HW 1 Solutions

(Points - 1.2, 1.4, 1.6, 1.7 each carry 20 points. Q.II Parts 1 and 2 each carry 10 points. Q.III Parts (a),(b),(c) carry 4,3,3 points respectively)

1.2 Figure 1 shows contrived data simulated according to this model. The modulating functions are also plotted. The code is given for part (a); for part (b), replace 20 by 200. For part (c), remove the `cos()` part.

```
s = c(rep(0,100), 10*exp(-(1:100)/200)*cos(2*pi*1:100/4))   # part (b)
x = ts(s + rnorm(200, 0, 1))
plot(x)
```

(c) The first signal bears a striking resemblance to the two arrival phases in the explosion. The second signal decays more slowly and looks more like the earthquake. The periodic behavior is emulated by the cosine function which will make one cycle every four points. If we assume that the data are sampled at 4 points per second, the data will make 1 cycle in a second, which is about the same rate as the seismic series.

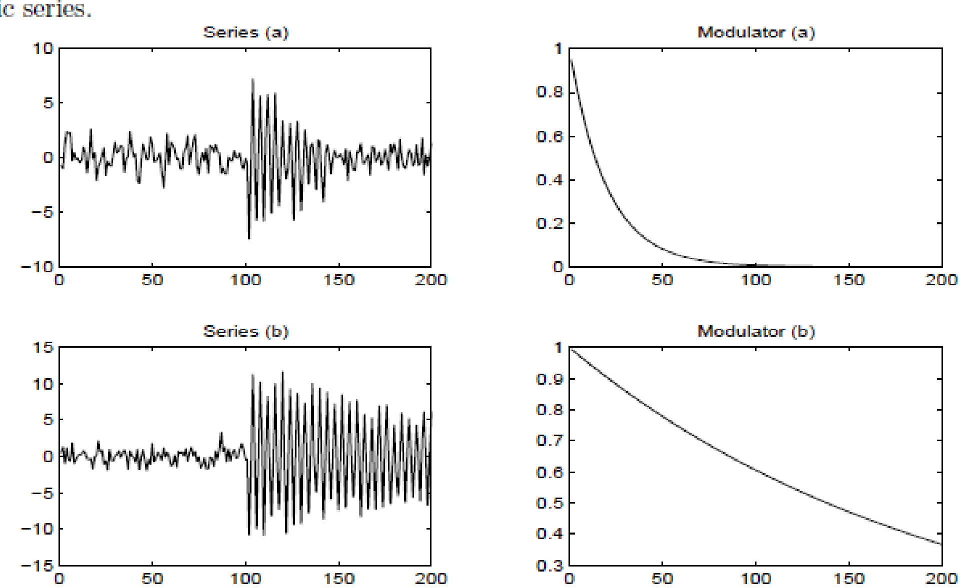

Figure 1: Simulated series with exponential modulations

1.4 Simply expand the binomial product inside the expectation and use the fact that μ_t is a nonrandom constant, i.e.,

$$
\begin{aligned}
\gamma(s,t) &= E[(x_s x_t - \mu_s x_t - x_s \mu_t + \mu_s \mu_t] \\
&= E(x_s x_t) - \mu_s E(x_t) - E(x_s)\mu_t + \mu_s \mu_t \\
&= E(x_s x_t) - \mu_s \mu_t - \mu_s \mu_t + \mu_s \mu_t
\end{aligned}
$$

1.6 (a) Since $Ex_t = \beta_1 + \beta_2 t$, the mean is not constant, i.e., does not satisfy (1.17). Note that

$$
\begin{aligned}
x_t - x_{t-1} &= \beta_1 + \beta_2 t + w_t - \beta_1 - \beta_2(t-1) - w_{t-1} \\
&= \beta_2 + w_t - w_{t-1},
\end{aligned}
$$

$|s-t|=0$

which is clearly stationary. Verify that the mean is β_2 and the autocovariance is 2 for $\cancel{s =}\ 2$ and -1 for $|s-t| = 1$ and is zero for $|s-t| > 1$.

(b) First, write

$$
\begin{aligned}
E(y_t) &= \frac{1}{2q+1} \sum_{j=-q}^{q} [(\beta_1 + \beta_2(t-j)] \\
&= \frac{1}{2q+1} \left[(2q+1)(\beta_1 + \beta_2 t) - \beta_2 \sum_{j=-q}^{q} j \right] \\
&= \beta_1 + \beta_2 t
\end{aligned}
$$

because the positive and negative terms in the last sum cancel out. To get the covariance write the process as

$$
y_t = \sum_{j=-\infty}^{\infty} a_j w_{t-j},
$$

where $a_j = 1, j = -q, \ldots, 0, \ldots, q$ and is zero otherwise. To get the covariance, note that we need

