Chapter 7: Estimation of Time Series Models

7.0 Fitting Time-Series Models in the Time Domain

Previous chapter introduced several different types of probability models that may be used to describe time series. This chapter discusses the problem of fitting a suitable model to an observed time series.

7.1 Estimating Autocovariance, Autocorrelation and Partial Autocorrelation Functions

Recall that the autocovariance, autocorrelation and partial autocorrelation functions can be estimated using the following functions:

1. Sample autocovariance function

$$\gamma_k = \frac{1}{n}\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

2. Sample autocorrelation function (ACF)

$$r_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

3. Sample partial autocorrelation function ($PACF$)

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} r_{k-1,j}\, r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j}\, r_j} & \text{if } k = 2, 3, \ldots \end{cases}$$

where $r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$ for $j = 1, 2, 3, \ldots, k-1$

7.1.1 The sampling distribution of autocorrelation

For a white noise model,

$$r_k \sim N\left(0, \frac{1}{n}\right)$$

This information can be used to develop tests of hypotheses and confidence interval.

For example, approximately 95% of all sample autocorrelation must lie within $\pm 1.96/\sqrt{n}$. If this is not the case, the series is probably not white noise. For this reason, it is common to plot lines at $\pm 1.96/\sqrt{n}$ when plotting ACF .

Example 7.1:
Consider the time series consisting of the 36 observations which was constructed using uncorrelated random number between 0 and 100. The data was plotted in **Figure 7.1**. Suppose that this fact were not known. It could be determined by applying autocorrelation analysis. For uncorrelated data, we would expect each autocorrelation to be close to zero. **Figure 7.2** shows the autocorrelation function for time lags of $1, 2, 3, \ldots, 10$.

Figure 7.1:

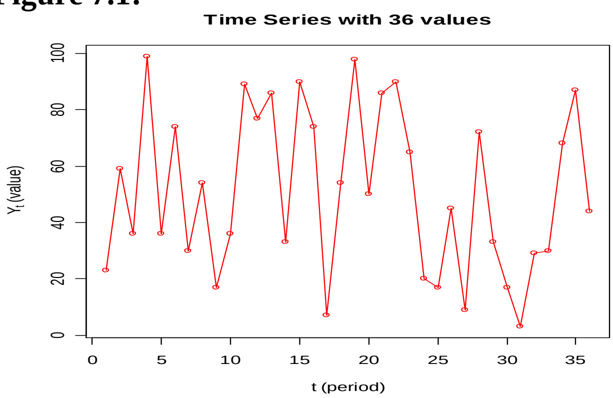
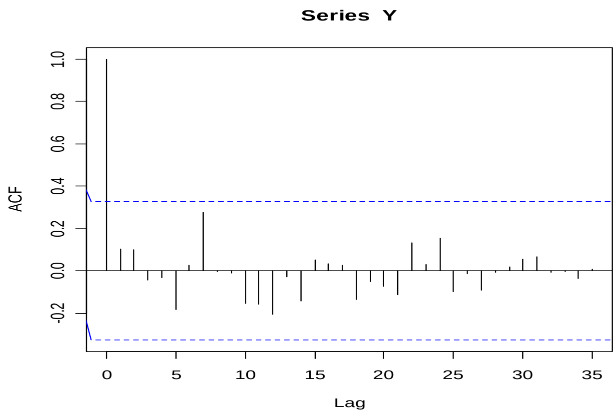


Figure 7.2:

$r_1 = 0.103$ $r_2 = 0.099$ $r_3 = -0.043$ $r_4 = -0.031$ $r_5 = -0.183$ $r_6 = 0.025$ $r_7 = 0.275$ $r_8 = -0.004$ $r_9 = -0.011$ $r_{10} = -0.152$	

The autocorrelation function is a valuable tool for investigating properties of an empirical time series. However, the statistical theory underlying r_k is quite complicated, and in some cases it is intractable. For the special case of a "white noise" series, the sampling distribution of r_k is known and can be used to practical advantage.

