Chapter 8: Forecasting

8.0 Introduction

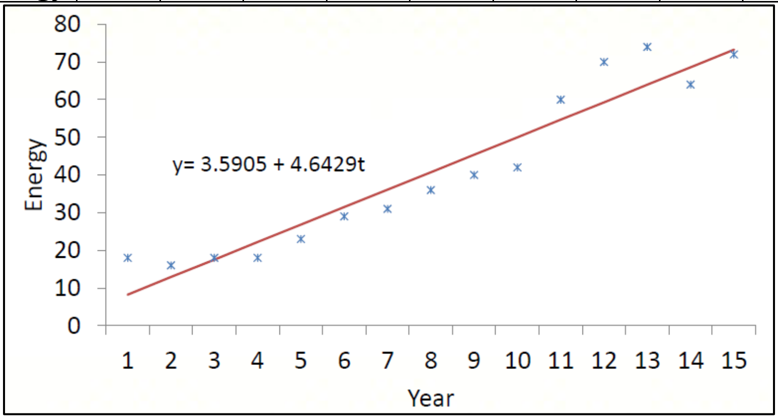
In this chapter we are again interested in construction models and using them for forecasting, the prediction solely on the past behavior of that variable.

8.1 Extrapolation of Trend Curves

For long-term forecasting of non-seasonal data, it is often useful to fit a **trend curve** (or **growth curve**) to successive values and then extrapolate. This approach is most often used when the data are yearly totals, and hence clearly non-seasonal. A variety of curves may be tried including polynomial, exponential, logistic and Gompertz curves. When the data are annual totals, at least 7 to 10 years of historical data are required to fit such curves. The method is worth considering for short annual series where fitting a complicated model to past data is unlikely to be worthwhile. Although primarily intended for long-term forecasting, it is inadvisable to make forecasts for a longer period ahead than about half the number of past years for which data are available.

Example: Linear model on the number of energy consumption (in quardrillion BTUs)

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Energy	18	16	18	18	23	29	31	36	40	42	60	70	74	64	72



Find the forecast value that made in the year 2020.

8.2 Averaging Methods

If a time series is generated by a constant process subject to random error, then the mean is a useful statistic and can be used as a forecast for the next periods.

<u>Simple Moving Averages</u>

A moving average forecast of order k, or $MA(k)$, is given by

$$F_{t+1} = \frac{1}{k}\sum_{i=t-k+1}^{t} Y_i$$

The term "moving average" is used because as one new observation becomes available, a new average can be computed by dropping the oldest observation and including the newest one. This moving average will then be the forecast for the next period.

The moving average of order k has the following characteristics:

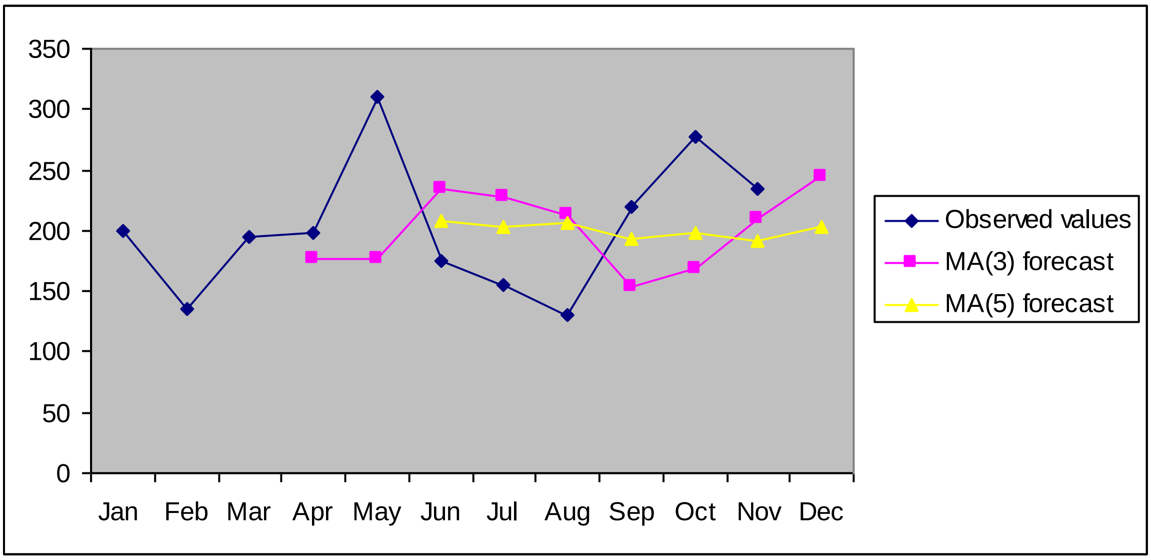
1. It deals only with the latest k period of known data (includes the most recent observation).
2. The number of data points in each average does not change as time goes on.

But it also has the following disadvantages:

1. It requires more storage because all of the k latest observations must be store, not just the average.
2. It cannot handle trend and seasonality very well.

Month	Time, t	Observed Values	MA(3)	MA(5)
Jan	1	200		
Feb	2	135		
Mar	3	195		
Apr	4	197.5		
May	5	310		
Jun	6	175		
Jul	7	155		
Aug	8	130	213.33	206.50
Sep	9	220	153.33	193.50
Oct	10	277.5	168.33	198.00
Nov	11	235	209.17	191.50
Dec	12		244.17	203.50

The $MA(3)$ values in column 4 are based on the values for the previous three months. For example, the forecast for April (the fourth month) is taken to be the average of Jan, Feb, and Mar. The Dec's $MA(3)$ forecast of 244.17 is the average for Sept, Oct, and Nov.



<u>Note:</u>

1. The use of a small value for k will allow the moving average to follow the pattern, but these MA forecast will nevertheless trail the pattern, lagging behind by one or more periods.
2. The more observations included in the moving average, the greater the smoothing effect.

8.3 Simple Exponential Smoothing (Weighted Moving Average)

With simple moving average forecasts, the mean of the past k observations was used as a forecast. This implies equal weights (equal to $1/k$) for all k data points.

The simple exponential smoothing method is used for forecasting a time series when there is no trend or seasonal pattern, but the mean of the time series is slowly changing over time. The simple exponential smoothing method gives the most recent observation the most weight. Older observations are given successively smaller. This is an obvious extension to the moving average by forecasting using *weighted moving average*.

Suppose that the time series $Y_t = \{y_1, y_2, ..., y_n\}$ has mean that may be slowly changing over time but has no trend or seasonal pattern. Then the forecast of the time series in time period t is given by the smoothing equation

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

where α is a smoothing constant between 0 and 1, and F_t is the forecast of the time series in time period t. $(Y_t - F_t)$ is the forecast error found when the observations Y_t becomes available.

Note:

1. It can be seen that the new forecast is simply the old forecast plus an adjustment for the error that occurred in the last forecast.
2. When α has a value close to 1, the new forecast will include a substantial adjustment for the error in the previous forecast. Conversely, when α is close to 0, new forecast will include very little adjustment.

Note that F_{t+1} can be rewritten as

$$\begin{aligned} F_{t+1} &= \alpha Y_t + (1-\alpha)F_t \\ &= \alpha Y_t + (1-\alpha)[\alpha Y_{t-1} + (1-\alpha)F_{t-1}] \\ &= \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + (1-\alpha)^2 F_{t-1} \\ &= \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \alpha(1-\alpha)^3 Y_{t-3} \\ &\quad + \alpha(1-\alpha)^4 Y_{t-4} + \cdots + \alpha(1-\alpha)^{t-1} Y_1 + (1-\alpha)^t F_1 \end{aligned}$$

1. The forecast (F_{t+1}) is based on weighting the most recent observations (Y_t) with a weight value (α) and weighting the most recent forecast (F_t) with a weight of $(1-\alpha)^t$
2. The weights for all past data sum approximately to one.
3. As the observations get older, the weights decreasing exponentially, hence the name exponential smoothing.

