case. However, we must stress that these comparisons are based upon a small sample for a single series and no general conclusions should be drawn. The only solid conclusion that can be drawn is the intuitively obvious one that SES should not be used for strongly trending series; whether to use LES on the original or transformed series, or to use SES on growth rates, remains a question for further examination in any particular study.

In Exercise 4.15 it is shown that this expression is close to the growth rate that we discussed in section 4.4.

	Fitted using MSE of log series	Fitted using MSE of original series
RMSE	3.66	2.80
MAE	2.93	1.90
MAPE	8.57	5.98
Value of α	1.00	1.00
Value of β	0.97	0.49

Table 4.11: Summary measures for LES for Netflix Sales, using a log transform. The error measures refer to the original units of the observations [Netflix.xlsx]

4.5.2 Use of growth rates

So far, we have assumed that the data series has not exhibited a strong trend. However, many products have a sales history of growth, possibly later followed by a decline as they are superseded by new (and hopefully improved) items. Similarly, many macroeconomic series, such as consumer expenditure, show a strong tendency to increase over time. A third category of time series relates to the returns on an investment, where the return is defined exactly as in equation (4.17) below. There are two ways to forecast such series:

1. Convert the series to growth over time and forecast the growth rate, then convert back to the original series.
2. Develop forecasting methods that account for trends.

Where the variable to be forecast is the future level of the original data series, approach (2) is more widely applicable and more commonly used. However, before embarking on a detailed analysis of that topic, we will explore the use of growth rates which are often of interest in themselves. These may be forecast using the approaches already discussed.

We begin by defining the growth rate. Assuming that the variable has a natural origin, the growth rate from one time period to the next is:

$$G_t = 100 * \frac{Y_t - Y_{t-1}}{Y_{t-1}}. \tag{4.17}$$

Note that growth, by this definition can be negative. We first compute the single period growth rates and then use SES to predict the growth for the next period, which we denote by $g_{t+1|t} = g_{t+1}$, following our usual convention. The one-step-ahead forecast for the original series is given by:

$$F_{t+1|t} = F_{t+1} = Y_t * \left[1 + \frac{g_{t+1}}{100}\right]. \tag{4.18}$$

That is, we unscramble equation (4.17) to determine the forecast. The forecast updates the current level, Y_t, by the forecast growth rate. Netflix operates a video rental business that has grown very rapidly in recent years. The quarterly sales (in $ million) are shown for 2000, Q1 through 2003, Q4 in Table 4.12. The growth rates for one quarter over the immediately preceding quarter are then computed using equation (4.18). The one-step-ahead forecasts of the growth rate are generated using the first value to define F_1; the optimal value for $\alpha = 0.60$. The one-step-ahead sales forecasts are then given in the last column, generated using (4.18).

The growth rate is highly variable, as is clear from Table 4.12 and the optimal value for α varies considerably with the time span of the data chosen. Short series often lead to unstable estimates. Given the shortness of the series, we selected the optimal value for α by minimizing the MSE over all the data points. The best fitting value does not guarantee the best point forecast on any single occasion, but rather it minimizes the expected forecast MSE (or other fitting criterion) for a series whose generating process is unchanged. The calculation of the various error measures, as in Table 4.5 is left to the reader as Exercise 4.6. Further consideration of Netflix is provided in the end of chapter case study.

The forecasts are not wonderful, but the evaluation of growth rates seems intuitively more reasonable than trying to forecast rapidly growing dollar amounts using SES. In particular, we can look for a slow-down in growth as a company matures. Mini-case 1 at the end of this chapter examines the growth of Netflix using more recent data. We will provide an additional justification for this approach later, in section 6.x.

Year	Quarter	Quarterly Sales	Growth - percent	Growth Forecast	Sales Forecast		
2000	1	5.17					
2000	2	7.15	38.1	38.1			
2000	3	10.18	42.5	38.1	9.9	RMSE =	4.66
2000	4	13.39	31.5	41.9	14.1		
2001	1	17.06	27.4	33.0	19.0	MAE =	3.91
2001	2	18.36	7.6	28.2	22.7		
2001	3	18.88	2.8	10.5	23.5	MAPE =	12.08
2001	4	21.62	14.5	3.9	20.9		
2002	1	30.53	41.2	13.0	22.5	Alpha =	0.86

2002	2	36.36	19.1	37.2	34.5
2002	3	40.73	12.0	21.7	49.9
2002	4	45.19	10.9	13.4	49.6
2003	1	55.67	23.2	11.3	51.2
2003	2	63.19	13.5	21.5	62.0
2003	3	72.20	14.3	14.6	76.8
2003	4	81.19	12.4	14.3	82.8

Table 4.12: Growth rate analysis of Netflix quarterly sales, 2000-2003. [Quarterly revenue (Sales) is measured in $million; growth is measured as a percentage]

4.6 Exponential Smoothing with a Damped Trend

As we saw earlier, the growth rate for Netflix slowed as the company matured. This phenomenon is quite common in time series for sales where a product line matures and sales may then decline unless the product is upgraded in some way. Indeed, such a product life-cycle is a standard expectation in marketing. We can accommodate such effects by modifying the updating equations for the level and slope. We anticipate that, in the absence of random errors, the leel should flatten out unless the process encounters some new stimulus. In turn, this expectation means that the slope should approach zero. We achieve this represnetaion in our model by introducing a dampening factor so that equations (4.15) become

$$\begin{aligned} L_t &= L_{t-1} + \phi B_{t-1} + \alpha e_t \\ B_t &= \phi B_{t-1} + \alpha\beta e_t \end{aligned} \tag{4.19}$$

