B.Sc./Grad. Dip.: Probability Models and Time Series
MAS programmes: Stochastic Models and Time Series

Solutions 6

1. (a) The acf of $\{y_t\}$ dies away slowly, more or less linearly. This indicates the presence of trend. Thus it appears that $\{y_t\}$ is from a non-stationary process.

   (b) There appears to be a cut-off point at lag 2 in the acf of $\{\triangle y_t\}$, which indicates that $\{\triangle y_t\}$ is from an MA(2) process. To check this, we should use the approximate 95% probability limits for the $r_\tau,\ \tau > 2$ at

$$\pm\, 2\,\sqrt{\frac{1+2(r_1^2+r_2^2)}{T}},$$

   i.e., at

$$\pm\, 2\,\sqrt{\frac{1+2(0.145^2+0.362^2)}{399}},$$

   i.e., at ± 0.114. (Note that there are 399 observed differences.) None of the tabulated $r_\tau,\ \tau > 2$ falls outside these limits. This supports the hypothesis of an MA(2) process.

   Thus it appears that $\{y_t\}$ is from an ARIMA(0,1,2) process.

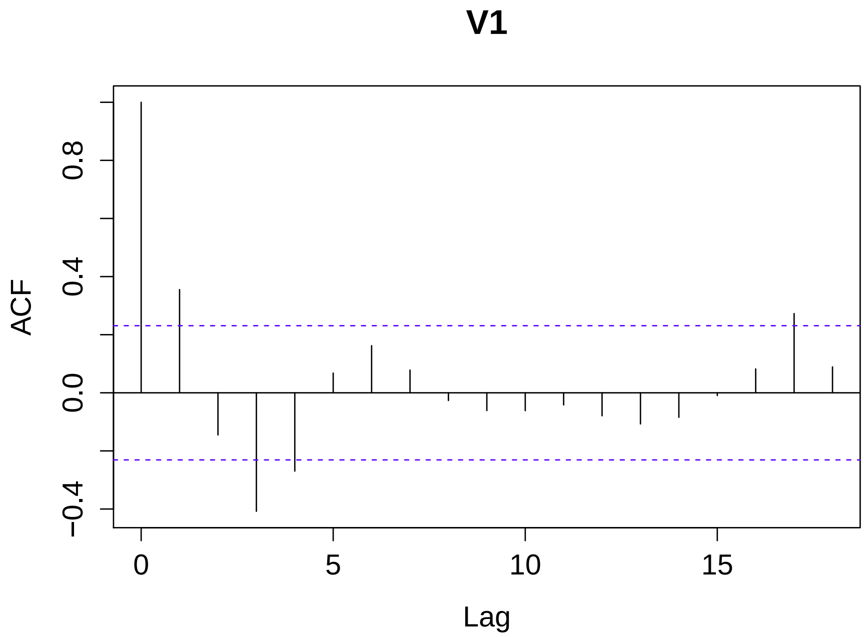
2. You will need to ensure that the file `sheepdata.txt` is saved into the working directory of RStudio or from whichever interface you are using to run R. You can download `sheepdata.txt` from moodle.

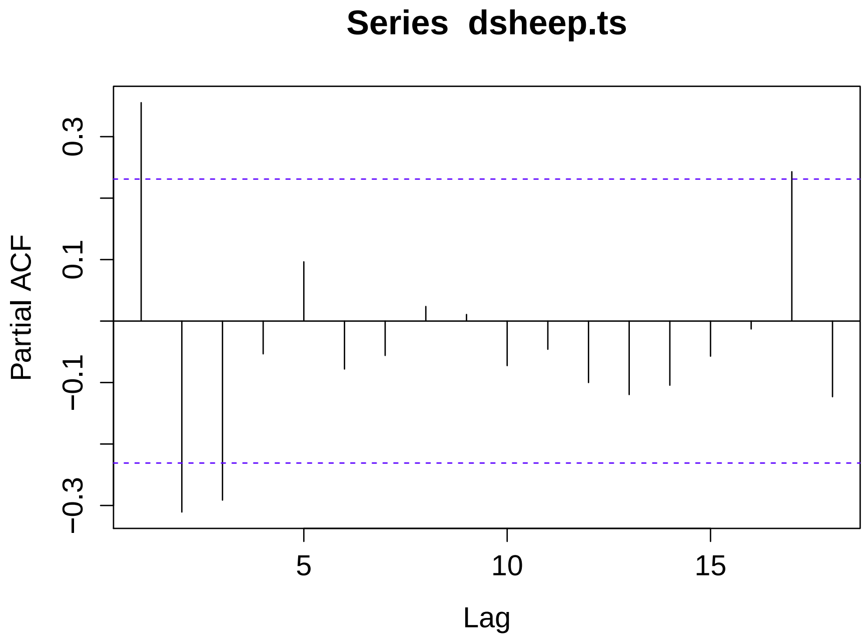
   The following is an outline of a suggested analysis.

```
> sheep <- read.table("sheepdata.txt")
> sheep.ts <- ts(sheep, start=1867)
> plot(sheep.ts, xlab = "", ylab = "",
+      main = "Sheep Pop. in England & Wales:1867 to 1939", las = 1)
> sheep.acf <- acf(sheep.ts, 18)
> dsheep.ts <- diff(sheep.ts)
> dsheep.acf <- acf(dsheep.ts, 18)
> dsheep.pacf <- acf(dsheep.ts, 18, type = "partial")
> sheep310 <- arima(sheep.ts, order = c(3, 1, 0))
> sheep310

Call:
arima(x = sheep.ts, order = c(3, 1, 0))

Coefficients:
         ar1      ar2      ar3
      0.4210  -0.2018  -0.3044
s.e.  0.1193   0.1363   0.1243

sigma^2 estimated as 4783:  log likelihood = -407.56,  aic = 823.12
> tsdiag(sheep310)
```