$$
\begin{aligned}
\gamma_y(h) &= E[(y_{t+h} - Ey_{t+h})(y_t - Ey_t)] \\
&= (2q+1)^{-2} \sum_j \sum_k a_j a_k E w_{t+h-j} w_{t-k} \\
&= \frac{\sigma^2}{(2q+1)^2} \sum_{j,k} a_j a_k \delta_{h+k-j}, \\
&= \sum_{j=-\infty}^{\infty} a_{j+h} a_j,
\end{aligned}
$$

where $\delta_{h+k-j} = 1, j = k+h$ and is zero otherwise. Writing out the terms in $\gamma_y(h)$, for $h = 0, \pm 1, \pm 2, \ldots$, we obtain

$$
\gamma_y(h) = \frac{\sigma^2(2q+1-|h|)}{(2q+1)^2}
$$

for $h = 0, \pm 1, \pm 2, \ldots, \pm 2q$ and zero for $|h| > q$.

1.7 By a computation analogous to that appearing in Example 1.17, we may obtain

$$
\gamma(h) = \begin{cases} 6\sigma_w^2 & h = 0 \\ 4\sigma_w^2 & h = \pm 1 \\ \sigma_w^2 & h = \pm 2 \\ 0 & |h| > 2. \end{cases}
$$

The autocorrelation is obtained by dividing the autocovariances by $\gamma(0) = 6\sigma_w^2$.

Question II :

1). $\sum_{i=1}^n (x_i - \bar{x}) = (\sum_{i=1}^n x_i) - n\bar{x} = n\bar{x} - n\bar{x} = 0.$

Also, $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})(y_i) - \bar{y} \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})(y_i).$

2). Using the previous part (II, Part 2), $S_x^{-2} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n \frac{1}{S_x^{-2}} (x_i - \bar{x}) y_i = \sum_{i=1}^n c_i y_i.$

3).(Bonus) (a). Given the linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, the least squares estimates of β_0 and β_1 can be obtained by minimizing RSS

$$RSS = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

Differentiating this w.r.t β_0, β_1 we get the normal equations as

$$\frac{\partial RSS}{\partial \beta_0} = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)(-1) = 0$$
$$\frac{\partial RSS}{\partial \beta_1} = 2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)(-x_i) = 0$$

Solving these equations we get the estimates as $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ and $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

(b). Using Question II - Part 2 and Question 3 - Part (a), we have $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i)}{S_x^2} = \sum_{i=1}^n c_i y_i$.

(c). First, we have

$$\sum_{i=1}^n \epsilon_i = \sum_{i=1}^n (y_i - \hat{y_i}) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) =$$
$$\sum_{i=1}^n (y_i - (\bar{y} - \hat{\beta}_1 \bar{x})) - n\hat{\beta}_1 \bar{x} = n\bar{y} - n\bar{y} = 0.$$

For proving the second identity we first look at the following term,

$$\sum_{i=1}^n \hat{y}_i x_i = \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i) x_i = \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 =$$
$$(\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = n\bar{x}\bar{y} + \hat{\beta}_1 (\bar{x} \sum_{i=1}^n x_i - \sum_{i=1}^n x_i^2) = n\bar{x}\bar{y} - n\bar{x}\bar{y} + \sum_{i=1}^n x_i y_i.$$

Hence we have $\sum_{i=1}^n \epsilon_i x_i = \sum_{i=1}^n (y_i - \hat{y}_i) x_i = \sum_{i=1}^n x_i y_i - x_i y_i = 0$. The second identity implies that the residuals vector $\epsilon = (\epsilon_1, ..., \epsilon_n)$ is orthogonal to the vector of observed values of the explanatory variable $x = (x_1, x_1, ..., x_n)$. $\sum_{i=1}^n \epsilon_i = 0$ would mean the residuals vector ϵ is orthogonal to the 1-vector (1,1,...,1).

HW 2 Solutions

(Points : Each question carries 20 points)

1.8 (a) Simply substitute $\delta s + \sum_{k=1}^{s} w_k$ for x_s to see that

$$\delta t + \sum_{k=1}^{t} w_k = \delta + \left(\delta(t-1) + \sum_{k=1}^{t-1} w_k\right) + w_t.$$

Alternately, the result can be shown by induction.

(b) For the mean,

$$Ex_t = E\left(\delta t + \sum_{k=1}^{t} w_k\right) = \delta t + \sum_{k=1}^{t} Ew_k = \delta t.$$

For the covariance, without loss of generality, consider the case $s \le t$, then

$$\begin{aligned}
\gamma(s,t) &= \text{cov}(x_s, x_t) = E\{(x_s - \delta s)(x_t - \delta t)\} \\
&= E\left\{\sum_{j=1}^{s} w_j \sum_{k=1}^{t} w_k\right\} \\
&= E\Big\{(w_1 + \cdots + w_s)(w_1 + \cdots + w_s + w_{s+1} + \ldots + w_t)\Big\} \\
&= \sum_{j=1}^{s} E(w_j^2) = s\,\sigma_w^2. \qquad [\text{or } \min(s,t)\,\sigma_w^2]
\end{aligned}$$

(c) The series is nonstationary because both the mean function and the autocovariance function depend on time, t.