In equations (4.19) we have inserted a *dampening factor* φ , with a value between zero and one, in front of every occurrence of the slope term B_{t-1}. We select φ to be positive but less than one so that the effect is to shift the slope term towards zero, or to dampen it. If we compute the forecast function for h-steps-ahead, we obtain:

$$F_{t+h|t} = L_t + (\phi + \phi^2 + \ldots + \phi^h)B_t \qquad (4.20)$$

This forecast levels out over time approaching the limiting value $L_t + B_t / (1-\phi)$ provided the dampening factor is less than one. This limiting value contrasts sharply with the case $\phi = 1$ when the forecast keeps increasing so that $F_{t+h|t} = L_t + hB_t$.

The damped trend model was introduced by Gardner and McKenzie (1985) and has proved surprisingly effective (Makridakis and Hibon, 2000).

Example 4.9: Forecasting for Netflix Sales using the damped trend method

We took the dampening factor to be $\phi = 0.95$ and repeated the forecasting exercise using damped LES, with the same fitting procedures are before. The results are summarized in Table 4.14. The forecasts are marginally inferior to those from LES without dampening; indeed the optimal value for $\phi = 1.00$ but if we believe that company expansion is slowing the damped form may be preferable for future use.

Period	Netflix Sales	Forecast		
1	5.17			
2	7.15			
3	10.18			
4	13.39	12.51	RMSE =	2.85
5	17.06	16.36		

6	18.36	20.37	MAE =	1.95
7	18.88	20.42		
8	21.62	19.99	MAPE =	6.04
9	30.53	23.66		
10	36.36	36.56	α =	1.00
11	40.73	42.15	β =	0.59
12	45.19	45.57	ϕ =	0.95
13	55.67	49.70		
14	63.19	63.60		
15	72.20	70.71		
16	81.19	80.45		

Table 4.13: One-step ahead fitted values and forecasts for Netflix Sales using the damped trend method (φ=0.95)

4.6.1 Choice of Method

The different forms of exponential smoothing apply to certain types of time series. Figure ? shows a categorization of series by the form of the trend (no trend, linear trend, damped trend) and seasonality (none, additive, multiplicative). When choosing a method the aim is to match it with the type of data. Often however, the form of series is unknown. By implication in our earlier discussion, we have recommended choosing the method that performs best on the hold-out sample, using one or more of the criteria we have discussed. When the series is too short to allow for a hold-out sample of reasonable size, we use the measures defined for the estimation sample. This approach must be treated with caution and we develop it further in Chapter 6.

4.7* Other Approaches to Trend Forecasting

Before we leave the subject of trend forecasting, several other methods are worthy of a brief mention. For a more comprehensive review of recent developments, see the excellent review article by Gardner (2006).

Brown's method of double exponential smoothing [DES]

As noted earlier, Robert Brown was the original developer of exponential smoothing methods and his books [Brown, 1959, 1963] have become classics. His initial derivation of exponential smoothing used a least squares argument which, for a local linear trend reduces to the use of LES with $\alpha = \beta$. Unless data are very limited there is no particular benefit to imposing this restriction and we will not consider it further. However, the discounted least squares approach is particularly useful when complex non-linear functions are involved and updating equations are not readily available.

SES with drift

If we set β=0, the updating equations (4.15) become:

$$\begin{aligned} L_t &= L_{t-1} + B + \alpha e_t \\ B_t &= B_{t-1} = B \end{aligned}$$

This version may be referred to as *SES with drift*, since the level increases by a fixed amount each period. Although this method is just a special case of LES, the simpler structure makes it easier to derive an optimal value for *B* using the estimation sample, rather than using the start-up values we have considered hitherto. This scheme is often surprisingly effective, as shown by Hyndman and Billah (2003) using the data of the M3 competition.

Tracking signals

Trigg and Leach (1967) introduced the concept of a tracking signal, whereby not only are the level and slope updated each time, but also the smoothing parameters. For example, for SES, we would use the updated value for α given by

$$\alpha_t = \frac{\sum E_t}{\sum M_t} \quad \alpha_t = \frac{E_{t-1}}{M_{t-1}} \tag{4.21}$$

Where E_t and M_t are smoothed values of the error and the absolute error respectively.

$$\begin{aligned} E_t &= \delta e_t + (1-\delta)E_{t-1} \\ M_t &= \delta |e_t| + (1-\delta)M_{t-1} \end{aligned} \tag{4.22}$$

Typically a value of δ in the range 0.1 to 0.2 is used. If a string of positive errors occurs, the value of α_t increases, to speed up the adjustment process; the reverse occurs for negative errors. Over the years there has been considerable debate over the benefits of such signals; for example Gardner (1985) found no real evidence that forecasts based upon tracking signals provided any improvements. A generally preferred approach is to update the parameter estimate regularly, which is no longer much of a computational problem even for large numbers of series. Fildes et al (1998) found that regular updating produced consistent gains over all forecast horizons. From the practical viewpoint, a forecasting system may incorporate data on a weekly or monthly basis and then at regular intervals the the forecasting analysis should be rerun to find revised parameters, so that updating is straightforward and should be unproblematic with modern forecasting systems.

