1. The TSA library contains the data set `co2`, which lists monthly carbon dioxide levels in northern Canada from 1/1994 to 12/2004. To load the data in R, remember that you need to first type

```
> library(TSA)
> data(co2)
```

(a) Construct a time series plot for these data. You can do this using the following R command:

```
> plot(co2,ylab="CO2 levels",xlab="Year",type="o")
```

Describe all systematic patterns you see in the plot.

(b) To enhance the usefulness of the plot, add monthly plotting symbols using the following R commands:

```
> plot(co2,ylab="CO2 levels",xlab="Year",type='l')
> points(y=co2,x=time(co2),pch=as.vector(season(co2)),cex=0.75)
```

Which months are consistently associated with highest CO_2 levels? the lowest? **Note:** The `cex=0.75` part controls the size of the plotting symbols specified in `pch`. Making `cex` larger increases the size of the plotting symbols.

(c) Consider fitting the simple straight line regression model (using the method of least squares)

$$Y_t = \beta_0 + \beta_1 t + e_t$$

to the data, where Y_t denotes the carbon dioxide level at time t. This model says that the CO_2 level is a linear function of time. You can do this using the R commands:

```
> model = lm(co2 ~ time(co2))
> summary(model)
```

What are the least squares estimates of β_0 and β_1? Write out the equation of the least squares regression line. In classical linear regression, what are the "usual" assumptions for the error terms e_t?

(d) Now construct a plot which superimposes the least squares fit over the time series plot from part (b). You can do this using the R commands:

```
> plot(co2,ylab="CO2 levels",xlab="Year",type='l')
> points(y=co2,x=time(co2),pch=as.vector(season(co2)),cex=0.75)
> abline(model)
```

(e) As with any regression analysis, we can use the residuals to assess the quality of the model. Use the following R commands to create a plot of the residuals from the least squares fit versus time:

```
> plot(y=rstudent(model),x=as.vector(time(co2)),xlab="Year",
    ylab="Standardised residuals",type='l')
> points(y=rstudent(model),x=as.vector(time(co2)),
      pch=as.vector(season(co2)),cex=0.75)
```

Comment on the adequacy of the straight line regression model in part (c). What other types of models might do a better job in capturing the systematic part of this time series?

2. Suppose that Z_1, Z_2, Z_3 are zero mean random variables with

$$\text{var}(Z_1) = 1 \quad \text{var}(Z_2) = 2 \quad \text{var}(Z_3) = 3$$

$$\text{cov}(Z_1, Z_2) = -0.5 \quad \text{cov}(Z_2, Z_3) = 2.5 \quad \text{cov}(Z_1, Z_3) = 0.$$

Calculate each of the following:
(a) $E(Z_1^2 - Z_2 - Z_2 Z_3)$
(b) $\text{var}(2Z_1 + 3Z_2 - Z_3)$
(c) $\text{cov}(3Z_1 - Z_2, Z_2 - 2Z_3)$
(d) $\text{corr}(Z_1, 2Z_2 + Z_3)$

3. (a) Simulate and plot a white noise process $e_t \sim \text{iid } \mathcal{N}(0,1)$ of length $n = 100$ using the following commands in R:

```
> wn.n01 = rnorm(100,0,1)
> plot(wn.n01,ylab="White noise process",xlab="Time",type="o")
```

(b) Repeat part (a) under the assumption that

- $e_t \sim \text{iid } t(1)$
- $e_t \sim \text{iid } \chi^2(4)$

To do this, just replace the first line of the code above with `wn.t1 = rt(100,1)` and `wn.chisq4 = rchisq(100,4)`, respectively. Comment on the differences among the 3 simulated white noise processes.
(c) Repeat part (a) using $n = 200$, $n = 300$, and $n = 500$. With your plot from part (a), take your 4 standard normal white noise processes and put them in a 2×2 matrix of plots using the `par(mfrow=c(2,2))` command in R. Label each plot in the matrix according to the sample size used, e.g.,

```
plot(wn.n01.100,ylab="WN",xlab="Time",main="Sample.size=100",type="o").
```

4. Do the following problems in Chapter 2 from Cryer and Chan: 2.1, 2.2, and 2.4.