It was assumed that the forecast horizon was just one period ahead. For longer forecasts, it is assumed that the forecast function is 'flat', that is, a point forecast made in time period t is

$$F_{t+h} = F_{t+1} \ (h = 2, 3, ...)$$

If $h=1$, then a $(1-\alpha)100\%$ prediction interval computed in time period t for F_{t+1} is computed by

$$F_{t+1} \pm z_{\alpha/2}\sqrt{MS_E}$$

If $h=2$, then a $(1-\alpha)100\%$ prediction interval computed in time period t for F_{t+2} is computed by

$$F_{t+1} \pm z_{\alpha/2}\sqrt{MS_E(1+\alpha^2)}$$

In general, for any h, a $(1-\alpha)100\%$ prediction interval computed in time period t for F_{t+h} is computed by

$$F_{t+1} \pm z_{\alpha/2}\sqrt{MS_E(1+(h-1)\alpha^2)}$$

where $MS_E = \dfrac{SS_E}{n} = \dfrac{\sum_{t=1}^{n}(y_t - F_t)^2}{n}$.

Example: Simple exponential smoothing is applied to the series

t	1	2	3	4	5
y_t	1000	900	990	909	982

You are given the smoothing coefficient is 0.9.

a) Find the smoothed value F_5 if the started smoothing value is $y_1 = 1000$.

Example: Electric can opener shipments

The following data obtained in a time-series forecasting analysis

$\alpha = 0.1,\ SS_E = 30154.65$

Month	Time period, t	Actual shipments, y_t	Smoothed estimate, F_t	Forecast Error	Squared Forecast Error
Jan	1	200			
Feb	2	135			
Mar	3	195			
Apr	4	197.5	193.65	3.85	14.8225
May	5	310	194.04	115.96	13446.7
Jun	6	175	205.63	-30.63	938.197
Jul	7	155	202.57	-47.57	2262.9
Aug	8	130	197.81	-67.81	4598.2
Sep	9	220	191.03	28.97	839.261
Oct	10	277.5	193.93	83.57	6983.94
Nov	11	235	202.28	32.72	1070.6
Dec	12				

Find

a) The point forecast in period 12 (December) F_{12}.

b) A 95% prediction interval made in month 11(November) for F_{12}.

c) A 95% prediction interval made in month 11(November) for F_{13}.

R-Codes:

```
# To enter the data
s <- c(200,135,195,197.5,310,175,155,130,220,277.5,235)
# Estimating the level of time series using simple exponential smoothing
es1 <- HoltWinters(s, alpha=.1, beta=FALSE, gamma=FALSE)
es1
```

Output:

```
Holt-Winters exponential smoothing without trend and without seasonal
component.
Call:
HoltWinters(x = s, alpha = 0.1, beta = 0, gamma = 0)

Smoothing parameters:
alpha:  0.1
beta :  FALSE
gamma:  FALSE

Coefficients:
      [,1]
a 205.5561
```

```
# Computes predictions and prediction intervals
predict(es1, n.ahead=3, prediction.interval = T, level=.95)
Time Series:
Start = 12
End = 14
Frequency = 1
        fit      upr      lwr
12 205.5561 326.1547 84.95752
13 205.5561 326.7562 84.35603
14 205.5561 327.3548 83.75751
```

Note:

1. In general, note that the smoothing equation

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

Substituting recursively for F_{t-1}, …, F_2 and F_1, we obtain

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots + \alpha(1-\alpha)^{t-1}Y_1 + (1-\alpha)^t F_1$$

The coefficient measuring the contribution of the observations Y_t, Y_{t-1}, …, Y_1 are α, $\alpha(1-\alpha)$, …, $\alpha(1-\alpha)^{t-1}$, respectively, and they are decreasing exponentially with age. For this reason we refer this procedure as *simple exponential smoothing*.

2. One point of concern relates to the initializing phase of exponential smoothing. For example, to get the forecasting system started we need F_1 because $F_2 = \alpha y_1 + (1-\alpha)F_1$. Since the value of F_1 is not known, we can use the first observed value (y_1) as the first forecast ($F_1 = y_1$) and then proceed using the smoothing equation.

   Another possibility would be to average the first four or five values in the data set and use this as the initial forecast.

3. The weight of α can be chosen by minimizing the value of MS_E (through trial and error) or some other criterions.

4. Note that the last term is $(1-\alpha)^t F_1$. So the initial forecast F_1 plays a role in all subsequent forecasts. But the weight attached to F_1 is $(1-\alpha)^t$ which is usually small. When a small value of α is chosen, the initial forecast plays a more prominent role

than when a larger α is used. Also, when more data are available t is larger and so the weight attached to F_1 is smaller.

5. If the smoothing parameter α is not close to zero, the influence of the initialization process rapidly becomes of less significance as time goes by. However, if α is close to zero, the initialization process can play a significant role for many time periods ahead.

8.4 <u>Additive Holt-Winters Method</u>

Suppose that the time series $Y_t = \{y_1, y_2, ..., y_n\}$ exhibits a linear trend locally and has a seasonal pattern with constant (additive) seasonal variation and that the level, growth rate, and seasonal pattern may be changing. Then the estimate L_t for the level, the estimate b_t for the growth rate, and the estimate S_t for the seasonal factor of the time series in time period t is given by the smoothing equations

Level: $L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$

Trend: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$

Seasonal: $S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}$

$$F_{t+m} = L_t + b_t m + S_{t-s+m}$$

where α, β and γ are smoothing constants between 0 and 1, and L_{t-1} and b_{t-1} are estimate at time $t-1$ for the level and growth rate, respectively, and S_{t-s} is the estimate in time period $t-s$ for the seasonal factor. Here s denotes the number of seasons in a year (s = 12 for monthly data, and s =4 for quarterly data).

<u>Note:</u>
L_t is a smoothed value of the series that does not include seasonality (or the data have been seasonally adjusted), while Y_t, on the other hand, do contain seasonality.

A point forecast made in time period t is

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \qquad (m = 1, 2, 3, ...)$$

A $(1-\alpha)100\%$ prediction interval computed in time period t for F_{t+m} is computed by

$$F_{t+m} \pm z_{\alpha/2}\sqrt{c_m MS_E}$$

If $m = 1$, then $c_1 = 1$.

If $2 \le m \le s$, then $c_m = \left(1 + \sum_{j=1}^{m-1} \alpha^2 (1 + j\beta)^2\right)$

If $m > s$, then $c_m = 1 + \sum_{j=1}^{m-1} \left(\alpha(1 + j\beta) + d_{j,s}(1-\alpha)\gamma\right)^2$

where $d_{j,s} = 1$ if j is an integer multiple of s and 0 otherwise

and $MS_E = \dfrac{SS_E}{n-3} = \dfrac{\sum_{t=1}^{n} \left(y_t - (L_{t-1} + b_{t-1} + S_{t-s+1})\right)^2}{n-3}$.

Initialization

To determine initial estimates of the seasonal indices we need to use at least one complete season's data (i.e. s period). Therefore, we initialize trend and level at period s.