R-codes:

```
Y <- c(23, 59, 36, 99, 36, 74, 30, 54, 17, 36, 89, 77, 86, 33, 90,
       74, 7, 54, 98, 50, 86, 90, 65, 20, 17, 45, 9, 72, 33, 17,
       3, 29, 30, 68, 87, 44)
plot.ts(x = Y, ylab = expression(paste(Y[t], " (value)")), xlab = "t
(period)", type = "o", col = "red", main = "Time Series with 36 values")
# expression(paste(Y[t], " (value)")) = Y_t(Value)

# The function "acf" computes (and by default plots) estimates of the
autocovariance or autocorrelation function.
ACF <- acf(x=Y, lag.max = 40, type = c("correlation"), plot = TRUE)
ACF$"acf"[2:11]
```

Note:
Sample *ACF* can be used to identify the order of moving average, q. If the process is of order q, the *ACF* should all be close to zero for lags greater than q.

7.1.2 Portmanteau tests

Rather than study the r_k values one at a time, an alternative approach is to consider a whole set of r_k values, say the first 15 of them (r_1 through r_{15}) at one time, and develop a test to see whether the set is significantly different from a zero set. Test of this sort are known as portmanteau tests.

Two common portmanteau test are:

1. Box – Pierce Q statistic:

$$Q = n\sum_{k=1}^{h} r_k^2$$

where h is the maximum lag being considered and n is the number of observations in the series. Usually $h \approx 20$.

Clearly, if each r_k is close to zero, Q will be relatively small whereas if some r_k values are large (either positive or negative), the Q statistic will be relatively large.

The hypotheses are:

$$H_0 : \varepsilon_t \sim NID(0, \sigma^2)$$
$$H_1 : \varepsilon_t \text{ do not follow } NID(0, \sigma^2)$$

Under H_0, $Q \sim \chi^2(h-m)$ where m is the number of parameters in the model which has been fitted to the data.

Using the data from *Example 7.1*:

$$Q = 36\sum_{k=1}^{10} r_k^2 = 5.62.$$

R-code:

```
Box.test(Y, lag=10)
Box-Pierce test
data:  Y
X-squared = 5.6193, df = 10, p-value = 0.8462
```

2. Ljung – Box test:

$$Q^* = n(n+2)\sum_{k=1}^{h} (n-k)^{-1} r_k^2$$

Under H_0, $Q^* \sim \chi^2(h-m)$ where m is the number of parameters in the model which has been fitted to the data.

Using the data from *Example 7.1*:

$$Q^* = 36(38)\sum_{k=1}^{10} \frac{1}{(36-k)} r_k^2$$
$$= 36(38)\left[\frac{1}{35}r_1^2 + \frac{1}{34}r_2^2 + \cdots + \frac{1}{26}r_{10}^2\right]$$
$$= 7.22$$

R-codes:

```
Box.test(Y, lag = 10, type = "Ljung")

Box-Ljung test
data:  Y
X-squared = 7.2174, df = 10, p-value = 0.7048
```

7.1.3 Interpreting an Autocorrelation Chart

Example: *Recognizing seasonality in a time series*

Seasonality is defined as a pattern that repeats itself over fixed intervals of times. The sales of heating oil, for example, are high in winter and low in summer, indicating a 12-month seasonal pattern.

In general, seasonality can be found by identifying a large autocorrelation coefficient or a large partial autocorrelation coefficient at the seasonal lag. Often autocorrelations at multiples of the seasonal lag will also be significant. So, for monthly data, large autocorrelations might also be seen at lag 24 and even lag 36.

Example: *Examining stationarity of time series data*

The autocorrelations of stationary data drop to zero relatively quickly, while for a non-stationary series they are significantly different from zero for several time lags. When represented graphically, the autocorrelations of non-stationary data decrease slowly as the number of time lags increases.