Linear moving averages

In section 4.2 we considered a simple moving average as an alternative to SES, albeit not an approach that we recommended. We could also look at the successive differences in the series $Y_t - Y_{t-1}$ and take a moving average of these values to estimate the slope. The net effect is to estimate the slope by $\frac{Y_t - Y_{t-K}}{K}$ for a K-term moving average. Again, LES usually provides better forecasts.

4.8 Prediction Intervals

Our discussion thus far has been purely in terms of point forecasts, yet in section 2.9 we stressed the importance of interval forecasts. At this stage we restrict attention to the construction of prediction intervals for one-step-ahead forecasts. Intervals for forecasts made two or more periods ahead require the specification of an underlying statistical model and so we defer that discussion to Chapter 9. Indeed the intervals for one-step-ahead also imply an underlying model but we can finesse that requirement for now by using the approximate interval given in section 2.9:

$$\text{Forecast} \pm z_{\alpha/2}*(\text{RMSE})$$

i.e the outcome is expected to life within the range,

(Forecast - $z_{\alpha/2}$*(RMSE), Forecast + $z_{\alpha/2}$*(RMSE), (1-α)% of the time.

Example 4.10: Prediction interval for Netflix

From Table 4.11, the point forecast for the next period (2003.4) was 80.53 with RMSE = 2.91. The resulting (approximate) 95 percent prediction interval is:

$$80.53 \pm 1.96 * 2.91 = 80.53 \pm 5.70 = [74.83, 86.23]$$

As the forecasting horizon increases, the intervals get progressively wider, as we shall see in section 9.x.

4.9 *The Box-Cox Transformations

The logarithmic transformation is appealing because it reflects proportionate change rather than absolute change. For many series in business and economics, the notion of proportional or percentage change is a natural framework; we often encounter statements such as "We expect GDP to grow by 2.5% next year". However, from the empirical perspective, the logarithmic transform is sometimes too strong, in that it may project future growth patterns in excess of reasonable expectations. An examination of Table 4.12 shows that the growth rate for Netflix moderated over time, and the extended series given in Table 4.14 (below) shows an even more marked slow-down.

How can we allow for the "irrational exuberance" sometimes shown by the logarithmic transformation? We have already discussed a modification of LES to allow for a damped trend; this can be applied after the log transform when appropriate. A second possibility is to select a transformation that is more moderate than the logarithmic form.

Instead of restricting the choice to the original units or the logarithmic transform, Box and Cox (1964) suggested using a power transformation, of which the square root and the cube root are the most obvious cases. On occasion, such transformations may have a natural interpretation as when considering the volume of a sphere; the cube root is proportional to the radius of the sphere. It is difficult to find any examples in business or economics where such transformations would have intuitive appeal. Nevertheless, from a purely empirical perspective, such transforms may provide better forecasts.

Given the original series Y_t we define the Box-Cox transform[6] as:

$$Z_t = Y_t^c \tag{4.23}$$

The parameter c is usually restricted to the range $-1 \le c \le 1$ where $c = 1$ corresponds to the original series, and $c = -1$ represents the reciprocal (think of miles per gallon versus gallons per mile). We will use only the square and cube root in our examples; in general, the choice among different values for c may be made by comparing the MSEs for the various options after transforming back to the original units. This reverse transformation is $Y_t = Z_t^{1/c}$.

Example 4.11: Forecasting for Netflix Sales Using the Box-Cox Transform

We considered the square and cube root transformations for the Netflix series. The results are given in Table 4.14 relate to fitting by minimizing the MSE for the transformed series. When the parameters are estimated using the MSE for the original data, the forecasting results are very similar and are omitted.

c	α	β	RMSE	MAE	MAPE
0.333	1.00	0.11	2.50	2.00	6.93
0.50	1.00	0.00	2.37	1.91	6.30

[6] Box and Cox (1964) used the slightly more complex form $Z_t = [Y_t^c - 1]/c$ which has numerical advantages when seeking an optimal value for c. For our purposes the simpler form in equation (4.23) will suffice.

Table 4.14: Summary measures LES for Netflix Sales Using the Box-Cox Transform

When we compare the results in Tables 4.10 and 4.14, the square root transformation seems to provide the best performance using RMSE, but the margin of superiority is very slight. We must stress again the very limited amount of data used in the analysis.

4.10 Principles for Extrapolative Models

To avoid undue repetition, we assume that the data series being forecast is appropriate for the problem to hand, in terms of relevance, timeliness and reporting accuracy. These assumptions are by no means trivial, but we have discussed them in earlier chapters and they remain critical to any forecasting exercise. We should always recall the maxim "Garbage in, garbage out". If the data do not satisfy these criteria, further analysis may be useless. As before, we cross-reference the principles listed in Armstrong (2001) where appropriate. A few principles are repeated from Chapter 2, since they are an integral part of the forecasting approach described in this chapter.