1. $L_s = \frac{1}{s}\left(Y_1 + Y_2 + \cdots + Y_s\right).$

2. $b_s = \frac{1}{s}\left[\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \cdots + \frac{Y_{s+s} - Y_s}{s}\right]$

3. $S_1 = Y_1 - L_s,\ S_2 = Y_2 - L_s,\ \ldots,\ S_s = Y_s - L_s$

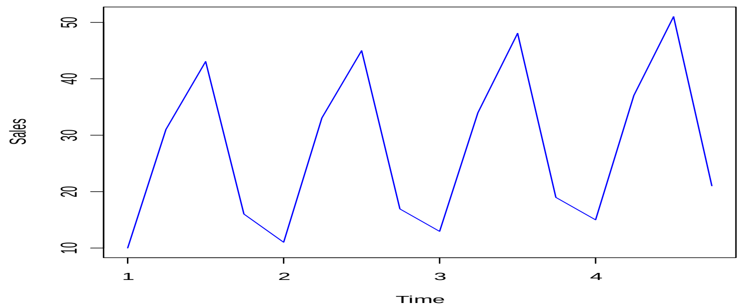
Several other methods for initializing are also available.

Example:

The quarterly sales of the TRK-50 mountain bike for the previous four years by a bicycle shop are given below:

Quarterly Sales of the TRK-50 Mountain Bike

	Year			
Quarter	1	2	3	4
1	10	11	13	15
2	31	33	34	37
3	43	45	48	51
4	16	17	19	21



The time plot above suggests that the mountain bike sales display a linear demand and a *constant* (additive) seasonal variation. Thus we apply the additive Holt-winters method to these data in order to find forecasts of future mountain bike sales.

Year	Time, t	Y_t	Level, L_t	Growth Rate, b_t	Seasonal Factor, S_t	Forecast, F_t
1	1	10			-15	
	2	31			6	
	3	43			18	
	4	16	25	0.375	-9	
2	5	11	25.5	0.3875	-14.95	10.9375
	6	33	26.11	0.4098	6.089	32.6088
	7	45	26.6158	0.4194	18.0384	45.0736
	8	17	26.8281	0.3987	-9.0828	18.1440
3	9	13	27.3714	0.4131	-14.8921	12.8924
	10	34	27.8098	0.4156	6.0991	34.3246
	11	48	28.5727	0.4504	18.1773	47.2004
	12	19	28.8350	0.4316	-9.1580	20.1085

4	13	15	29.3917	0.4441	-14.8421	14.9937
	14	37	30.0488	0.4654	6.1843	36.6985
	15	51	30.9759	0.5115	18.3620	49.8494
	16	21	31.2215	0.4850	-9.2644	22.4421

R-Codes:

```
Bike <- c(10,31,43,16,11,33,45,17,13,34,48,19,15,37,51,21)
Yt <- ts(Bike, frequency=4)      #function “ts” is used to create time series
object
Yt

R-output:
  Qtr1 Qtr2 Qtr3 Qtr4
1   10   31   43   16
2   11   33   45   17
3   13   34   48   19
4   15   37   51   21

plot(Yt, xlab="Time", ylab="Sales", col="Blue")
m<- HoltWinters(Yt, alpha=0.2, beta=0.1, gamma=0.1, seasonal="additive")
m
MSE <- m$"SSE"/(NROW(Bike)-3)
MSE
predict(object=m, n.ahead=3,prediction.interval=T, level=.95)

#Holt-Winters exponential smoothing with trend and additive seasonal
component.

Call:
 HoltWinters(x = Yt, alpha = 0.2, beta = 0.1, gamma = 0.1, seasonal =
"additive")

Smoothing parameters:
 alpha:  0.2
 beta :  0.1
 gamma:  0.1

Coefficients:
          [,1]
a   31.3827876
b    0.5282335
s1 -14.6456318          #1st quarter
s2   6.8377413          #2nd quarter
s3  18.8141442          #3rd quarter
s4  -9.9464101          #4th quarter
> MSE <- m$"SSE"/(NROW(Bike)-3)
> MSE
[1] 0.5846681

> predict(object=m, n.ahead=2,prediction.interval=T, level=.95)
          fit      upr      lwr
5 Q1 17.26539 18.67046 15.86032
5 Q2 39.27700 40.71566 37.83833
5 Q3 51.78163 53.25929 50.30397
```

Find

a) The point forecast for period 17, F_{17} made in period 16.

b) A 95% prediction interval for F_{17} made in period 16.

8.5 Multiplicative Holt-Winters Method

Suppose that the time series $Y_t = \{y_1, y_2, ..., y_n\}$ exhibits a linear trend locally and has a seasonal pattern with increasing (multiplicative) seasonal variation and that the level, growth rate, and seasonal pattern may be changing. Then the estimate L_t for the level, the estimate b_t for the growth rate, and the estimate S_t for the seasonal factor of the time series in time period t is given by the smoothing equations.

Level: $L_t = \alpha(Y_t / S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$

Trend: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$

Seasonal: $S_t = \gamma(Y_t / L_t) + (1-\gamma)S_{t-s}$

$$F_{t+m} = (L_t + b_t m)S_{t-s+m}$$

where α, β and γ are smoothing constants between 0 and 1, and L_{t-1} and b_{t-1} are estimate at time $t-1$ for the level and growth rate, respectively, and S_{t-s} is the estimate in time period $t-s$ for the seasonal factor. Here s denotes the number of seasons in a year (s = 12 for monthly data, and s =4 for quarterly data).

A point forecast made in time period t is

$$F_{t+m} = (L_t + b_t m)S_{t-s+m} \qquad (m = 1, 2, 3, ...)$$

Initialization

1. $L_s = \frac{1}{s}\left(Y_1 + Y_2 + \cdots + Y_s\right)$.

2. $b_s = \frac{1}{s}\left[\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \cdots + \frac{Y_{s+s} - Y_s}{s}\right]$

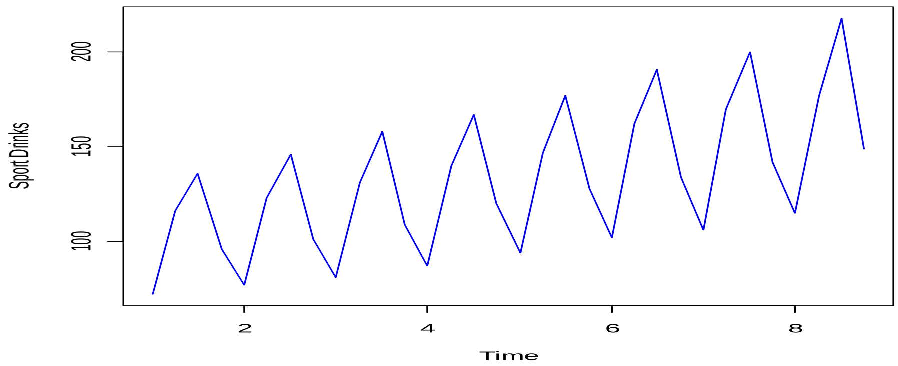
3. $S_1 = \frac{Y_1}{L_s},\ S_2 = \frac{Y_2}{L_s},\ \ldots,\ S_s = \frac{Y_s}{L_s}$

Several other methods for initializing are also available.

Example:
The quarterly sales of Tiger sport Drink for the last eight years are given below:

Quarterly Sales of Tiger Sport Drink (1000sCases)

	Year							
Quarter	1	2	3	4	5	6	7	8
1	72	77	81	87	94	102	106	115
2	116	123	131	140	147	162	170	177
3	136	146	158	167	177	191	200	218
4	96	101	109	120	128	134	142	149



The time plot above indicates that there is a linear increase in sales over the eight-year period and that the *seasonal pattern is increasing as the level of time series increases*. This pattern suggests that multiplicative Holt-Winters might be employed to forecast future sales.