Example: *Removing non-stationarity in a time series*

For example, Weekly Sales over 100,000 Rolls of Absorbent Paper Towels

R-codes:

```
y <- c(15,14.4064,14.9383,16.0374,15.632,14.3975,13.8959,14.0765,16.375,16.5342,
       16.3839,17.1006,17.7876,17.7354,17.001,17.7485,18.1888,18.5997,17.5859,15.7389,
       13.6971,15.0059,16.2574,14.3506,11.9515,12.0328,11.2142,11.7023,12.5905,12.1991,
       10.7752,10.1129,9.933,11.7435,12.259,12.5009,11.5378,9.6649,10.1043,10.3452,
       9.2835,7.7219,6.83,8.2046,8.5289,8.8733,8.7948,8.1577,7.9128,8.7978,
       9.0775,9.3234,10.4739,10.6943,9.8367,8.1803,7.2509,5.0814,1.8313,-0.9127,
       -1.3173,-0.6021,0.14,1.403,1.928,3.5626,1.9615,4.8463,6.5454,8.0141,
       7.9746,8.4959,8.4539,8.7114,7.378,8.1905,9.972,9.693,9.4506,11.2088,
       11.4986,13.2778,13.591,13.4297,13.3125,12.7445,11.7979,11.7319,11.6523,11.3718,
       10.5502,11.4741,11.5568,11.7986,11.8867,11.2951,12.7847,13.9435,13.6859,14.1136,
       13.8949,14.2853,16.3867,17.0884,15.8861,14.8227,15.9479,15.0982,13.877,14.2746,
       15.1682,15.3818,14.1863,13.9996,15.2463,17.0179,17.2929,16.6366,15.341,15.6453)
y <- ts(y)    # The function ts is used to create time-series objects
z <- ts(diff(y, lag=1))  # function diff = Returns suitably lagged and iterated
differences
z2 <- ts(diff(z, lag=1))
ts.plot(y,gpars=list(main="Original Values", xlab="Week", ylab="Sales", lty=1))
acf(y)        #acf computes estimates of autocorrelation function
pacf(y)       #pacf computes estimates of partial autocorrelation function

ts.plot(z,gpars=list(main= "First Differences", xlab="Week", ylab="Sales", lty=1))
acf(z, main="z")
pacf(z, main="z")
Box.test(z, lag=24)   #Box-pierce Q test
Box.test(z, lag=24,type="Ljung")        #Ljung-Box test

ts.plot(z2,gpars=list(main= "Second Differences", xlab="Week", ylab="Sales", lty=1))
acf(z2, main="z2")
pacf(z2, main="z2")
Box.test(z2, lag=24)
Box.test(z2, lag=24,type="Ljung")
```

Figure 7.3:

