3

Forecasting Strategies

3.1 Purpose

Businesses rely on forecasts of sales to plan production, justify marketing decisions, and guide research. A very efficient method of forecasting one variable is to find a related variable that leads it by one or more time intervals. The closer the relationship and the longer the lead time, the better this strategy becomes. The trick is to find a suitable lead variable. An Australian example is the Building Approvals time series published by the Australian Bureau of Statistics. This provides valuable information on the likely demand over the next few months for all sectors of the building industry. A variation on the strategy of seeking a leading variable is to find a variable that is associated with the variable we need to forecast and easier to predict.

In many applications, we cannot rely on finding a suitable leading variable and have to try other methods. A second approach, common in marketing, is to use information about the sales of similar products in the past. The influential Bass diffusion model is based on this principle. A third strategy is to make extrapolations based on present trends continuing and to implement adaptive estimates of these trends. The statistical technicalities of forecasting are covered throughout the book, and the purpose of this chapter is to introduce the general strategies that are available.

3.2 Leading variables and associated variables

3.2.1 Marine coatings

A leading international marine paint company uses statistics available in the public domain to forecast the numbers, types, and sizes of ships to be built over the next three years. One source of such information is *World Shipyard Monitor*, which gives brief details of orders in over 300 shipyards. The paint company has set up a database of ship types and sizes from which it can

P.S.P. Cowpertwait and A.V. Metcalfe, *Introductory Time Series with R*, Use R, DOI 10.1007/978-0-387-88698-5_3,

forecast the areas to be painted and hence the likely demand for paint. The company monitors its market share closely and uses the forecasts for planning production and setting prices.

3.2.2 Building approvals publication

Building approvals and building activity time series

The Australian Bureau of Statistics publishes detailed data on building approvals for each month, and, a few weeks later, the Building Activity Publication lists the value of building work done in each quarter. The data in the file `ApprovActiv.dat` are the total dwellings approved per month, averaged over the past three months, labelled "Approvals", and the value of work done over the past three months (chain volume measured in millions of Australian dollars at the reference year 2004–05 prices), labelled "Activity", from March 1996 until September 2006. We start by reading the data into `R` and then construct time series objects and plot the two series on the same graph using `ts.plot` (Fig. 3.1).

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/ApprovActiv.dat"
> Build.dat <- read.table(www, header=T) ; attach(Build.dat)
> App.ts <- ts(Approvals, start = c(1996,1), freq=4)
> Act.ts <- ts(Activity, start = c(1996,1), freq=4)
> ts.plot(App.ts, Act.ts, lty = c(1,3))
```

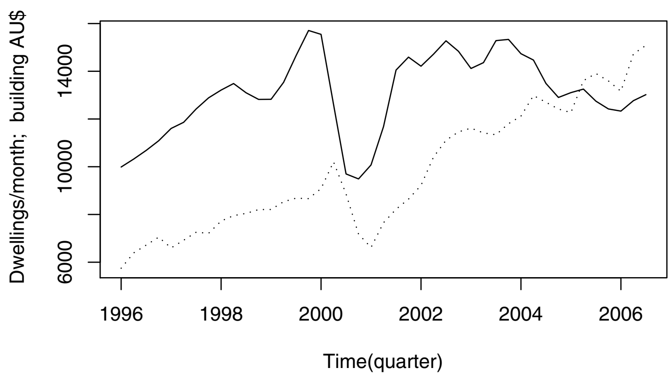

Fig. 3.1. Building approvals (solid line) and building activity (dotted line).

In Figure 3.1, we can see that the building activity tends to lag one quarter behind the building approvals, or equivalently that the building approvals appear to lead the building activity by a quarter. The *cross-correlation function*, which is abbreviated to *ccf*, can be used to quantify this relationship. A plot of the cross-correlation function against lag is referred to as a *cross-correlogram*.

Cross-correlation

Suppose we have time series models for variables x and y that are stationary in the mean and the variance. The variables may each be serially correlated, and correlated with each other at different time lags. The combined model is *second-order stationary* if all these correlations depend only on the lag, and then we can define the *cross covariance function* (*ccvf*), $\gamma_k(x, y)$, as a function of the lag, k:

$$\gamma_k(x, y) = E\left[(x_{t+k} - \mu_x)(y_t - \mu_y)\right] \tag{3.1}$$

This is not a symmetric relationship, and the variable x is lagging variable y by k. If x is the input to some physical system and y is the response, the cause will precede the effect, y will lag x, the ccvf will be 0 for positive k, and there will be spikes in the ccvf at negative lags. Some textbooks define ccvf with the variable y lagging when k is positive, but we have used the definition that is consistent with `R`. Whichever way you choose to define the ccvf,

$$\gamma_k(x, y) = \gamma_{-k}(y, x) \tag{3.2}$$

When we have several variables and wish to refer to the acvf of one rather than the ccvf of a pair, we can write it as, for example, $\gamma_k(x, x)$. The lag k cross-correlation function (*ccf*), $\rho_k(x, y)$, is defined by

$$\rho_k(x, y) = \frac{\gamma_k(x, y)}{\sigma_x \sigma_y} \quad . \tag{3.3}$$

The ccvf and ccf can be estimated from a time series by their sample equivalents. The sample ccvf, $c_k(x, y)$, is calculated as

$$c_k(x, y) = \frac{1}{n} \sum_{t=1}^{n-k} \left(x_{t+k} - \overline{x}\right)\left(y_t - \overline{y}\right) \tag{3.4}$$

The sample acf is defined as

$$r_k(x, y) = \frac{c_k(x, y)}{\sqrt{c_0(x, x) c_0(y, y)}} \tag{3.5}$$

Cross-correlation between building approvals and activity

The `ts.union` function binds time series with a common frequency, padding with 'NA's to the union of their time coverages. If `ts.union` is used within the `acf` command, `R` returns the correlograms for the two variables and the cross-correlograms in a single figure.

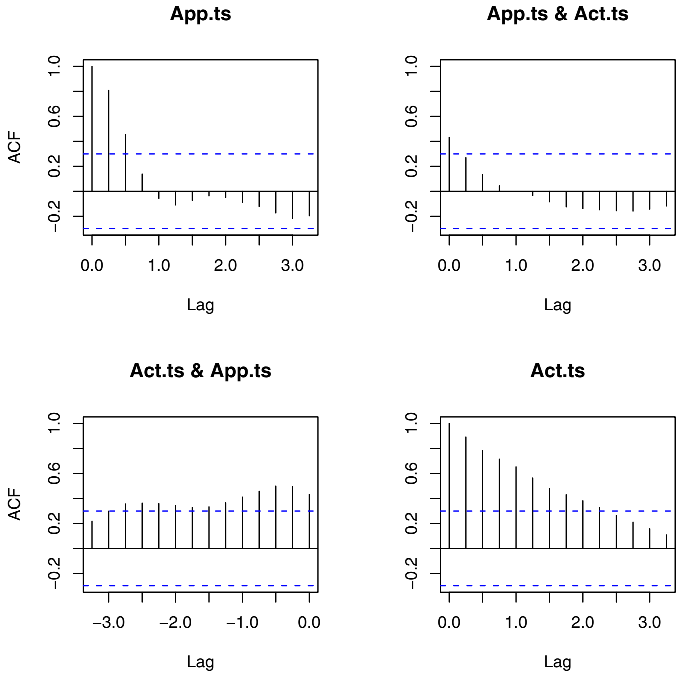

Fig. 3.2. Correlogram and cross-correlogram for building approvals and building activity.