4.1 [5.8] Plot the series (see 2.4)

Data plotting should be the first step in any analysis. If a large number of series is involved, plot a selection of them. Such plots will often serve to identify data recording errors, missing values and unusual events.

4.2 [5.1] Clean the data (see 2.2 and 2.5)

Data plots and simple screening procedures (checks for outliers) provide the basis for making adjustments for anomalous values. Make sure the adjustments are for valid data-recording reasons and keep a record of all such changes.

4.3[5.2] Use transformations as required by process expectations (see 2.3)

Such transformations may involve a conversion current to real dollar values, a logarithmic transformation to reflect proportional growth or a switch to growth rates to account for trends. Intelligent use of knowledge related to the phenomena being studied will help to avoid "crazy" forecasts.

4.4[6.6] Select simple methods unless empirical evidence calls for greater complexity

The set of exponential smoothing methods described in this chapter rely upon only the past history of the series in question. Such extrapolative methods often suffice in the short to medium term unless measurements on key explanatory variables are available.

4.5 [see 6.10, 9.3] Examine the value of alternative methods, preferably using out-of-sample data.

Methods that use more parameters will often "fit" better within the estimation sample, but this advantage may be illusory. Out-of-sample testing provides an even playing field for comparing the performance of different methods. Note that the estimation sample can be used to make such comparisons provided due care is taken; see Chapter 9.

4.6 [9.5] Update the estimates frequently

Regular updating of the parameter estimates is found to improve forecasting performance as it helps to take into account any changes in the behavior of the series. Once the database has been established and updated with the most recent information, such updates are straightforward.

Summary

Extrapolative forecasts are useful in the short to medium term where the recent behavior of the series under study is sufficient to provide a framework for forecasting. When a series does not display marked changes in level over time, simple exponential smoothing (SES) as described in section 4.3 will usually suffice. However, many series do contain systematic trends and in such circumstances linear exponential smoothing (LES) should be considered, as described in section 4.4. A series may display non-linear behavior, either in the growth pattern or due to some kind of life cycle, as is the case for the sales of many products. In sections 4.5 and 4.6 we explore the use of transformations and of damped methods to handle these issues. Other extrapolative methods are reviewed briefly in section 4.7. The construction of one-step-ahead prediction intervals is examined in section 4.8. After an optional section 4.9 on Box-Cox transforms, some basic principles for forecasting using extrapolative methods are summarized in section 4.10.

Exercises

4.1 Table 4.15 shows the growth rate in the U.S. Gross Domestic Product (GDP).

a. Use three and seven term moving averages to generate one-step-ahead forecasts for 2002 to the end of the series. Graph the results and comment on the differences between the two moving averages.

b. Compare the performance of the two procedures by calculating the RMSE and MAE. Why is the MAPE inappropriate in this case?

Year	GDP Growth		Year	GDP Growth
1963	4.4		1986	3.5
1964	5.8		1987	3.4
1965	6.4		1988	4.1
1966	6.5		1989	3.5
1967	2.5		1990	1.9
1968	4.8		1991	-0.2
1969	3.1		1992	3.3
1970	0.2		1993	2.7
1971	3.4		1994	4.0
1972	5.3		1995	2.5
1973	5.8		1996	3.7
1974	-0.5		1997	4.5
1975	-0.2		1998	4.2
1976	5.3		1999	4.5
1977	4.6		2000	3.8
1978	5.6		2001	0.8
1979	3.2		2002	1.9
1980	-0.2		2003	3.0
1981	2.5		2004	4.4
1982	-1.9		2005	2.9
1983	4.5		2006	2.8
1984	7.2		2007	2.0
1985	4.1			

Table 4.15: Percentage change in U.S. Gross Domestic Product, adjusted for inflation. Source: Bureau of Economic Analysis, U.S. Department of Commerce

[GDP change.xlsx]

4.2 Table 4.16 shows the annual percentage change in the Consumer Price Index (CPI) from 1991 – 2007.

a. Use three and seven term moving averages to generate one-step-ahead forecasts for 1998 to the end of the series.

b. Compare the performance of the two procedures by calculating the MSE and MAE.

c. Calculate the MASE for the one-step-ahead forecasts. Why is the MAPE inappropriate in this case?

Year	Percent Increase
1991	3.1
1992	2.9
1993	2.7
1994	2.7
1995	2.5
1996	3.3
1997	1.7
1998	1.6
1999	2.7
2000	3.4
2001	1.6
2002	2.4
2003	1.9
2004	3.3
2005	3.4
2006	2.5
2007	4.1

Table 4.16: Percentage change (December to December) in the Consumer Price Index Source: Bureau of Labor Statistics, U.S. Department of Labor. [CPI change.xlsx]

4.3 Use the data in Table 4.15 to generate forecasts for GDP growth using simple exponential smoothing (SES).

a. Use the observed value for 1963 as the start value and compute the one-step-ahead forecasts for subsequent years using each of α = 0.2, 0.5 and 0.8 in turn.

b. Compare the forecasting performance for these different values of α by calculating the MSE and MAE.

c. How does this method compare with the moving average procedures used in Exercise 4.1? [Be careful to make comparisons over the same time periods.]

4.4 Rework the analysis in Exercise 4.3 using the optimal value of α (chosen using the SES program described in Appendix 4A, or otherwise).