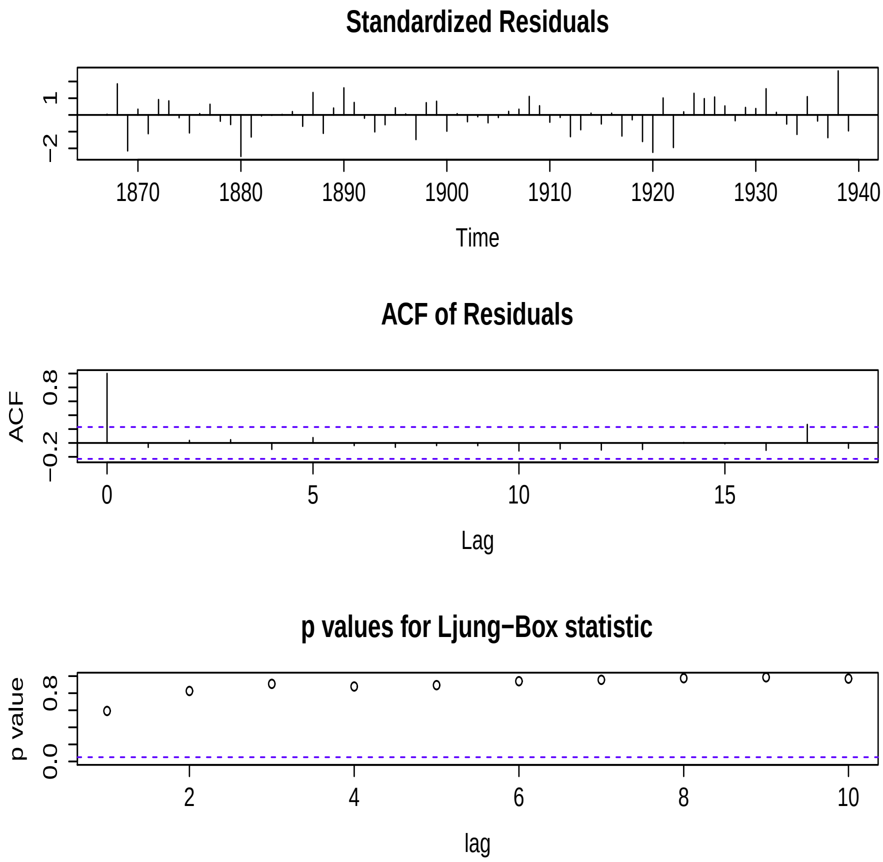
   (a) A plot of the data and an examination of the acf, which dies away rather slowly, indicates that there is a trend, so that differencing of the data is probably appropriate.

   (b) The following acf and pacf are for the differenced data.


V1
ACF
0.8
0.4
0.0
−0.4
0
5
10
15
Lag


Series dsheep.ts
Partial ACF
0.3
0.1
−0.1
−0.3
5
10
15
Lag

From examination of the acf and pacf, it seems that an AR(3) model might be a reasonable one to try for the differenced data, i.e., an ARIMA(3,1,0) model for the undifferenced data. (Note that 95% probability limits appropriate to the differenced data are given by $\pm 2/\sqrt{T} = \pm 2/\sqrt{72} = \pm 0.236$.) The freak significant values at lag 17 need not be taken too seriously – they seem to reflect the fact that there happen to be peaks in the sheep population at intervals of 17 years.



Nothing is indicated by the goodness-of-fit diagnostics which suggests that the ARIMA$(3, 1, 0)$ model (which has no process mean included) is inappropriate.

The fitted model equation is

$$\triangle Y_t = 0.4210 \triangle Y_{t-1} - 0.2018 \triangle Y_{t-2} - 0.3044 \triangle Y_{t-3} + \epsilon_t,$$

i.e.,

$$Y_t = 1.4210 Y_{t-1} - 0.6228 Y_{t-2} - 0.1026 Y_{t-3} + 0.3044 Y_{t-4} + \epsilon_t.$$