```
> acf(ts.union(App.ts, Act.ts))
```

In Figure 3.2, the acfs for x and y are in the upper left and lower right frames, respectively, and the ccfs are in the lower left and upper right frames. The time unit for lag is one year, so a correlation at a lag of one quarter appears at 0.25. If the variables are independent, we would expect 5% of sample correlations to lie outside the dashed lines. Several of the cross-correlations at negative lags do pass these lines, indicating that the approvals time series is leading the activity. Numerical values can be printed using the `print()` function, and are 0.432, 0.494, 0.499, and 0.458 at lags of 0, 1, 2, and 3, respectively. The ccf can be calculated for any two time series that overlap, but if they both have trends or similar seasonal effects, these will dominate (Exercise 1). It may be that common trends and seasonal effects are precisely what we are looking for, but the population ccf is defined for stationary random processes and it is usual to remove the trend and seasonal effects before investigating cross-correlations. Here we remove the trend using `decompose`, which uses a centred moving average of the four quarters (see Fig. 3.3). We will discuss the use of `ccf` in later chapters.

```
> app.ran <- decompose(App.ts)$random
> app.ran.ts <- window (app.ran, start = c(1996, 3) )
> act.ran <- decompose (Act.ts)$random
> act.ran.ts <- window (act.ran, start = c(1996, 3) )
> acf (ts.union(app.ran.ts, act.ran.ts))
> ccf (app.ran.ts, act.ran.ts)
```

We again use `print()` to obtain the following table.

```
> print(acf(ts.union(app.ran.ts, act.ran.ts)))

     app.ran.ts     act.ran.ts
  1.000 ( 0.00)  0.123 ( 0.00)
  0.422 ( 0.25)  0.704 (-0.25)
 -0.328 ( 0.50)  0.510 (-0.50)
 -0.461 ( 0.75) -0.135 (-0.75)
 -0.400 ( 1.00) -0.341 (-1.00)
 -0.193 ( 1.25) -0.187 (-1.25)
...