Year	Time, t	Y_t	Level, L_t	Growth Rate, b_t	Seasonal Factor, S_t	Forecast, F_t
1	1	72			0.6857	
	2	116			1.1048	
	3	136			1.2952	
	4	96	105.0000	1.6875	0.9143	
2	5	77	107.8083	1.7996	0.6886	75.4723
	6	123	109.9536	1.8341	1.1062	123.6541
	7	146	111.9743	1.8528	1.2961	147.5315
	8	101	113.1554	1.7856	0.9121	104.8395
3	9	81	115.4800	1.8395	0.6899	80.9330
	10	131	117.5414	1.8617	1.1070	132.1776
	11	158	119.9033	1.9117	1.2983	158.1481
	12	109	121.3525	1.8655	0.9107	112.2176
4	13	87	123.7972	1.9234	0.6911	86.8909
	14	140	125.8704	1.9384	1.1075	141.5499
	15	167	127.9737	1.9549	1.2989	168.7685
	16	120	130.2955	1.9916	0.9118	120.6128
5	17	94	133.0310	2.0660	0.6927	93.5801
	18	147	134.6235	2.0186	1.1060	151.1201
	19	177	136.5669	2.0111	1.2986	179.9638
	20	128	138.9402	2.0473	0.9127	128.6795
6	21	102	142.2405	2.1726	0.6951	100.3858
	22	162	144.8264	2.2139	1.1072	162.8057
	23	191	147.0475	2.2147	1.2987	193.8426
	24	134	148.7731	2.1658	0.9115	137.5810
7	25	106	151.2491	2.1968	0.6957	106.7521
	26	170	153.4643	2.1986	1.1073	172.3610
	27	200	155.3310	2.1654	1.2976	204.3614
	28	142	157.1545	2.1312	0.9107	145.0628
8	29	115	160.4889	2.2515	0.6978	113.5579
	30	177	162.1628	2.1938	1.1057	181.7280
	31	218	165.0868	2.2668	1.2999	217.5358
	32	149	166.6047	2.1919	0.9091	153.4480

R-Codes:

```
Y <- c(72, 116, 136, 96, 77, 123, 146, 101, 81, 131, 158, 109, 87, 140, 167,
120, 94, 147, 177, 128, 102, 162, 191, 134, 106, 170, 200, 142, 115, 177,
218, 149)
Yt <- ts(Y, frequency=4)
plot(Yt, xlab="Time", ylab="Sport Drinks", col="Blue")
m<- HoltWinters(Yt, alpha=0.2, beta=0.1, gamma=0.1, seasonal="mult")
m
```

```
predict(object=m, n.ahead=2,prediction.interval=T, level=.95)

#Holt-Winters exponential smoothing with trend and multiplicative seasonal
component.
Call:
HoltWinters(x = Yt, alpha = 0.2, beta = 0.1, gamma = 0.1, seasonal = "mult")
Smoothing parameters:
 alpha: 0.2
 beta : 0.1
 gamma: 0.1
Coefficients:
          [,1]
a  166.9252797
b    2.2379389
s1   0.7080902          #1st quarter
s2   1.1083459          #2nd quarter
s3   1.2971180          #3rd quarter
s4   0.8994015          #4th quarter
> predict(object=m, n.ahead=4,prediction.interval=T, level=.95)
          fit      upr      lwr
9 Q1 119.7828 120.9675 118.5981
9 Q2 189.9718 191.8496 188.0939
9 Q3 225.2304 227.6902 222.7706
9 Q4 158.1841 159.9937 156.3745
```

Find the forecast values, F_{33}, F_{34} and F_{35} that made in period 32

8.6 The Box-Jenkins procedure

This section gives an outline of the forecasting procedure, based on autoregressive integrated moving average (ARIMA) models, which is usually known as the Box-Jenkins approach. Box and Jenkins showed how the use of differencing can extend ARMA models to ARIMA models and hence cope with non-stationary series. It also showed how to incorporate seasonal terms into seasonal ARIMA (SARIMA) models.

Box and Jenkins provides a general strategy for time-series forecasting, the main stages in setting up a Box-Jenkins forecasting model are as follow:

1. **Model identification**. Examine the data to see which member of the class of ARIMA processes appears to be most appropriate.
2. **Estimation**. Estimate the parameters of the chosen model.
3. **Diagnostic checking**. Examine the residuals from the fitted model to see if it is adequate.
4. **Consideration of alternative models if necessary**. If the first model appears to be inadequate for some reason, then alternative ARIMA models may be tried until a satisfactory model is found. When such a model has been found, it is usually relatively straightforward to calculate forecasts as conditional expectations.

8.6.1 Identification

Steps to identify which model is appropriate for a given set of data:

1. Plot the data. Identify any unusual observations. Decide if a transformation is necessary to stabilize the variance. If necessary, transform the data to achieve stationarity in the variance.
2. Consider if the data appear stationary from the time plot and the ACF and $PACF$. If the time plot shows the data scattered horizontally around a constant mean, or equivalently, the ACF and $PACF$ drop to or near zero quickly, it indicates that the data are stationary. Otherwise, non-stationarity is implied.
3. When the data appear non-stationary, they can be made stationary by differencing. For non-seasonal data, take first differences of the data. For seasonal data, take seasonal differences of the data. If they are still non-stationary, take the first differences of the differenced data.
4. When stationary has been achieved, examine the autocorrelation to see if any pattern remains. There are three possibilities to consider:
    a) Seasonality- Autocorrelation and/or partial autocorrelations at the seasonal lags are large and significantly different from zero.
    b) If there are no significant autocorrelations after lag q, a $MA(q)$ model may be appropriate. If there are no significant partial autocorrelations after lag p, an $AR(p)$ model may be appropriate.
    c) If there is no clear, MA, or AR model suggested, a mixture model may be necessary.

Expected patterns in the ACF and $PACF$ for simple AR and MA models

Process	ACF	$PACF$
$AR(1)$	Exponential decay: on positive side if $\phi_1 > 0$ and alternating in sign starting on negative side if $\phi_1 < 0$	Spike at lag 1, then cuts off to zero: spike positive if $\phi_1 > 0$, negative if $\phi_1 < 0$.
$AR(p)$	Exponential decay or damped sine-wave. The exact pattern depends on the signs and sizes of $\phi_1, \phi_2 \cdots, \phi_p$.	Spikes at lags 1 to *p*, then cuts off to zero.
$MA(1)$	Spike at lag 1then cuts off to zero: spike positive if $\theta_1 < 0$, negative if $\theta_1 > 0$.	Exponential decay: on negative side if $\theta_1 > 0$ and alternating in sign starting on positive side if $\theta_1 < 0$
$MA(q)$	Spikes at lags 1 to *q*, then cuts off to zero.	Exponential decay or damped sine-wave. The exact pattern depends on the signs and sizes of $\theta_1, \theta_2 \cdots, \theta_q$.