4.5 Use the data in Table 4.16 to generate forecasts for changes in the CPI using simple exponential smoothing (SES).

a. Use the observed value for 1991 as the start value and compute the one-step-ahead forecasts for subsequent years using each of $\alpha = 0.2$, 0.5 and 0.8 in turn.
b. Compare the performance for these different values of α by calculating the MSE and MAE.
c. How does this method compare with the moving average procedures used in Exercise 4.2? [Be careful to make comparisons over the same time periods.]

4.6 Rework the analysis in Exercise 4.3 using the optimal value of α (chosen using the SES program described in Appendix 4A, or otherwise).

4.7 Table 4.17 gives the Spot Price of Saudi Arabian Light Oil (in U.S. dollars per barrel). The prices are quoted for early January in each year.

a. Use the observed value for 1989 as the start value and compute the one-step-ahead forecasts for subsequent years using each of $\alpha = 0.2$, 0.5 and 0.8 in turn.
b. Compute the MSE, MAE and MAPE for each case and contrast the results.
c. Using the same start value, find the optimal level for α, using the data for the period 1990-2003 as the estimation sample. Use *SES.xlsm* as described in Appendix 4A or other suitable software.
d. Generate the forecasts for 2004 – 2006.

Year	Spot Price (Dollars per Barrel)
1989	13.15
1990	18.40
1991	24.00
1992	15.90
1993	16.80
1994	12.40
1995	16.63
1996	18.20
1997	22.98
1998	15.50
1999	10.03
2000	23.45
2001	20.90
2002	18.90
2003	27.39
2004	27.08
2005	34.05
2006	55.01

Table 4.17: Free on Board (FOB) Spot Price of Saudi Arabian Light Oil (in U.S. dollars per barrel). Source: U.S. Department of Energy, Energy Information Administration.

[SA Oil spot.xlsx]

4.8 Repeat the analysis in Exercise 4.7 using linear exponential smoothing [use *LES.xlsm* or other suitable software]. Compare the forecasting performance of the two methods.

4.9 Repeat the analysis in Exercise 4.8 using 1990 – 2004 and then 1990 – 2005 as the estimation samples. Generate the 95 percent prediction intervals for the one-step-ahead forecasts in each case. Do the prediction intervals include the observed values?

4.10 Compute the summary error measures for the Netflix series given in Table 4.7 and summarize the results. Use different time ranges for the estimation sample and determine how the optimal value of α varies.

4.11 Use LES to forecast domestic passengers at Dulles airport for 2004 – 2007 (see Table 2.12, *Dulles.xlsx*), with the years 1963 – 2003 as the estimation sample. Then use the complete data set to make forecasts for 2008 – 2012.

4.12 Dulles Airport was greatly expanded in the mid-eighties, which produced a significant increase in the number of passengers. Rerun the analyses of Exercise 4.11 using the years 1986 – 2003 as the estimation sample. Compare your results with those for the complete sample.

4.13 Rerun the analyses of Exercise 4.10 for the damped trend and log transform versions of LES and compare the results of the four methods using MSE, MAE and MAPE.

4.14 Rerun the analyses of Exercise 4.12 for the damped trend and log transform versions of LES and compare the results of the four methods using MSE, MAE and MAPE.

4.15* Show that: $\ln(Y_t / Y_{t-1}) \doteq \frac{Y_t - Y_{t-1}}{Y_{t-1}}$ so that the log-transform and growth rate methods will often produce similar results. [Hint: ln(1+x)≑x for small x]

4.16* Demonstrate that equation (4.16) may be rewritten as the second equation in (4.15), by substituting the expression for the level at time *t*.

4.17* Repeat the analysis for the data on oil price in Table 4.17 using the Box=Cox transform with c = 0.333 and c = 0.5. Compare the results with the original forecasts (c = 1.0).

Mini-case 4.1: The Growth of Netflix

Home entertainment has become a major element of the modern economy. Whereas 25 years ago, television was the primary form of in-home entertainment, the VCR and then the DVD player have become widely owned alternatives to a night at the movies. In addition, PCs are increasingly used for movie-watching.

In response to these technological developments, a variety of organizations now provide movie rentals. At first these services were restricted to videotape rentals but now focus almost exclusively on DVD rentals and sales. Many such business units are conventional “bricks and mortar” stores including some national chains.

Netflix initiated a different business plan. Subscribers sign up for a given number of movies and they provide Netflix with a list of movies they would like to view. Once one DVD is returned, the next one on the list is mailed to the subscriber. Moreover, since technology does not stand still, Netflix is extending its business to include on-line rentals. The Netflix Annual Report for 2007 indicates that they have over 10 million subscribers and annual total rental revenues in excess of $1.5 billion (as of the end of 2007).

As a market analyst, you would like to know how the company is faring in the marketplace and whether it is likely to continue to grow at the same rate. Table 4.14 is an extension of Table 4.7 and provides the following quarterly data over the period 2000:1 to 2007:4

- S = Quarterly revenue (or Sales) in $million
- G = Sales growth rate over previous quarter (percent)
- N = Number of subscribers at end of quarter (000s)
- R = Revenue per subscriber per month ($).