     app.ran.ts     act.ran.ts
  0.123 ( 0.00)  1.000 ( 0.00)
 -0.400 ( 0.25)  0.258 ( 0.25)
 -0.410 ( 0.50) -0.410 ( 0.50)
 -0.250 ( 0.75) -0.411 ( 0.75)
  0.071 ( 1.00) -0.112 ( 1.00)
  0.353 ( 1.25)  0.180 ( 1.25)
...
```

The `ccf` function produces a single plot, shown in Figure 3.4, and again shows the lagged relationship. The Australian Bureau of Statistics publishes the building approvals by state and by other categories, and specific sectors of the building industry may find higher correlations between demand for their products and one of these series than we have seen here.

3.2.3 Gas supply

Gas suppliers typically have to place orders for gas from offshore fields 24 hours ahead. Variation about the average use of gas, for the time of year, depends on temperature and, to some extent, humidity and wind speed. Coleman et al. (2001) found that the weather accounts for 90% of this variation in the United Kingdom. Weather forecasts for the next 24 hours are now quite accurate and are incorporated into the forecasting procedure.

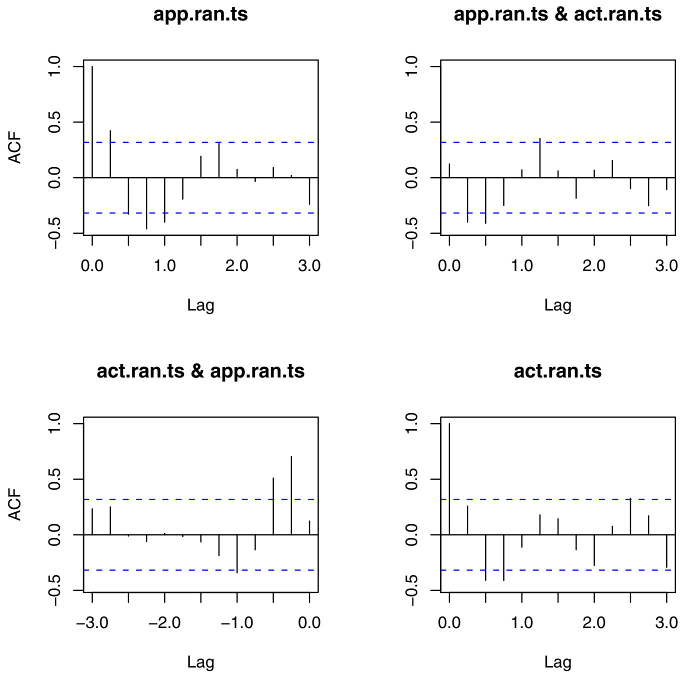

Fig. 3.3. Correlogram and cross-correlogram of the random components of building approvals and building activity after using `decompose`.

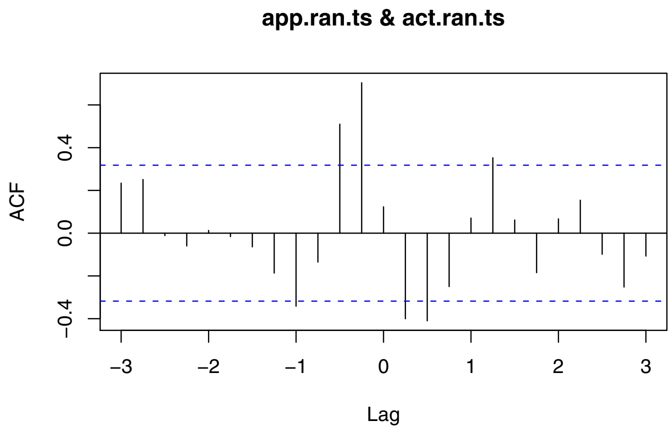

Fig. 3.4. Cross-correlogram of the random components of building approvals and building activity after using `decompose`.

3.3 Bass model

3.3.1 Background

Frank Bass published a paper describing his mathematical model, which quantified the theory of adoption and diffusion of a new product by society (Rogers, 1962), in *Management Science* nearly fifty years ago (Bass, 1969). The mathematics is straightforward, and the model has been influential in marketing. An entrepreneur with a new invention will often use the Bass model when making a case for funding. There is an associated demand for market research, as demonstrated, for example, by the Marketing Science Centre at the University of South Australia becoming the Ehrenberg-Bass Institute for Marketing Science in 2005.

3.3.2 Model definition

The Bass formula for the number of people, N_t, who have bought a product at time t depends on three parameters: the total number of people who eventually buy the product, m; the *coefficient of innovation*, p; and the *coefficient of imitation*, q. The Bass formula is

$$N_{t+1} = N_t + p(m - N_t) + qN_t(m - N_t)/m \tag{3.6}$$

According to the model, the increase in sales, $N_{t+1} - N_t$, over the next time period is equal to the sum of a fixed proportion p and a time varying proportion $q\frac{N_t}{m}$ of people who will eventually buy the product but have not yet done so. The rationale for the model is that initial sales will be to people who are interested in the novelty of the product, whereas later sales will be to people who are drawn to the product after seeing their friends and acquaintances use it. Equation (3.6) is a difference equation and its solution is

$$N_t = m\frac{1 - \mathrm{e}^{-(p+q)t}}{1 + (q/p)\mathrm{e}^{-(p+q)t}} \tag{3.7}$$

It is easier to verify this result for the continuous-time version of the model.

3.3.3 Interpretation of the Bass model*

One interpretation of the Bass model is that the time from product launch until purchase is assumed to have a probability distribution that can be parametrised in terms of p and q. A plot of sales per time unit against time is obtained by multiplying the probability density by the number of people, m, who eventually buy the product. Let $f(t)$, $F(t)$, and $h(t)$ be the density, cumulative distribution function (cdf), and hazard, respectively, of the distribution of time until purchase. The definition of the hazard is

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{3.8}$$

The interpretation of the hazard is that if it is multiplied by a small time increment it gives the probability that a random purchaser who has not yet made the purchase will do so in the next small time increment (Exercise 2). Then the continuous time model of the Bass formula can be expressed in terms of the hazard:

$$h(t) = p + qF(t) \tag{3.9}$$

Equation (3.6) is the discrete form of Equation (3.9) (Exercise 2). The solution of Equation (3.8), with $h(t)$ given by Equation (3.9), for $F(t)$ is

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + (q/p)e^{-(p+q)t}} \tag{3.10}$$

Two special cases of the distribution are the *exponential distribution* and *logistic distribution*, which arise when $q = 0$ and $p = 0$, respectively. The logistic distribution closely resembles the normal distribution (Exercise 3). Cumulative sales are given by the product of m and $F(t)$. The pdf is the derivative of Equation (3.10):

$$f(t) = \frac{(p+q)^2 e^{-(p+q)t}}{p\left[1 + (q/p)e^{-(p+q)t}\right]^2} \tag{3.11}$$

Sales per unit time at time t are

$$S(t) = mf(t) = \frac{m(p+q)^2 e^{-(p+q)t}}{p\left[1 + (q/p)e^{-(p+q)t}\right]^2} \tag{3.12}$$

The time to peak is

$$t_{\text{peak}} = \frac{\log(q) - \log(p)}{p + q} \tag{3.13}$$

3.3.4 Example

We show a typical Bass curve by fitting Equation (3.12) to yearly sales of VCRs in the US home market between 1980 and 1989 (Bass website) using the `R` non-linear least squares function `nls`. The variable `T79` is the year from 1979, and the variable `Tdelt` is the time from 1979 at a finer resolution of 0.1 year for plotting the Bass curves. The cumulative sum function `cumsum` is useful for monitoring changes in the mean level of the process (Exercise 8).

```
> T79 <- 1:10
> Tdelt <- (1:100) / 10
> Sales <- c(840,1470,2110,4000, 7590, 10950, 10530, 9470, 7790, 5890)
> Cusales <- cumsum(Sales)
> Bass.nls <- nls(Sales ~ M * ( ((P+Q)^2 / P) * exp(-(P+Q) * T79) ) /
  (1+(Q/P)*exp(-(P+Q)*T79))^2, start = list(M=60630, P=0.03, Q=0.38))
> summary(Bass.nls)
```

```
Parameters:
   Estimate Std. Error t value Pr(>|t|)
M 6.798e+04  3.128e+03   21.74 1.10e-07 ***
P 6.594e-03  1.430e-03    4.61  0.00245 **
Q 6.381e-01  4.140e-02   15.41 1.17e-06 ***

Residual standard error: 727.2 on 7 degrees of freedom
```

The final estimates for m, p, and q, rounded to two significant places, are 68000, 0.0066, and 0.64 respectively. The starting values for P and Q are p and q for a typical product. We assume the sales figures are prone to error and estimate the total sales, m, setting the starting value for M to the recorded total sales. The data and fitted curve can be plotted using the code below (see Fig. 3.5 and 3.6):