Note:

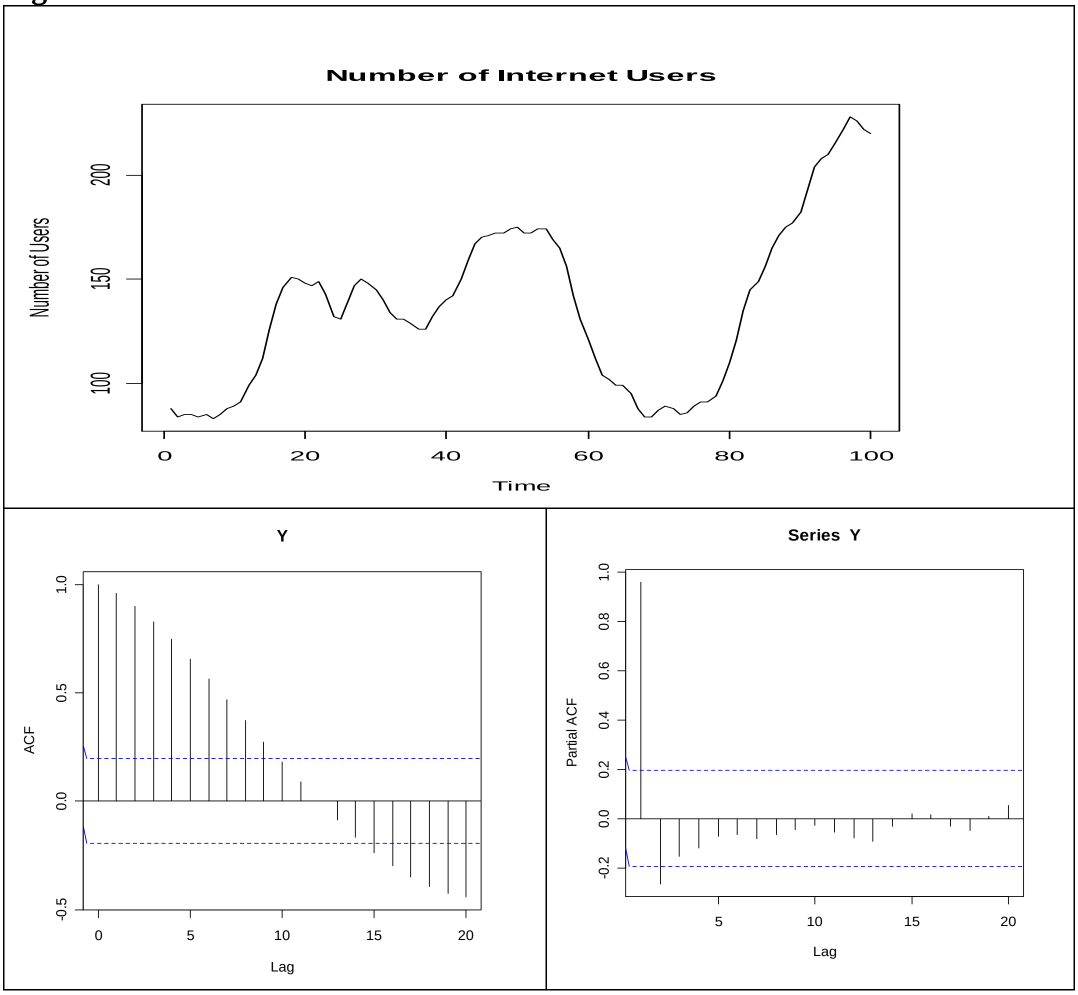
ARIMA	The general ARIMA(*p*,*d*,*q*) models yields a tremendous variety of patterns in the ACF and $PACF$, so that it is not practical to state rules for identifying general $ARIMA$ models. However, the simpler $AR(p)$ and $MA(q)$ models do provide some identifying features that can help a forecaster zero in on a particular $ARIMA$ model identification. It is possible to obtain several different models that might yield almost the same quality forecasts.

ARIMA $(p, d, q)(P,D,Q)_s$	The seasonal part of an *AR* or *MA* model will be seen in the seasonal lags of the *PACF* and *ACF*. For example, the seasonal *MA* model *ARIMA* $(0, 0, 0)(0,0,1)_{12}$ will show a spike at lag 12 in the *ACF* but no other significant spikes. The *PACF* will show exponential decay in the seasonal lags, i.e., at lags 12, 24, 36,…. Similarly an *ARIMA* $(0, 0, 0)(1,0,0)_{12}$ (a seasonal *AR* model) will show exponential decay in the seasonal lags of the *ACF*, and a single significant spike at lag 12 in the *PACF*.

<u>Example:</u> A non-seasonal time series (Internet user)

Internet user dataset contains the number of users logged onto an Internet server each minute over a 100-minute period.

Figure 8.1: Plot of data from InternetUser.txt



R-Codes:

```
Y <- c(88, 84, 85, 85, 84, 85, 83, 85, 88, 89, 91, 99,
       104, 112, 126, 138, 146, 151, 150, 148, 147, 149, 143, 132, 131,
       139, 147, 150, 148, 145, 140, 134, 131, 131, 129, 126, 126, 132,
       137, 140, 142, 150, 159, 167, 170, 171, 172, 172, 174, 175, 172,
       172, 174, 174, 169, 165, 156, 142, 131, 121, 112, 104, 102, 99,
       99, 95, 88, 84, 84, 87, 89, 88, 85, 86, 89, 91, 91, 94, 101, 110,
       121, 135, 145, 149, 156, 165, 171, 175, 177, 182, 193, 204, 208,
       210, 215, 222, 228, 226, 222, 220)
Y <- ts(Y)
plot(Y, ylab="Number of Users", main="Number of Internet Users")
```

```
acf(Y)
pacf(Y)
```

Figure 8.1 shows an initial analysis of the data. The autocorrelation plot gives indications of non-stationary and the data plot make this clear too. The first partial autocorrelation is very dominant and close to 1 – also showing the non-stationary.

So, we take the first differences of the data and reanalyze.