Year	Quarter	Quarterly Sales	Growth -percent	Subscribers end of period (000s)	Monthly income per subscriber ($/head)
2000	1	5.17	*		
2000	2	7.15	38.1		
2000	3	10.18	42.5		
2000	4	13.39	31.5		
2001	1	17.06	27.4		
2001	2	18.36	7.6		
2001	3	18.88	2.8		
2001	4	21.62	14.5		
2002	1	30.53	41.2		
2002	2	36.36	19.1		
2002	3	40.73	12.0		
2002	4	45.19	10.9		
2003	1	55.67	23.2	1052	
2003	2	63.19	13.5	1147	
2003	3	72.20	14.3	1291	
2003	4	81.19	12.4	1487	
2004	1	99.82	22.9	1932	
2004	2	119.71	19.9	2093	
2004	3	140.41	17.3	2229	
2004	4	140.66	0.2	2610	
2005	1	152.45	8.4	3018	18.91
2005	2	164.03	7.6	3196	18.24
2005	3	172.74	5.3	3592	17.63

2005	4	193.00	11.7	4179	17.27
2006	1	224.13	16.1	4866	17.06
2006	2	239.35	6.8	5169	16.36
2006	3	255.95	6.9	5662	16.24
2006	4	277.23	8.3	6316	15.87
2007	1	305.32	10.1	6797	15.86
2007	2	303.69	-0.5	6742	15.24
2007	3	293.97	-3.2	7028	14.57
2007	4	302.36	2.9	7479	14.22

Table 4.15: Netflix sales data for the period 2000:Q1 20007:Q4

(Source: Annual Reports available on http://ir.netflix.com)

In order to assess the company's prospects, carry out the following analyses and comment upon the results.

1. Develop forecasting methods for each of these series using data through 2006:4 and use them to generate forecasts for 2007 and 2008.
2. Although the definitions do not permit an exact match, it is approximately true that S = N*R*3 (Why?). The use of N and R to forecast S represents an example of disaggregated forecasts, discussed in section 2.x. Would you regard the direct forecasts or the disaggregated forecasts as more reliable in this case? Why?
3. What other factors would you wish to take into account before making forecasts one quarter ahead? Eight quarters ahead?

Mini-case 4.2: The evolution of Wal-Mart

Consider the data on Wal-Mart stores previously given in Table 2.11 and available in *Walmart.xlsx*. In mini-case 2.2 we examined the changing composition of store types and also the company's growth using basic statistical tools. The objective now is to use exponential smoothing methods to generate forecasts.

1. Forecast store composition for the next three years.
2. Forecast sales for the next eight quarters.

In each case plot the fitted and observed values of the series and identify any possible shortcomings in your forecasts. What steps, if any, might be taken to improve the forecasts *without collecting any new data*?

Mini-case 4.3: Volatility in the Dow-Jones Index

The Efficient Markets Hypothesis (EMH) in finance embodies the notion that financial markets rapidly process any new information, so that the best forecast of a stock's price in the next (short) time period is the current price. The EMH is also known as the random walk hypothesis, since any movement from the present state is essentially unpredictable. The errors, or stock returns in this context are $e_t = Y_t - Y_{t-1}$, so defined since "yesterday's" closing price is "today's" forecast. The EMH corresponds to the assumption that the successive errors are independent. There is a vast literature on the EMH in its various forms (see, for example, http://en.wikipedia.org/wiki/Efficient-market_hypothesis) and it is not our purpose to discuss the ideas in any detail. Rather, we note that, in forecasting terms, the EMH translates into the use of SES with $\alpha = 1$, essentially of no real value to the stock trader.

Despite many efforts to the contrary, it is essentially not possible for any purely statistical forecasting method to beat the random walk. Insider information can of course be of considerable value, but those forecasters tend not to be at liberty (to talk about their methods). Nevertheless, there are some interesting questions that can be answered that relate to *volatility*. That is, we can examine the pattern of the forecast errors to determine whether the inherent variability in those forecasts varies over time. When variability is high, considerable opportunities exist for traders to buy or sell options; see Hull (2002) for details. Conversely, when variability is very low, there is little room in the market for such contracts.

We may examine the absolute values of the one-step-ahead errors $|e_t| = |Y_t - Y_{t-1}|$. The spreadsheet *DowJones.xlsx* provides daily closing prices for the period 2001 – 2005, along with the values of the returns and their absolute values.

1. Examine the validity of the EMH for this series.
2. Use the SES to develop a forecasting method for the absolute values of the returns. Is volatility to some extent predictable?

References

Armstrong , J.S. (ed., 2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Kluwer, Boston and Dordrecht.

Box, G.E.P. & Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, series, B, 26,* 211 – 252

Brown, R.G. (1959). *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York, NY

Brown, R.G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series.* Prentice-Hall, Englewood Cliffs, NJ

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting, 8,* 81 – 98

Fildes, R., Hibon, M., Makridakis, S. & Meade, N. (1998). Generalising about univariate forecasting methods: Further empirical evidence. *International Journal of Forecasting, 14,* 339 – 358

Gardner, E.S. Jr (1985). Exponential smoothing: The state of the art. *Journal of Forecasting, 4,* 1 – 28.

Gardner, E. S. Jr (2006). Exponential smoothing: The state of the art – Part II. *International Journal of Forecasting, 22,* 637 – 677.