```
> Bcoef <- coef(Bass.nls)
> m <- Bcoef[1]
> p <- Bcoef[2]
> q <- Bcoef[3]
> ngete <- exp(-(p+q) * Tdelt)
> Bpdf <- m * ( (p+q)^2 / p ) * ngete / (1 + (q/p) * ngete)^2
> plot(Tdelt, Bpdf, xlab = "Year from 1979",
                    ylab = "Sales per year", type='l')
> points(T79, Sales)
> Bcdf <- m * (1 - ngete)/(1 + (q/p)*ngete)
> plot(Tdelt, Bcdf, xlab = "Year from 1979",
                    ylab = "Cumulative sales", type='l')
> points(T79, Cusales)
```

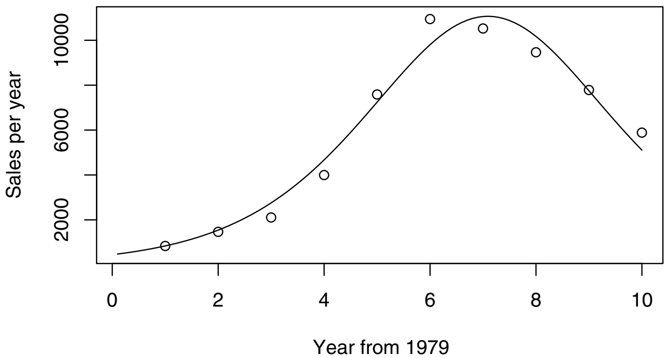

Fig. 3.5. Bass sales curve fitted to sales of VCRs in the US home market, 1980–1989.

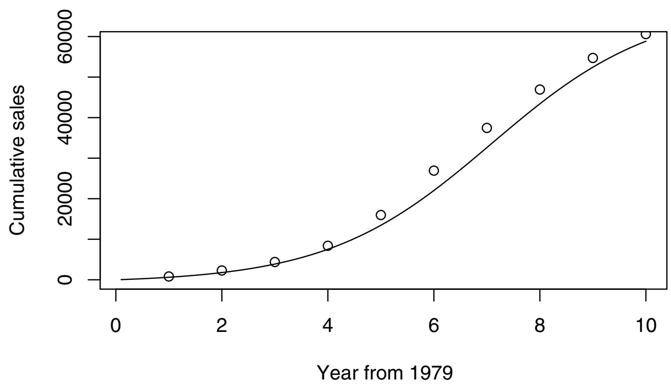

Fig. 3.6. Bass cumulative sales curve, obtained as the integral of the sales curve, and cumulative sales of VCRs in the US home market, 1980–1989.

It is easy to fit a curve to past sales data. The importance of the Bass curve in marketing is in forecasting, which needs values for the parameters m, p, and q. Plausible ranges for the parameter values can be based on published data for similar categories of past inventions, and a few examples follow.

Product	m	p	q	Reference
Typical product	-	0.030	0.380	VBM[1]
35 mm projectors, 1965–1986	3.37 million	0.009	0.173	Bass[2]
Overhead projectors, 1960–1970	0.961 million	0.028	0.311	Bass
PCs, 1981–2010	3.384 billion	0.001	0.195	Bass

[1]Value-Based Management; [2]Frank M. Bass, 1999.

Although the forecasts are inevitably uncertain, they are the best information available when making marketing and investment decisions. A prospectus for investors or a report to the management team will typically include a set of scenarios based on the most likely, optimistic, and pessimistic sets of parameters.

The basic Bass model does not allow for replacement sales and multiple purchases. Extensions of the model that allow for replacement sales, multiple purchases, and the effects of pricing and advertising in a competitive market have been proposed (for example, Mahajan et al. 2000). However, there are several reasons why these refinements may be of less interest to investors than you might expect. The first is that the profit margin on manufactured goods, such as innovative electronics and pharmaceuticals, will drop dramatically once patent protection expires and competitors enter the market. A second reason is that successful inventions are often superseded by new technology, as

VCRs have been by DVD players, and replacement sales are limited. Another reason is that many investors are primarily interested in a relatively quick return on their money. You are asked to consider Bass models for sales of two recent 3G mobile communication devices in Exercise 4.

3.4 Exponential smoothing & the Holt-Winters method

3.4.1 Exponential smoothing

Our objective is to predict some future value x_{n+k} given a past history $\{x_1, x_2, \ldots, x_n\}$ of observations up to time n. In this subsection we assume there is no systematic trend or seasonal effects in the process, or that these have been identified and removed. The mean of the process can change from one time step to the next, but we have no information about the likely direction of these changes. A typical application is forecasting sales of a well-established product in a stable market. The model is

$$x_t = \mu_t + w_t \tag{3.14}$$

where μ_t is the non-stationary mean of the process at time t and w_t are independent random deviations with a mean of 0 and a standard deviation σ. We will follow the notation in `R` and let a_t be our estimate of μ_t. Given that there is no systematic trend, an intuitively reasonable estimate of the mean at time t is given by a weighted average of our observation at time t and our estimate of the mean at time $t - 1$:

$$a_t = \alpha x_t + (1 - \alpha) a_{t-1} \qquad 0 < \alpha < 1 \tag{3.15}$$

The a_t in Equation (3.15) is the *exponentially weighted moving average (EWMA)* at time t. The value of α determines the amount of smoothing, and it is referred to as the *smoothing parameter*. If α is near 1, there is little smoothing and a_t is approximately x_t. This would only be appropriate if the changes in the mean level were expected to be large by comparison with σ. At the other extreme, a value of α near 0 gives highly smoothed estimates of the mean level and takes little account of the most recent observation. This would only be appropriate if the changes in the mean level were expected to be small compared with σ. A typical compromise figure for α is 0.2 since in practice we usually expect that the change in the mean between time $t - 1$ and time t is likely to be smaller than σ. Alternatively, `R` can provide an estimate for α, and we discuss this option below. Since we have assumed that there is no systematic trend and that there are no seasonal effects, forecasts made at time n for any lead time are just the estimated mean at time n. The forecasting equation is

$$\hat{x}_{n+k|n} = a_n \qquad k = 1, 2, \ldots \tag{3.16}$$

Equation (3.15), for a_t, can be rewritten in two other useful ways. Firstly, we can write the sum of a_{t-1} and a proportion of the one-step-ahead forecast error, $x_t - a_{t-1}$,

$$a_t = \alpha(x_t - a_{t-1}) + a_{t-1} \tag{3.17}$$

Secondly, by repeated back substitution we obtain

$$a_t = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \dots \tag{3.18}$$

When written in this form, we see that a_t is a linear combination of the current and past observations, with more weight given to the more recent observations. The restriction $0 < \alpha < 1$ ensures that the weights $\alpha(1-\alpha)^i$ become smaller as i increases. Note that these weights form a geometric series, and the sum of the infinite series is unity (Exercise 5). We can avoid the infinite regression by specifying $a_1 = x_1$ in Equation (3.15).

For any given α, the model in Equation (3.17) together with the starting value $a_1 = x_1$ can be used to calculate a_t for $t = 2, 3, \dots, n$. One-step-ahead prediction errors, e_t, are given by

$$e_t = x_t - \hat{x}_{t|t-1} = x_t - a_{t-1} \tag{3.19}$$

By default, `R` obtains a value for the smoothing parameter, α, by minimising the sum of squared one-step-ahead prediction errors ($SS1PE$):

$$SS1PE = \sum_{t=2}^{n} e_t^2 = e_2^2 + e_3^2 + \dots + e_n^2 \qquad a_1 = x_1 \tag{3.20}$$

However, calculating α in this way is not necessarily the best practice. If the time series is long and the mean has changed little, the value of α will be small. In the specific case where the mean of the process does not change, the optimum value for α is $\frac{1}{n}$. An exponential smoothing procedure set up with a small value of α will be slow to respond to any unexpected change in the market, as occurred in sales of videotapes, which plummeted after the invention of DVDs.

Complaints to a motoring organisation

The number of letters of complaint received each month by a motoring organisation over the four years 1996 to 1999 are available on the website. At the beginning of the year 2000, the organisation wishes to estimate the current level of complaints and investigate any trend in the level of complaints. We should first plot the data, and, even though there are only four years of data, we should check for any marked systematic trend or seasonal effects.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/motororg.dat"
> Motor.dat <- read.table(www, header = T); attach(Motor.dat)
> Comp.ts <- ts(complaints, start = c(1996, 1), freq = 12)
> plot(Comp.ts, xlab = "Time / months", ylab = "Complaints")
```