Figure 8.2: Plot of the differences of the data from Internet user

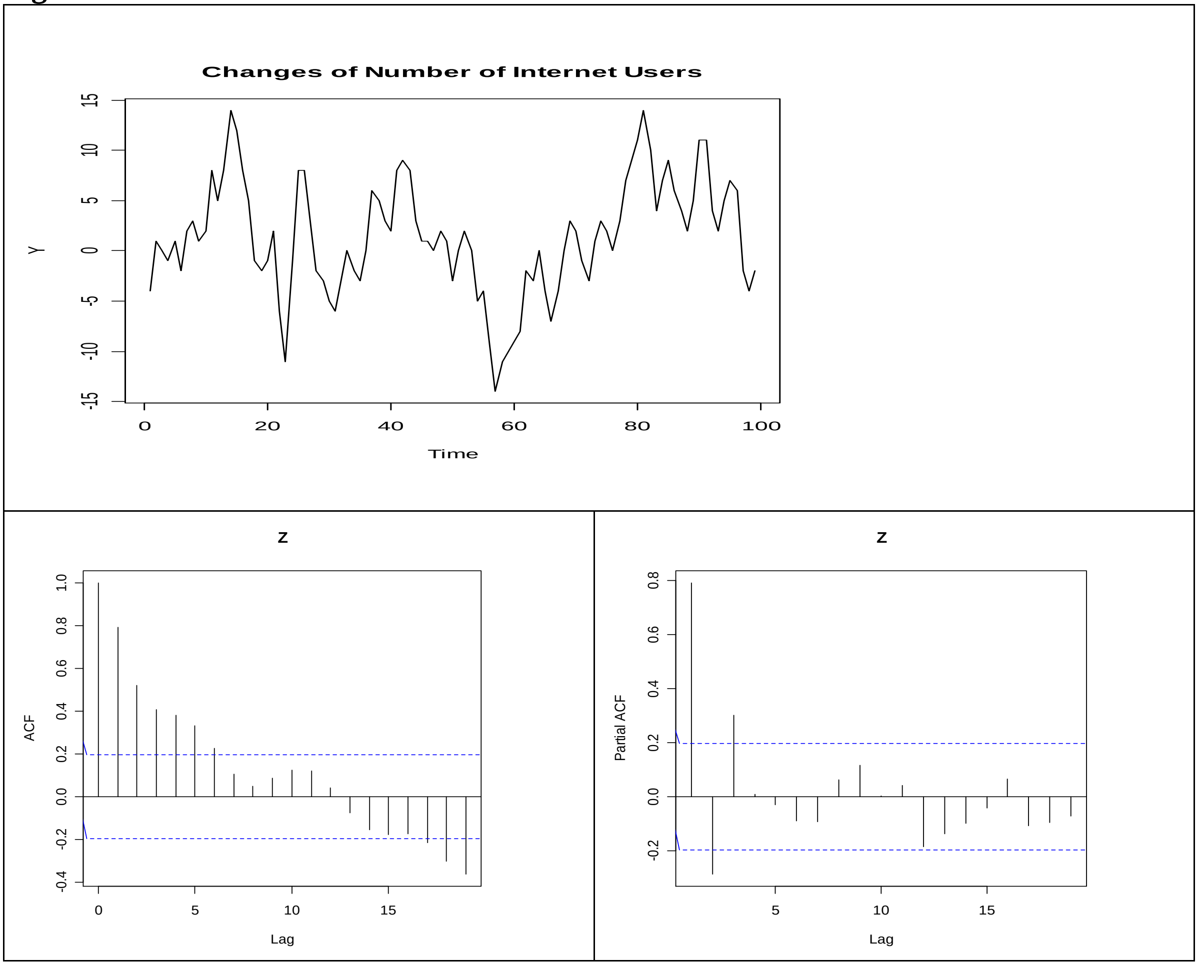


Figure 8.2 shows that the autocorrelations is a mixture of the exponential decay and sine-wave pattern and there are three significant partial autocorrelation. This suggests an $AR(3)$ model is operating.

So, for the original series, we have identified a $ARIMA(3,1,0)$ model:

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)(1-B)Y_t=\varepsilon_t$$

In terms of the Box-Jenkins stages, the identification of a tentative model has been completed.

<u>Example:</u> A seasonal time series
Writing dataset contains monthly sales for printing and writing paper.

Figure 8.3

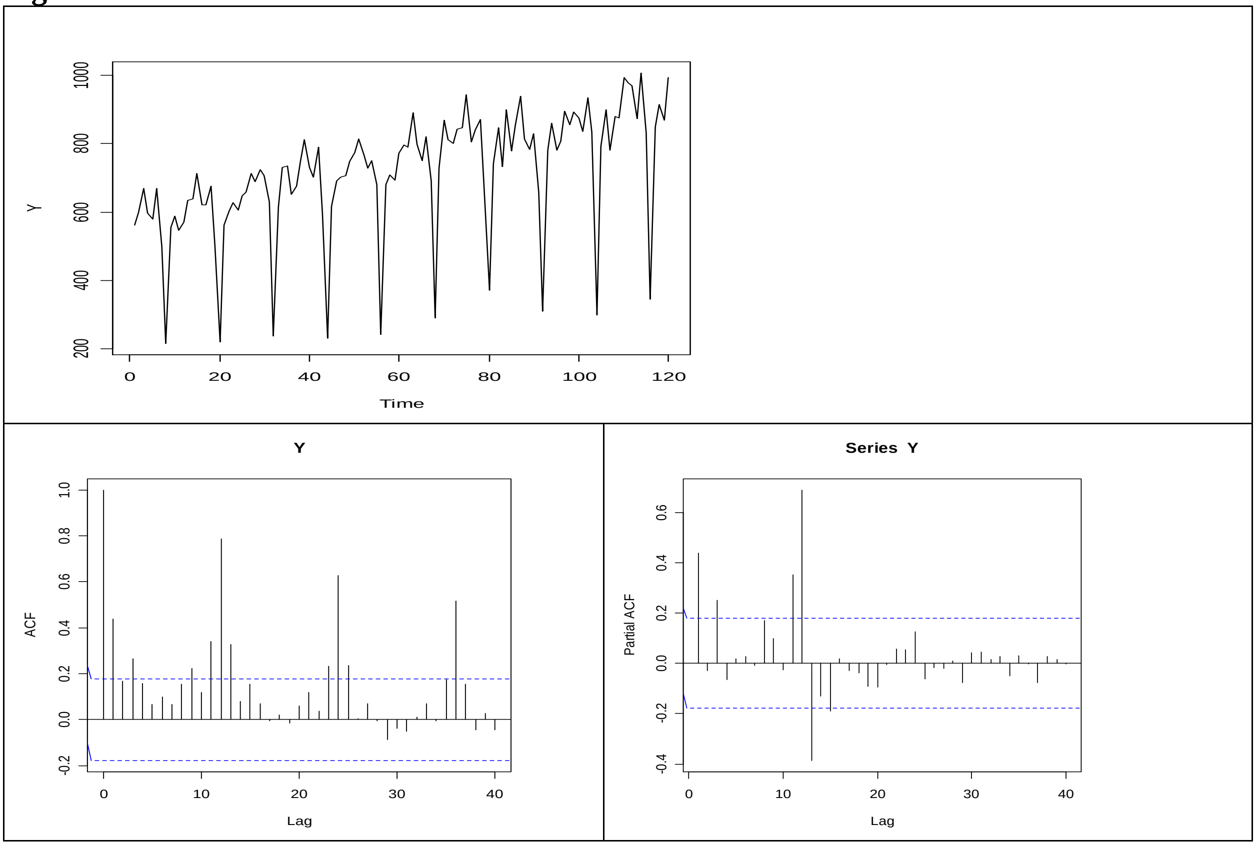
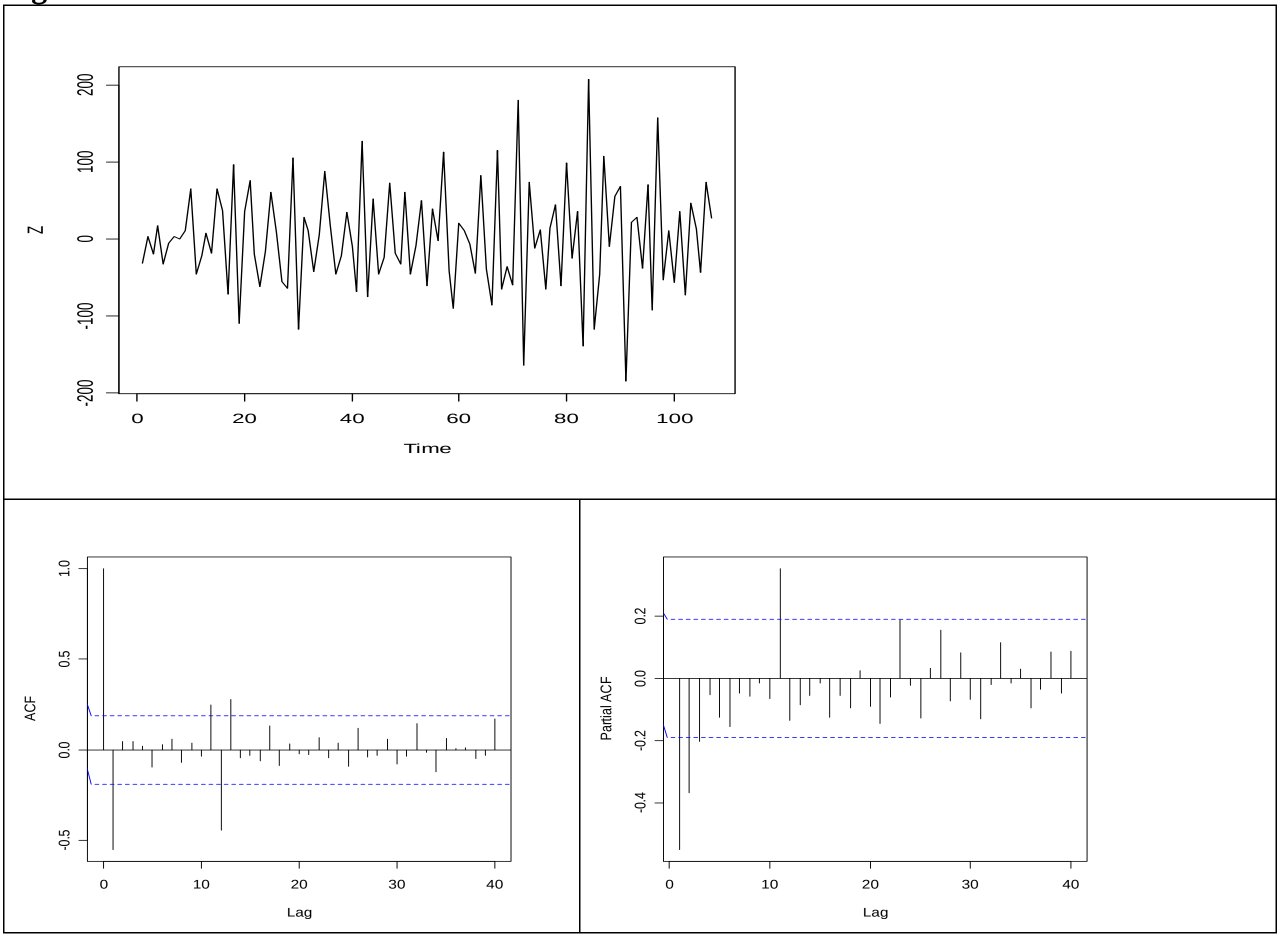