Gardner, E.S. Jr. and McKenzie, E. (1985). Forecasting trends in time series. *Management Science, 31,* 1237-1246.

Hull, J.C. (2002). *Fundamentals of Futures and Options Markets*. Prentice-Hall, Upper Saddle River, NJ. 4th edition

Hyndman, R.J. & Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting, 19,* 287 – 290

Makridakis, S. & Hibon, M. (2000).The M3-competition: results, conclusions and implications. *International Journal of Forecasting, 16,* 451 – 476

Trigg, D.W. & Leach, D.H. (1967). Exponential smoothing with an adaptive response rate. *Operational Research Quarterly, 18,* 53 - 59

Appendix 4A; Excel Macro for Simple Exponential Smoothing

This appendix provides details of the Excel macro (*SES.xlsm)* for simple exponential smoothing (SES). The upper part of the displayed spreadsheet, which occupies columns A – H summarizes the data entry and basic calculations. Data for the first 15 periods of the WFJ sales data are shown by way of illustration.

Period	Year/Obs	DATA	Level	Error	Error sq	|Error|	APE
1		23056					
2		24817	23056				
3		24300	24339	-39	1535	39	0
4		23242	24310	-1068	1140046	1068	5
5		22862	23533	-670	449321	670	3
6		22863	23044	-182	33041	182	1
7		23391	22912	479	229389	479	2
8		22469	23261	-791	626352	791	4
9		22241	22685	-443	196367	443	2
10		24367	22362	2005	4018968	2005	8
11		29457	23822	5635	31750893	5635	19
12		31294	27925	3368	11344621	3368	11
13		38713	30378	8335	69476256	8335	22
14		35749	36448	-699	488746	699	2
15		39768	35939	3829	14661921	3829	10

The values in these columns are as follows:

Period	the observation number
Year/Obs	the identifying value for the observation (e.g. November 1942), placed in a single column [optional]
Data	the data series; a new series may be pasted in along with its name
Level	the level of the series, computed in accordance with equation (4.8)

Error	the difference between the observed value and the fitted value (for the estimation sample) or one-step-ahead forecast (for the hold-out sample)
Error sq	the squared errors
|Error|	the absolute values of the errors
APE	The absolute percentage errors

The number of decimal places displayed in each column may be adjusted in the usual way.

The second part of the display, covering columns I – N provides the initial settings, the operational keys and the summary statistics.

Restore Default Values

level_1=	23,056
Alpha (α) =	0.300

	MSE	MAE	MAPE	RMSE
Estimation sample	11458923	2240	7	3385.10
Hold-out sample	23689514	3963	11	4867.19

Start-level Yr 1

Run Solver!

Holdout

25

Total Obs.	**Max Holdout**
62	**43**

The various elements of this display are as follows:

level_1	Starting value for the level
Alpha	Value of the smoothing constant (α=0.3 is the default)

Estimation sample	Values in this row of the table relate to that part of the series used to estimate the smoothing parameter
Hold-out sample	Corresponding values for the hold-out sample
MSE	Mean square error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
RMSE	Root mean square error
Start-level 1	Select the first observation to set level_1 (default) or to the average of the first four observations (Start-level 4).
Holdout	Size of the hold-out sample (default is about 40 percent)
Total Obs.	Number of observations
Max Holdout	Maximum size of the hold-out sample (about 70 percent)
Restore Default Values	Click to reset values to defaults (holdout and alpha)
Run Solver!	Select optimal value for alpha using Solver

Running the Macro

1. Open the spreadsheet *SES.xlsm* and be sure to **Enable Macros.**
2. If a new data set is to be entered copy the values into columns B (if observations have labels) and C.
3. If the new series is shorter than the old series, be sure to delete the old values from the end of the series.
4. Click on **Restore Default Values.**
5. Select **Start-level** and **Holdout** values.
6. **Run Solver**! Retain solution or return to original values to run the macro again with different settings.

*If Solver does not run please go to the troubleshooting section at the end of this appendix.

Output

The macro generates the MSE, MAE, MAPE and RMSE for both estimation and hold-out samples. The optimal value for alpha is obtained by minimizing the MSE for the estimation sample. The one-step-ahead forecasts for the hold-out sample are generated using the optimal alpha value.

To fit using minimum MAE or minimum MAPE, go to the appropriate sheet in the macro (e.g. SES-first 1) and change the target cell to L6 or M6 respectively. The new criterion will stay in place until another change is made.

Appendix 4B: Excel Macro for Linear Exponential Smoothing

This appendix provides details of the Excel macro (*LES.xlsm)* for linear exponential smoothing (LES). Four variations are considered:

- Linear exponential smoothing (section 4.4)
- LES with a logarithmic transformation (section 4.5)
- Damped linear exponential smoothing (section 4.6)
- LES with a Box-Cox transformation* (section 4.9)

The data set in the example refers to the number of domestic passengers at Dulles airport (see Table 2.12 and Exercise 4.11). The layout of upper part of the displayed spreadsheet refers to LES without dampening. Extra columns are required for the cases involving transformations. Columns A – J summarize the data entry and basic calculations.