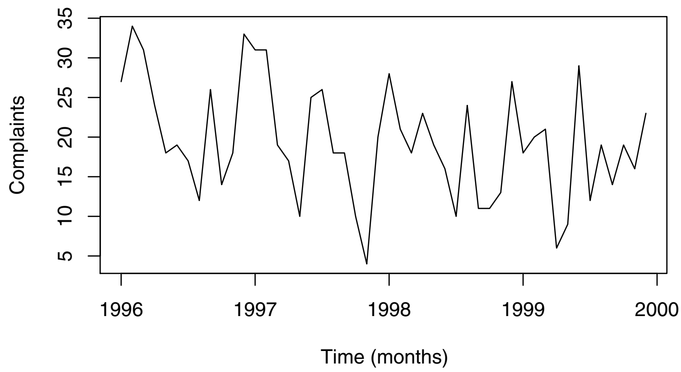

Fig. 3.7. Monthly numbers of letters of complaint received by a motoring organisation.

There is no evidence of a systematic trend or seasonal effects, so it seems reasonable to use exponential smoothing for this time series. Exponential smoothing is a special case of the Holt-Winters algorithm, which we introduce in the next section, and is implemented in `R` using the `HoltWinters` function with the additional parameters set to 0. If we do not specify a value for α, `R` will find the value that minimises the one-step-ahead prediction error.

```
> Comp.hw1 <- HoltWinters(complaints, beta = 0, gamma = 0) ; Comp.hw1
> plot(Comp.hw1)

Holt-Winters exponential smoothing without trend and without seasonal
component.

Smoothing parameters:
 alpha:  0.143
 beta :  0
 gamma:  0

Coefficients:
    [,1]
 a 17.70

> Comp.hw1$SSE

[1] 2502
```

The estimated value of the mean number of letters of complaint per month at the end of 1999 is 17.7. The value of α that gives a minimum SS1PE, of 2502, is 0.143. We now compare these results with those obtained if we specify a value for α of 0.2.

```
> Comp.hw2 <- HoltWinters(complaints, alpha = 0.2, beta=0, gamma=0)
> Comp.hw2

...
 alpha:  0.2
 beta :  0
 gamma:  0

Coefficients:
   [,1]
a 17.98

> Comp.hw2$SSE

[1] 2526
```

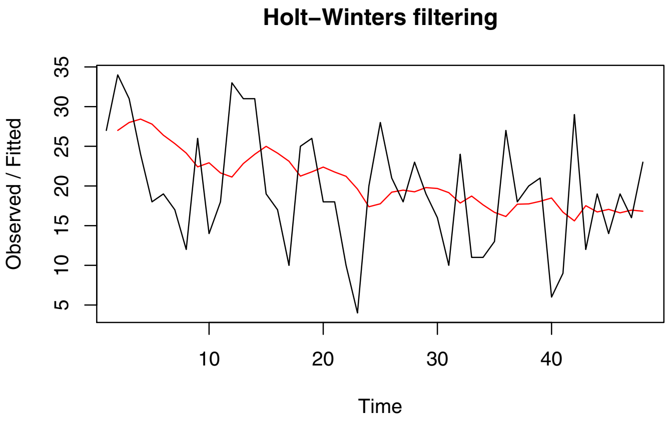

Fig. 3.8. Monthly numbers of letters and exponentially weighted moving average.