Figure 8.3 shows a very clear seasonal pattern in the data plot and a general increasing trend. The autocorrelations are almost all positive, and the dominant seasonal pattern shows clearly in the large values or r_{12}, r_{24}, and r_{36}.

This evidence suggests taking a seasonal difference. The seasonally difference data also appear non-stationary (plot is not shown) and so is differenced again at lag 1.

```
Y <- c(562.674, 599.000, 668.516, 597.798, 579.889,
       668.233, 499.232, 215.187, 555.813, 586.935, 546.136, 571.111,
       634.712, 639.283, 712.182, 621.557, 621.000, 675.989, 501.322,
       220.286, 560.727, 602.530, 626.379, 605.508, 646.783, 658.442,
       712.906, 687.714, 723.916, 707.183, 629.000, 237.530, 613.296,
       730.444, 734.925, 651.812, 676.155, 748.183, 810.681, 729.363,
       701.108, 790.079, 594.621, 230.716, 617.189, 691.389, 701.067,
       705.777, 747.636, 773.392, 813.788, 766.713, 728.875, 749.197,
       680.954, 241.424, 680.234, 708.326, 694.238, 772.071, 795.337,
       788.421, 889.968, 797.393, 751.000, 821.255, 691.605, 290.655,
       727.147, 868.355, 812.390, 799.556, 843.038, 847.000, 941.952,
       804.309, 840.307, 871.528, 656.330, 370.508, 742.000, 847.152,
       731.675, 898.527, 778.139, 856.075, 938.833, 813.023, 783.417,
       828.110, 657.311, 310.032, 780.000, 860.000, 780.000, 807.993,
       895.217, 856.075, 893.268, 875.000, 835.088, 934.595, 832.500,
       300.000, 791.443, 900.000, 781.729, 880.000, 875.024, 992.968,
       976.804, 968.697, 871.675, 1006.852, 832.037, 345.587, 849.528,
       913.871, 868.746, 993.733)
Y <- ts(Y)
plot(Y)
acf(Y, lag.max=40)
pacf(Y, lag.max=40)
```

Figure 8.4: A seasonal time series



The twice differenced data shown in **Figure 8.4** appear to be stationary, and a lot of the dominant seasonal spikes have disappeared.
From the *PACF* in **Figure 8.4**, the exponential decay of first few lags suggesting a non-seasonal $MA(1)$ model. The value r_1 in the *ACF* is significant reinforce a non-seasonal $MA(1)$ model and r_{12} is significant suggesting a seasonal $MA(1)$ model. Thus we end up with the tentative identification:

$$ARIMA\ (0,1,1)(0,1,1)_{12}.$$

Or

$$(1 - B)(1 - B^{12})Y_t = (1 - \theta_1 B)\ (1 - \Theta_1 B^{12})\varepsilon_t.$$

$\begin{pmatrix}\text{Non-}\\ \text{seasonal}\\ \text{difference}\end{pmatrix}$ $\begin{pmatrix}\text{seasonal}\\ \text{difference}\end{pmatrix}$ $\begin{pmatrix}\text{Non-}\\ \text{seasonal}\\ \text{MA(1)}\end{pmatrix}$ $\begin{pmatrix}\text{seasonal}\\ \text{MA(1)}\end{pmatrix}$

Recapitulation
The process of identifying a Box-Jenkins *ARIMA* model requires experience and good judgment, but there are some helpful guiding principles.

a) *Make the series stationary*:
An initial analysis of the raw data can quite readily show whether the time series is stationary in the mean and the variance. Differencing, (non-seasonal and/or seasonal) will usually take care of any non-stationarity in the mean. Logarithmic or power transformations will often take care of non-stationary variance.

b) *Consider non-seasonal aspects*:
An examination of the *ACF* and *PACF* of the stationary series obtained in step (a) can reveal whether a *AR* or *MA* model is feasible.

c) *Consider seasonal aspects*:
An examination of the *ACF* and *PACF* at the seasonal lags can help identify AR and MA models for the seasonal aspects of the data, but the indications are not easy to find as in the case of the non-seasonal aspects. For quarterly data, the forecaster should try to see the pattern of r_4, r_8, r_{12}, r_{16}, and so on, in the *ACF* and *PACF*. For monthly data, r_{12}, r_{24}, and possibly r_{36} could be examined.

8.6.2 Estimating the parameter

Having made a tentative model identification, the *AR* and *MA* parameters, seasonal and non-seasonal, have to be determined.
The method of least squares and maximum likelihood can be used for *ARIMA* models. However, for models involving an *MA* component, no simple formula that can be applied to obtain the estimates as there is in regression. An iterative method must be used. A preliminary estimate is chosen and a computer program refines the estimate iteratively until the sum of squared errors is minimized.

Example:
In Example Internet User, an $ARIMA(3,1,0)$ was identified:

$$Y_t' = \varphi_1 Y_{t-1}' + \phi_2 Y_{t-2}' + \phi_3 Y_{t-3}' + \varepsilon_t$$

where $Y_t' = Y_t - Y_{t-1}$. Using maximum likelihood estimation, the following information was produced by R:

```
Y <- c(88, 84, 85, 85, 84, 85, 83, 85, 88, 89, 91, 99,
       104, 112, 126, 138, 146, 151, 150, 148, 147, 149, 143, 132, 131,
       139, 147, 150, 148, 145, 140, 134, 131, 131, 129, 126, 126, 132,
       137, 140, 142, 150, 159, 167, 170, 171, 172, 172, 174, 175, 172,
       172, 174, 174, 169, 165, 156, 142, 131, 121, 112, 104, 102, 99,
       99, 95, 88, 84, 84, 87, 89, 88, 85, 86, 89, 91, 91, 94, 101, 110,
       121, 135, 145, 149, 156, 165, 171, 175, 177, 182, 193, 204, 208,
       210, 215, 222, 228, 226, 222, 220)
Y <- ts(Y)
arima(Y, order=c(3,1,0))

Call:
arima(x = Y, order = c(3, 1, 0))

Coefficients:
```

```
         ar1      ar2     ar3
      1.1513  -0.6612  0.3407
s.e.  0.0950   0.1353  0.0941

sigma^2 estimated as 9.363:  log likelihood = -252,  aic = 511.99
```
$VAr(\varepsilon_t) = 9.363$
```
#AIC = Akaike's Information Criterion
```

Hence the estimated model is

Note that this model now looks like an $AR(4)$ model. However, the parameters do not satisfy the conditions necessary to give a stationary series.