Observation	Year/Obs	DATA	Level	Slope	Forecast	Error	Error sq	|Error|	APE
1	1963	641							
2	1964	728							
3	1965	920							
4	1966	1079	763	140	903	176	30849	176	16.3
5	1967	1427	956	145	1101	327	106652	327	22.9
6	1968	1602	1199	155	1354	248	61739	248	15.5
7	1969	1928	1428	162	1591	337	113715	337	17.5

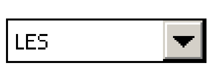
8	1970	1869	1692	173	1865	5	20	5	0.2
9	1971	1881	1866	173	2039	-157	24789	157	8.4
10	1972	1992	1992	168	2160	-167	27931	167	8.4
11	1973	2083	2109	163	2272	-189	35838	189	9.1
12	1974	2004	2216	157	2373	-369	135920	369	18.4
13	1975	2000	2262	146	2409	-408	166553	408	20.4
14	1976	2251	2286	134	2420	-169	28590	169	7.5
15	1977	2267	2369	129	2498	-231	53398	231	10.2

The values in these columns are as follows:

Period	the observation number
Year/Obs	the identifying value for the observation (e.g. November 1942), placed in a single column [optional]
Data	the data series; a new series may be pasted in along with its name
Level	the level of the series, computed in accordance with equation (4.15)
Slope	the level of the series, computed in accordance with equation (4.15)
Forecast	The forecast, computed using the slope and the level, as in equation (4.14)
Error	the difference between the observed value and the fitted value (for the estimation sample) or one-step-ahead forecast (for the hold-out sample)
Error sq	the squared errors
|*Error*|	the absolute values of the errors
APE	The absolute percentage errors

The number of decimal places displayed in each column may be adjusted in the usual way.

The second part of the display, covering columns I – N provides the initial settings, the operational keys and the summary statistics.



Select the Solving Method						# of Obs. Holdout	18	
			not relevant		not relevant			
Alpha (α) =	0.3000	Beta (β) =	0.1000	Phi =	0.9500	C =	0.5000	

Estimation	MSE	MAE	MAPE	ME	Hold-out	MSE	MAE	MAPE	ME
	2393357	841	19	526		4658067	1562	11	304
RMSE	1547				RMSE	2158			

The various elements of this display are as follows:

level_1	Starting value for the level
Alpha	Smoothing constant for the level (α=0.3 is the default)
Beta	Smoothing constant for the slope (β=0.1 is the default)
Phi	Amount of dampening (φ = 0.95 is the default when "Damped" is chosen; otherwise not relevant)
C	Power used in the Box-Cox transform (C=0.5 is the default when the transform is used; otherwise not relevant)
Estimation sample	Values in this row of the table relate to that part of the series used to estimate the smoothing parameter
Hold-out sample	Corresponding values for the hold-out sample
MSE	Mean square error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
RMSE	Root mean square error
# of Obs. Holdout	Size of the hold-out sample (default is about 40 percent)
*Total Obs. **	Number of observations
*Max Holdout **	Maximum size of the hold-out sample (about 70 percent)
*Restore Default Values **	Click to reset values to defaults (holdout and alpha)
*Run Solver! **	Select optimal value for alpha using Solver

- Not shown for reasons of space, but in the same form as for SES.

Running the Macro

1. Open the spreadsheet *LES.xlsm* and be sure to **Enable Macros.**
2. If a new data set is to be entered copy the values into columns B (if observations have labels) and C.
3. If the new series is shorter than the old series, be sure to delete the old values from the end of the series.
4. Click on **Restore Default Values.**
5. Select size of **Holdout** sample.
6. Select the method option: **LES, Damped, Log Form, C-transform** [additional columns appear when a transform is used]
7. **Run Solver**! Retain solution or return to original values to run the macro again with different settings.

*If Solver does not run please go to the troubleshooting section at the end of this appendix.

Output

The macro generates the MSE, MAE, MAPE and RMSE for both estimation and hold-out samples. The optimal values for the parameters are obtained by minimizing the MSE [cell C8] for the estimation sample of the (transformed) series. The one-step-ahead forecasts for the hold-out sample are generated using the optimal estimates.

Method	**Alpha**	**Beta**	**Phi**	**C**
LES	Fitted	Fitted	NR (phi = 1.0)	NR (C = 1)
Damped	Fitted	Fitted	Fitted	NR (C = 1)
Log Form	Fitted	Fitted	NR (phi = 1.0)	NR (C → 0)
C-transform	Fitted	Fitted	NR (phi = 1.0)	Pre-set (C = 0.5)

NR = Not relevant, implied default value in parentheses

To fit using minimum MAE or minimum MAPE, or to use the original values rather than the transforms, go to the appropriate sheet in the macro (e.g. LES) the target cell appropriately. The new criterion will stay in place until another change is made.

Troubleshooting

Solver doesn't run?

1) Go to Excel options
2) Select Add-ins
3) Browse for "Solver Add-in" and select it. (if inactive)
4) Click "Go"
5) Check "Solver Add-in" on the list and click ok.

I do have the Solver Add-in but still doesn't run.

1) Go to "Developer" Menu
2) Click on "Visual Basic" Icon (This should open visual basic)
3) Go to "Tools" Menu, and select "References."
4) On the pop-up window select "SOLVER's" checkbox. (If there is another line containing "Solver" un-check it)