The estimated value of the mean number of letters of complaint per month at the end of 1999 is now 18.0, and the SS1PE has increased slightly to 2526. The advantage of letting R estimate a value for α is that it is optimum for a practically important criterion, SS1PE, and that it removes the need to make a choice. However, the optimum estimate can be close to 0 if we have a long time series over a stable period, and this makes the EWMA unresponsive to any future change in mean level. From Figure 3.8, it seems that there was a decrease in the number of complaints at the start of the period and a slight rise towards the end, although this has not yet affected the exponentially weighted moving average.

3.4.2 Holt-Winters method

We usually have more information about the market than exponential smoothing can take into account. Sales are often seasonal, and we may expect trends to be sustained for short periods at least. But trends will change. If we have a successful invention, sales will increase initially but then stabilise before declining as competitors enter the market. We will refer to the change in level from one time period to the next as the *slope*.[1] Seasonal patterns can also change due to vagaries of fashion and variation in climate, for example. The Holt-Winters method was suggested by Holt (1957) and Winters (1960), who were working in the School of Industrial Administration at Carnegie Institute of Technology, and uses exponentially weighted moving averages to update estimates of the seasonally adjusted mean (called the *level*), slope, and seasonals.

The Holt-Winters method generalises Equation (3.15), and the additive seasonal form of their updating equations for a series $\{x_t\}$ with period p is

$$\left.\begin{aligned} a_t &= \alpha(x_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \\ b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(x_t - a_t) + (1-\gamma)s_{t-p} \end{aligned}\right\} \tag{3.21}$$

where a_t, b_t, and s_t are the estimated level,[2] slope, and seasonal effect at time t, and α, β, and γ are the smoothing parameters. The first updating equation takes a weighted average of our latest observation, with our existing estimate of the appropriate seasonal effect subtracted, and our forecast of the level made one time step ago. The one-step-ahead forecast of the level is the sum of the estimates of the level and slope at the time of forecast. A typical choice of the weight α is 0.2. The second equation takes a weighted average of our previous estimate and latest estimate of the slope, which is the difference in the estimated level at time t and the estimated level at time $t-1$. Note that the second equation can only be used after the first equation has been applied to get a_t. Finally, we have another estimate of the seasonal effect, from the difference between the observation and the estimate of the level, and we take a weighted average of this and the last estimate of the seasonal effect for this season, which was made at time $t-p$. Typical choices of the weights β and γ are 0.2. The updating equations can be started with $a_1 = x_1$ and initial slope, b_1, and seasonal effects, $s_1, \ldots, s_p$, reckoned from experience, estimated from the data in some way, or set at 0. The default in `R` is to use values obtained from the `decompose` procedure.

The forecasting equation for x_{n+k} made after the observation at time n is

$$\hat{x}_{n+k|n} = a_n + kb_n + s_{n+k-p} \qquad k \le p \tag{3.22}$$

[1] When describing the Holt-Winters procedure, the `R` help and many textbooks refer to the slope as the trend.

[2] The mean of the process is the sum of the level and the appropriate seasonal effect.

where a_n is the estimated level and b_n is the estimated slope, so $a_n + kb_n$ is the expected level at time $n+k$ and s_{n+k-p} is the exponentially weighted estimate of the seasonal effect made at time $n = k - p$. For example, for monthly data ($p = 12$), if time $n + 1$ occurs in January, then s_{n+1-12} is the exponentially weighted estimate of the seasonal effect for January made in the previous year. The forecasting equation can be used for lead times between $(m-1)p+1$ and mp, but then the most recent exponentially weighted estimate of the seasonal effect available will be $s_{n+k-(m-1)p}$.

The Holt-Winters algorithm with multiplicative seasonals is

$$\left.\begin{aligned} a_n &= \alpha\left(\frac{x_n}{s_{n-p}}\right) + (1-\alpha)(a_{n-1} + b_{n-1}) \\ b_n &= \beta(a_n - a_{n-1}) + (1-\beta)b_{n-1} \\ s_n &= \gamma\left(\frac{x_n}{a_n}\right) + (1-\gamma)s_{n-p} \end{aligned}\right\} \tag{3.23}$$

The forecasting equation for x_{n+k} made after the observation at time n becomes

$$\hat{x}_{n+k|n} = (a_n + kb_n)s_{n+k-p} \qquad\qquad k \le p \tag{3.24}$$

In R, the function `HoltWinters` can be used to estimate smoothing parameters for the Holt-Winters model by minimising the one-step-ahead prediction errors (SS1PE).

Sales of Australian wine

The data in the file `wine.dat` are monthly sales of Australian wine by category, in thousands of litres, from January 1980 until July 1995. The categories are fortified white, dry white, sweet white, red, rose, and sparkling. The sweet white wine time series is plotted in Figure 3.9, and there is a dramatic increase in sales in the second half of the 1980s followed by a reduction to a level well above the starting values. The seasonal variation looks as though it would be better modelled as multiplicative, and comparison of the SS1PE for the fitted models confirms this (Exercise 6). Here we present results for the model with multiplicative seasonals only. The Holt-Winters components and fitted values are shown in Figures 3.10 and 3.11 respectively.

```
> www <- "http://www.massey.ac.nz/~pscowper/ts/wine.dat"
> wine.dat <- read.table(wine, header = T) ; attach (wine.dat)
> sweetw.ts <- ts(sweetw, start = c(1980,1), freq = 12)
> plot(sweetw.ts, xlab= "Time (months)", ylab = "sales (1000 litres)")
> sweetw.hw <- HoltWinters (sweetw.ts, seasonal = "mult")
> sweetw.hw ; sweetw.hw$coef ; sweetw.hw$SSE

...
Smoothing parameters:
 alpha:  0.4107
 beta :  0.0001516
```

```
 gamma:  0.4695
...

> sqrt(sweetw.hw$SSE/length(sweetw))
[1] 50.04
> sd(sweetw)
[1] 121.4

> plot (sweetw.hw$fitted)
> plot (sweetw.hw)
```

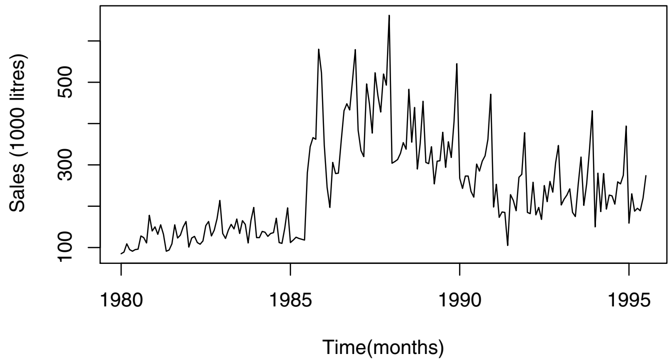

Fig. 3.9. Sales of Australian sweet white wine.