Example: Sales of printing/writing paper

```
> arima(Y, order=c(0,1,1), seasonal=list(order =c(0,1,1), period=12))

Call:
arima(x = Y, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period
= 12))

Coefficients:
          ma1     sma1
      -0.8402  -0.6360
s.e.   0.0611   0.0929

sigma^2 estimated as 1809:  log likelihood = -556.91,  aic = 1119.83
```

The model for this series is the $ARIMA$ (0,1,1)(0,1,1)12

$$(1-B)(1-B^{12})Y_t = (1-\theta_1 B)(1-\Theta_1 B^{12})\varepsilon_t$$

The estimated model is:

Note:
It is a convention to include the minus sign on the coefficients θ_k of MA parts.

8.6.3 Identification revisited

Having estimated an $ARIMA$ model, it is necessary to revisit the question of identification to see if the selected model can be improved. There are three aspects of model identification that can arise at this point in the modeling process.

1. Some of the estimated parameters may have been insignificant ($p-$value may have been larger than 0.05). If so revised model with the insignificant terms omitted may be considered.
2. The ACF and $PACF$ provide some guidance on how to select pure AR or pure MA models. But mixture models are much harder to identify. Therefore it is normal to begin with either a pure AR or a pure MA model, then considering extending the selected model to a mixed $ARMA$ model.

3. There may have been more than one plausible model identified, and we need a method to determine which of them is preferred. Because of these considerations it is common to have several competing models for the series and we need a method for selecting the best of the models.

The most commonly used criteria is to choosing the smallest Akaike's Information Criterion (AIC):

$$AIC = -2\ln L + 2m$$

where $m = p + q + P + Q$, the number of terms estimated in the model, L denotes the likelihood.

Note:

1. The AIC does not have much meaning by itself. It is only useful in comparison to the AIC value for another model fitted to the same data.
2. A difference in AIC values of 2 or less is not regarded as substantial and we may wish to choose a simpler model either for simplicity, or for the sake of getting a better model fit.

Example: Internet Usage

We previously identified that a first difference makes this non-seasonal series stationary. Therefore, we need consider only $ARIMA(p,1,q)$ models. Table below shows the AIC values for these models. The $ARIMA(3,1,0)$ selected initially still the best model since it has the smallest AIC value.

```
Y <- c(88, 84, 85, 85, 84, 85, 83, 85, 88, 89, 91, 99,
       104, 112, 126, 138, 146, 151, 150, 148, 147, 149, 143, 132, 131,
       139, 147, 150, 148, 145, 140, 134, 131, 131, 129, 126, 126, 132,
       137, 140, 142, 150, 159, 167, 170, 171, 172, 172, 174, 175, 172,
       172, 174, 174, 169, 165, 156, 142, 131, 121, 112, 104, 102, 99,
       99, 95, 88, 84, 84, 87, 89, 88, 85, 86, 89, 91, 91, 94, 101, 110,
       121, 135, 145, 149, 156, 165, 171, 175, 177, 182, 193, 204, 208,
       210, 215, 222, 228, 226, 222, 220)
Y <- ts(Y)
AIC <- matrix(0, 6,6)
for (p in 0:5)
  for (q in 0:5)
  {
    mod.fit <- arima(Y, order=c(p,1,q))
    AIC[p+1,q+1] <- mod.fit$aic
    p
  }
AIC
         [q=0]    [q=1]    [q=2]    [q=3]    [q=4]    [q=5]
[p=0] 630.9950 549.8055 519.8749 520.2717 519.3800 518.8573
[p=1] 529.2378 514.2995 516.2519 514.5763 515.1001 516.2762
[p=2] 522.1782 516.2914 517.3604 515.7733 513.2413 518.0892
[p=3] 511.9940 513.9377 515.6208 514.4139 514.7583 516.4277
[p=4] 513.9298 515.9558 516.1818 519.0777 515.3952 518.1249
[p=5] 515.8617 517.6386 513.5433 521.6405 511.1393 512.7706
```

8.6.4 Forecasting with *ARIMA* models

The notation we used is compact and convenient. An *ARIMA* (0,1,1)(0,1,1)$_{12}$ model is described as

$$(1-B)(1-B^{12})Y_t = (1-\theta_1 B)(1-\Theta_1 B^{12})\varepsilon_t$$

for example. However, in order to use an identified model for forecasting, it is necessary to expand the equation and make it look like a more conventional regression equation. For the model above, the form is

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \Theta_1\varepsilon_{t-12} + \theta_1\Theta_1\varepsilon_{t-13}$$

In order to use this equation to forecast 1 period ahead, i.e. Y_{t+1}, we increase the subscripts by one, throughout:

$$Y_{t+1} = Y_t + Y_{t-11} - Y_{t-12} + \varepsilon_{t+1} - \theta_1\varepsilon_t - \Theta_1\varepsilon_{t-11} + \theta_1\Theta_1\varepsilon_{t-12}$$

The term ε_{t+1} will not be known because the expected value of future random errors has to be taken as zero, but from the fitted model it will be possible to replace the values ε_t, ε_{t-11}, ε_{t-12} by their empirical determined values.

Example:

Sales of printing/writing paper

Period, t	Sales, Y_t	Error, ε_t
101	835.088	8.28
102	934.595	61.83
103	832.500	112.80
104	300.000	–88.64
105	791.443	–5.91
106	900.000	14.80
107	781.729	–40.66
108	880.000	12.80
109	875.024	–12.85
110	992.968	98.02
111	976.804	0.98
112	968.697	67.20
113	871.675	– 17.35
114	1006.852	55.97
115	832.037	13.01
116	345.587	–59.64
117	849.528	5.98
118	913.871	–26.62
119	868.746	17.02
120	993.733	70.87

The fitted model is

$$(1-B)(1-B^{12})Y_t = (1-0.84B)(1-0.636B^{12})\varepsilon_t$$

or

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \varepsilon_t - 0.84\varepsilon_{t-1} - 0.636\varepsilon_{t-12} + 0.534\varepsilon_{t-13}$$

To forecast the sales in period 121,

note: $\hat{\varepsilon}_{121} = 0$

For period 122, the forecast is

note: $\hat{\varepsilon}_{122} = 0$

8.6.5 Confidence Interval for l – period forecast

The $(1-\alpha)100\%$ confidence interval for the l – period forecast is given by

$$\hat{Y}_{t+l} \pm z_{\alpha/2}\sqrt{V(e_l)} \text{ if } \sigma_\varepsilon^2 \text{ is known,}$$

where $e_l = y_{t+l} - \hat{Y}_{t+l}$

Example:

a) The $AR(2)$ model $Y_t = 56 + 0.6Y_{t-1} - 0.3Y_{t-2} + \varepsilon_t$ is being used for forecasting. You are given $y_t = 60$, $y_{t-1} = 55$ and $V(\varepsilon_t) = 2$. Find the upper limit of the 95% 2–period ahead forecast interval.

b) The following ARMA(1, 1) model is used for forecasting:
Given that $\phi_1 = -0.6$, $\theta_1 = 0.4$, $y_{60} = 80$, $\mu = 50$, $\hat{\varepsilon}_{60} = -2$, $\sigma_\varepsilon^2 = 1$, find the 95% forecast interval for Y_{61}.