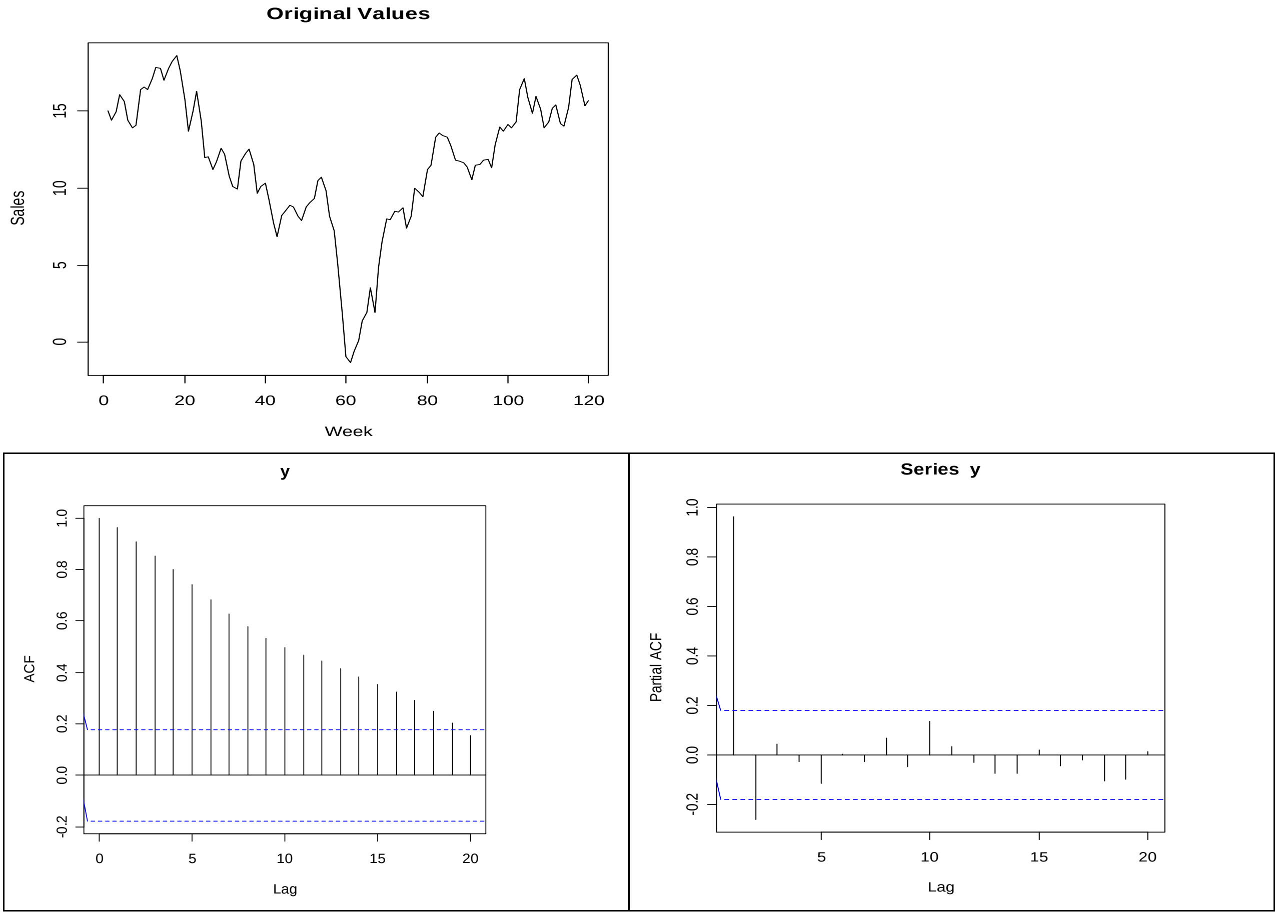
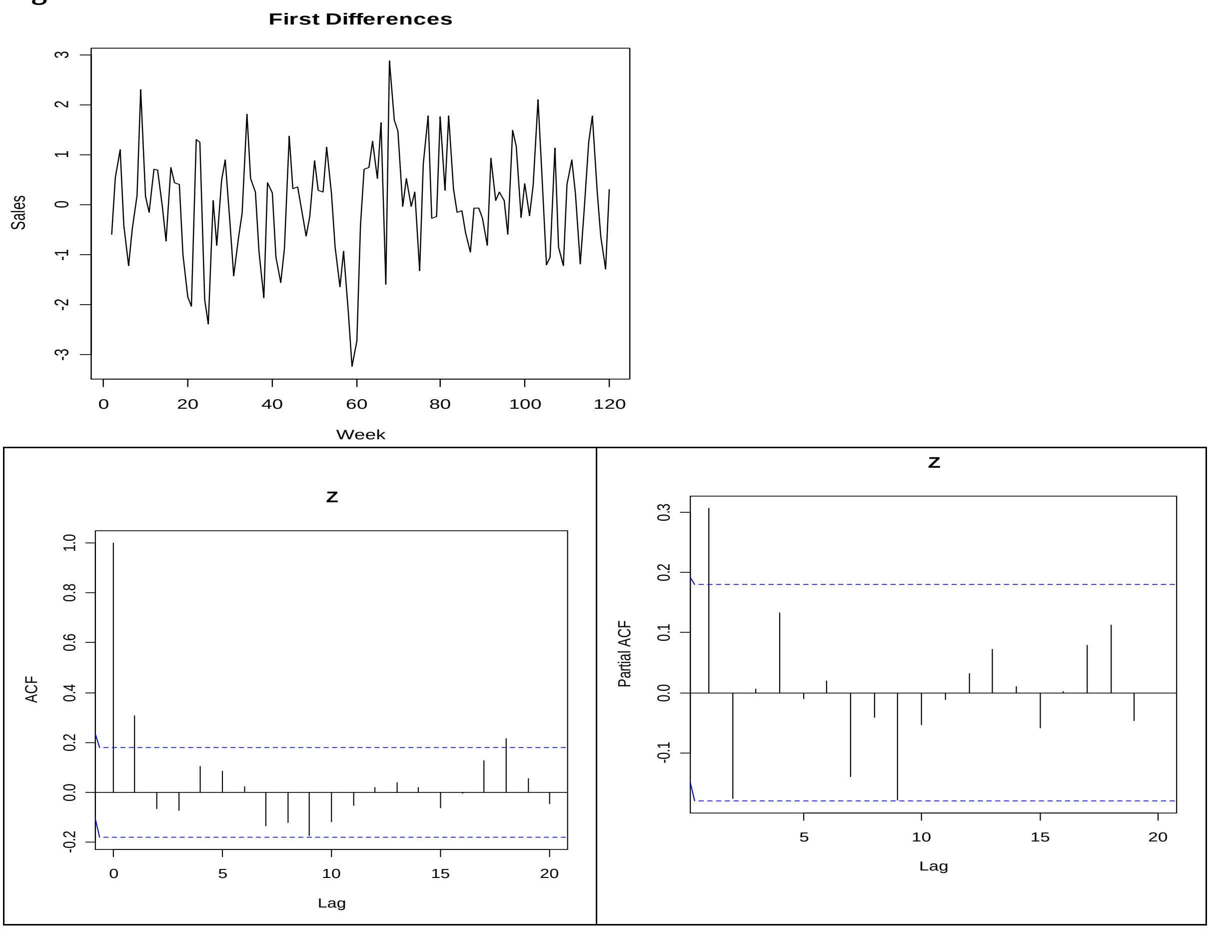