The optimum values for the smoothing parameters, based on minimising the one-step ahead prediction errors, are 0.4107, 0.0001516, and 0.4695 for α, β, and γ, respectively. It follows that the level and seasonal variation adapt rapidly whereas the trend is slow to do so. The coefficients are the estimated values of the level, slope, and multiplicative seasonals from January to December available at the latest time point ($t = n = 187$), and these are the values that will be used for predictions (Exercise 6). Finally, we have calculated the mean square one-step-ahead prediction error, which equals 50, and have compared it with the standard deviation of the original time series which is 121. The decrease is substantial, but a more testing comparison would be with the mean one-step-ahead prediction error if we forecast the next month's sales as equal to this month's sales (Exercise 6). Also, in Exercise 6 you are asked to investigate the performance of the Holt-Winters algorithm if the three smoothing parameters are all set equal to 0.2 and if the values for the parameters are optimised at each time step.

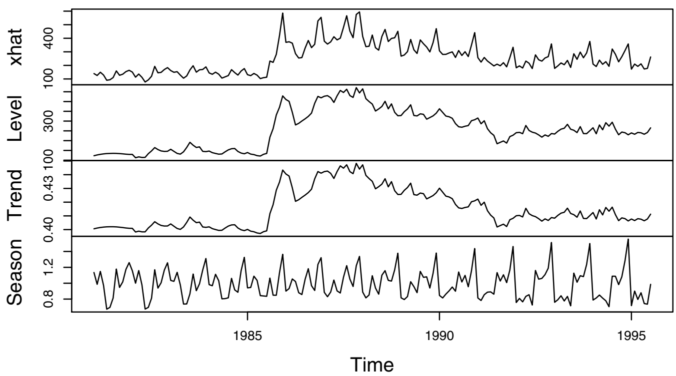

Fig. 3.10. Sales of Australian white wine: fitted values; level; slope (labelled trend); seasonal variation.

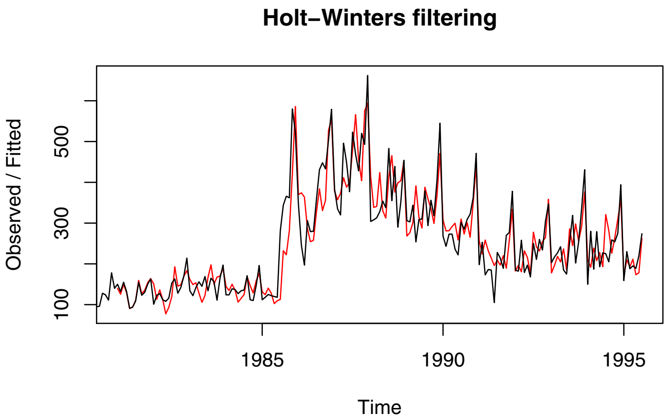

Fig. 3.11. Sales of Australian white wine and Holt-Winters fitted values.

3.4.3 Four-year-ahead forecasts for the air passenger data

The seasonal effect for the air passenger data of §1.4.1 appeared to increase with the trend, which suggests that a 'multiplicative' seasonal component be used in the Holt-Winters procedure. The Holt-Winters fit is impressive – see Figure 3.12. The `predict` function in `R` can be used with the fitted model to make forecasts into the future (Fig. 3.13).

```
> AP.hw <- HoltWinters(AP, seasonal = "mult")
> plot(AP.hw)
```

```
> AP.predict <- predict(AP.hw, n.ahead = 4 * 12)
> ts.plot(AP, AP.predict, lty = 1:2)
```

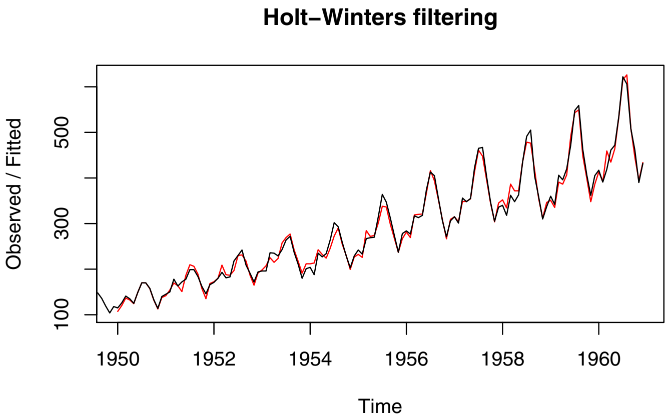

Fig. 3.12. Holt-Winters fit for air passenger data.

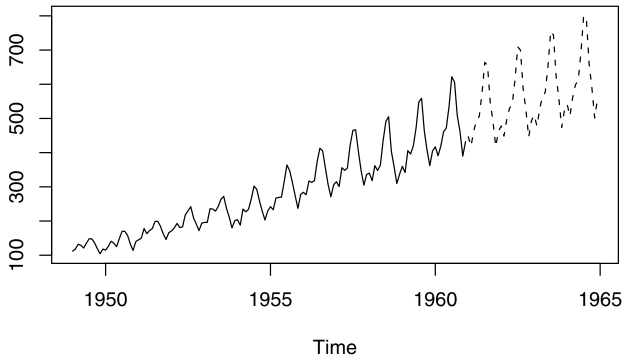

Fig. 3.13. Holt-Winters forecasts for air passenger data for 1961–1964 shown as dotted lines.

The estimates of the model parameters, which can be obtained from `AP.hw$alpha`, `AP.hw$beta`, and `AP.hw$gamma`, are $\hat{\alpha} = 0.274$, $\hat{\beta} = 0.0175$, and $\hat{\gamma} = 0.877$. It should be noted that the extrapolated forecasts are based entirely on the trends in the period during which the model was fitted and would be a sensible prediction assuming these trends continue. Whilst the extrapolation in Figure 3.12 looks visually appropriate, unforeseen events could lead to completely different future values than those shown here.

3.5 Summary of commands used in examples

`nls`	non-linear least squares fit
`HoltWinters`	estimates the parameters of the Holt-Winters or exponential smoothing model
`predict`	forecasts future values
`ts.union`	create the union of two series
`coef`	extracts the coefficients of a fitted model

3.6 Exercises

1. a) Describe the association and calculate the ccf between x and y for k equal to 1, 10, and 100.
```
> w <- 1:100
> x <- w + k * rnorm(100)
> y <- w + k * rnorm(100)
> ccf(x, y)
```
   b) Describe the association between x and y, and calculate the ccf.
```
> Time <- 1:370
> x <- sin(2 * pi * Time / 37)
> y <- sin(2 * pi * (Time + 4) / 37)
```
   Investigate the effect of adding independent random variation to x and y.

2. a) Let $f(t)$ be the density of time T to purchase for a randomly selected purchaser. Show that

$$P(\text{Buys in next time increment } \delta t |\text{ no purchase by time } t) = h(t)\delta t$$

   b) The survivor function $S(t)$ is defined as the complement of the cdf

$$S(t) = 1 - F(t)$$

   Show that $S(t) = \exp(-\int_0^t h(u)\,du)$ and $E[T] = \int_0^\infty S(t)\,dt$.

   c) Explain how Equation (3.6) is the discrete form of Equation (3.9).

3. a) Verify that the solution of Equation (3.8), with $h(t)$ given by Equation (3.9), for $F(t)$ is Equation (3.10).

b) The logistic distribution has the cdf: $F(t) = \{1 + \exp(-(t-\mu)/b)\}^{-1}$, with mean μ and standard deviation $b\pi/\sqrt{3}$. Plot the cdf of the logistic distribution with a mean 0 and standard deviation 1 against the cdf of the standard normal distribution.
c) Show that the time to peak of the Bass curve is given by Equation (3.13). What does this reduce to for the exponential and logistic distributions?

4. The *Independent* on July 11, 2008 reported the launch of Apple's iPhone. A Deutsche Bank analyst predicted Apple would sell 10.5 million units during the year. The company was reported to have a target of 10 million units worldwide for 2008. Initial demand is predicted to exceed supply. Carphone Warehouse reportedly sold their online allocations within hours and expect to sell out at most of their UK shops. The report stated that there were 60,000 applications for 500 iPhones on the Hutchison Telecommunications website in Hong Kong.
   a) Why is a Bass model without replacement or multiple purchases likely to be realistic for this product?
   b) Suggest plausible values for the parameters p, q, and m for the model in (a), and give a likely range for these parameters. How does the shape of the cumulative sales curve vary with the parameter values?
   c) How could you allow for high initial sales with the Bass model?

5. a) Write the sum of n terms in a geometric progression with a first term a and a common ratio r as

$$S_n = a + ar + ar^2 + \ldots + ar^{n-1}$$

   Subtract rS_n from S_n and rearrange to obtain the formula for the sum of n terms:

$$S_n = \frac{a(1-r^n)}{1-r}$$

   b) Under what conditions does the sum of n terms of a geometric progression tend to a finite sum as n tends to infinity? What is this sum?
   c) Obtain an expression for the sum of the weights in an EWMA if we specify $a_1 = x_1$ in Equation (3.15).
   d) Suppose x_t happens to be a sequence of independent variables with a constant mean and a constant variance σ^2. What is the variance of a_t if we specify $a_1 = x_1$ in Equation (3.15)?

6. Refer to the sweet white wine sales (§3.4.2).
   a) Use the `HoltWinters` procedure with α, β and γ set to 0.2 and compare the SS1PE with the minimum obtained with `R`.
   b) Use the `HoltWinters` procedure on the logarithms of sales and compare SS1PE with that obtained using sales.

c) What is the SS1PE if you predict next month's sales will equal this month's sales?
d) This is rather harder: What is the SS1PE if you find the optimum α, β and γ from the data available at each time step before making the one-step-ahead prediction?

7. Continue the following exploratory time series analysis using the global temperature series from §1.4.5.
   a) Produce a time plot of the data. Plot the aggregated annual mean series and a boxplot that summarises the observed values for each season, and comment on the plots.
   b) Decompose the series into the components trend, seasonal effect, and residuals, and plot the decomposed series. Produce a plot of the trend with a superimposed seasonal effect.
   c) Plot the correlogram of the residuals from question 7b. Comment on the plot, explaining any 'significant' correlations at significant lags.
   d) Fit an appropriate Holt-Winters model to the monthly data. Explain why you chose that particular Holt-Winters model, and give the parameter estimates.
   e) Using the fitted model, forecast values for the years 2005–2010. Add these forecasts to a time plot of the original series. Under what circumstances would these forecasts be valid? What comments of caution would you make to an economist or politician who wanted to use these forecasts to make statements about the potential impact of global warming on the world economy?

8. A cumulative sum plot is useful for monitoring changes in the mean of a process. If we have a time series composed of observations x_t at times t with a target value of τ, the CUSUM chart is a plot of the cumulative sums of the deviations from target, cs_t, against t. The formula for cs_t at time t is

$$cs_t = \sum_{i=1}^{t} (x_i - \tau)$$

   The `R` function `cumsum` calculates a cumulative sum. Plot the CUSUM for the motoring organisation complaints with a target of 18.

9. Using the motor organisation complaints series, refit the exponential smoothing model with weights $\alpha = 0.01$ and $\alpha = 0.99$. In each case, extract the last residual from the fitted model and verify that the last residual satisfies Equation (3.19). Redraw Figure 3.8 using the new values of α, and comment on the plots, explaining the main differences.