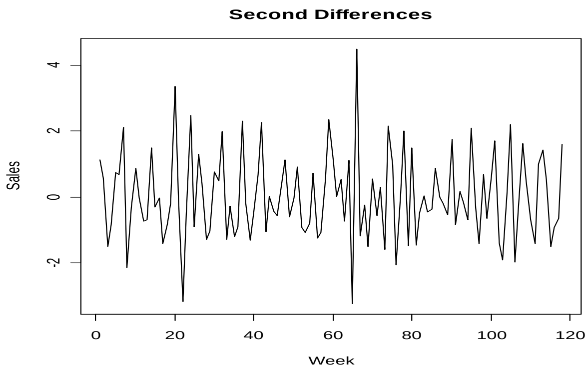
Figure 7.4:

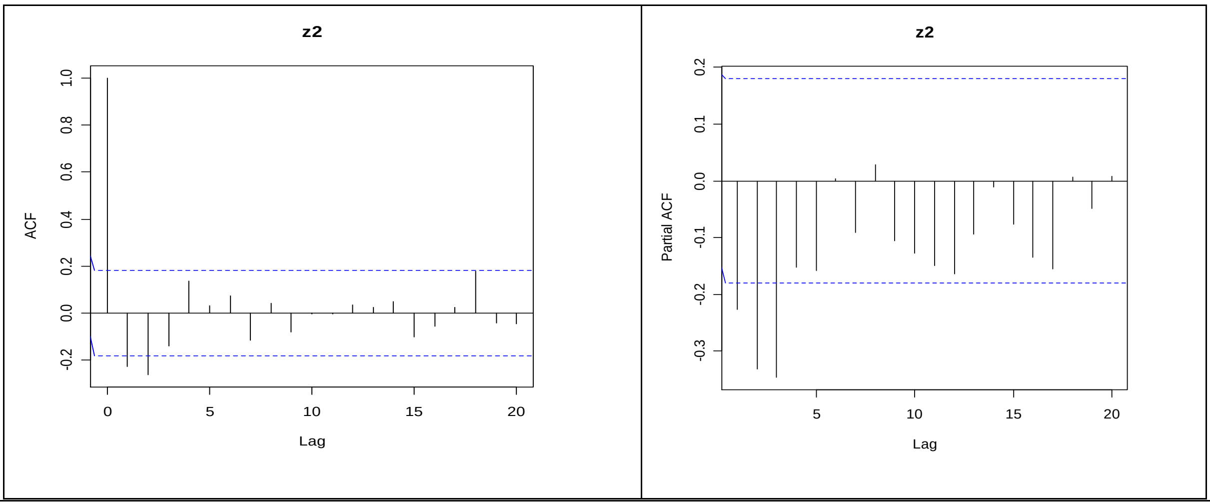


```
        Box-Pierce test
data:  z
X-squared = 11.1904, df = 1, p-value = 0.0008222
data:  z
X-squared = 33.866, df = 24, p-value = 0.08707

        Box-Ljung test
data:  z
X-squared = 11.4749, df = 1, p-value = 0.0007054
data:  z
X-squared = 37.217, df = 24, p-value = 0.04163
```

Figure 7.5:






```
> Box.test(z2, lag=24)
        Box-Pierce test
data:  z2
X-squared = 29.1928, df = 24, p-value = 0.2130

> Box.test(z2, lag=24, type="Ljung")
        Box-Ljung test

data:  z2
X-squared = 31.5537, df = 24, p-value = 0.1385
```

Figure 7.3 shows the analysis on the original sales data (Y), while **Figure 7.4** shows the analysis the sales data (Y) after taking the first difference. The $p-$values of Box–Pierce Q test and the Ljung – Box Q^* tests are both less than 10% indicating that the first differenced series is not white noise and hence non-stationary. We take a second difference of the series and shown in **Figure 7.5**, now, the series looks just like a white noise series, the $p-$values of Box–Pierce Q test and the Ljung–Box Q^* tests are both greater than 10% indicating that no significant evidence that the second differenced series is not a white noise.