(d) From (b), $\rho_x(t-1,t) = (t-1)\sigma_w^2 / \sqrt{(t-1)\sigma_w^2}\sqrt{t\sigma_w^2}$, which yields the result. The implication is that the series tends to change slowly.

(e) One possibility is to note that $\nabla x_t = x_t - x_{t-1} = \delta + w_t$, which is stationary.

1.9 Because $E(U_1) = E(U_2) = 0$, we have $E(x_t) = 0$. Then,

$$\begin{aligned}
\gamma(h) &= E(x_{t+h}x_t) \\
&= E\Big\{\big(U_1 \sin[2\pi\omega_0(t+h)] + U_2\cos[2\pi\omega_0(t+h)]\big)\big(U_1\sin[2\pi\omega_0 t] + U_2\cos[2\pi\omega_0 t]\big)\Big\} \\
&= \sigma_w^2\Big(\sin[2\pi\omega_0(t+h)]\sin[2\pi\omega_0 t] + \cos[2\pi\omega_0(t+h)]\cos[2\pi\omega_0 t]\Big) \\
&= \sigma_w^2 \cos[2\pi\omega_0(t+h) - 2\pi\omega_0 t] \\
&= \sigma_w^2 \cos[2\pi\omega_0 h]
\end{aligned}$$

by the standard trigonometric identity, $\cos(A-B) = \sin A \sin B + \cos A \cos B$.

1.10 (a) $MSE(A) = E\Big\{x_{t+\ell}^2 - 2AE(x_{t+\ell}x_t) + A^2E(x_t^2)\Big\} = \gamma(0) - 2A\gamma(\ell) + A^2\gamma(0)$. Setting the derivative with respect to A to zero yields

$$-2\gamma(\ell) + 2A\gamma(0) = 0$$

and solving gives the required value.

(b) $MSE(A) = \gamma(0)\left[1 - 2\frac{\rho(\ell)\gamma(\ell)}{\gamma(0)} + \rho^2(\ell)\right] = \gamma(0)\left[1 - 2\rho^2(\ell) + \rho^2(\ell)\right] = \gamma(0)\left[1 - \rho^2(\ell)\right]$.

(c) If $x_{t+\ell} = Ax_t$ with probability one, then

$$E(x_{t+\ell} - Ax_t)^2 = \gamma(0)\left[1 - \rho^2(\ell)\right] = 0$$

implying that $\rho(\ell) = \pm 1$. Since $A = \rho(\ell)$, the conclusion follows.

1.15 The process is stationary because

$$\mu_{x,t} = E(x_t) = E(w_t w_{t-1}) = E(w_t)E(w_{t-1}) = 0;$$

$$\gamma_x(0) = E(w_t w_{t-1} w_t w_{t-1}) = E(w_t^2)E(w_{t-1}^2) = \sigma_w^2 \sigma_w^2 = \sigma_w^4,$$

$$\gamma_x(1) = E(w_{t+1} w_t w_t w_{t-1}) = E(w_{t+1})E(w_t^2)E(w_{t-1}) = 0 = \gamma(-1),$$

and similar computations establish that $\gamma_x(h) = 0$, for $|h| \geq 1$. The series is white noise.

2.3 The code is given in the problem and below. The results should be all over the place and that's what should be taken away from this exercise.

```
par(mfcol = c(3,2)) # set up graphics
for (i in 1:6){
x = ts(cumsum(rnorm(100,.01,1))) # the data
reg = lm(x~0+time(x), na.action=NULL) # the regression
plot(x) # plot data
lines(.01*time(x), col="red", lty="dashed") # plot mean
abline(reg, col="blue") } # plot regression line
```

HW 3 Solutions

(Points: Part I carries 20 points, Each of 2.9,2.11,3.4,3.6,3.7 carries 16 points)

2.9 The R code for this problem is below.

```
summary(fit <- lm(soi~time(soi), na.action=NULL)) # part (a)
```

```
               Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  13.70367     3.18873     4.298  2.12e-05
time(soi)    -0.00692     0.00162    -4.272  2.36e-05 <- significant slope
```

```
soi.d = resid(fit)    # part (b), detrended data in soi.d
plot(soi.d)
(length(soi.d))   # = 453
per = as.vector(abs(fft(soi.d))^2/453)  # as.vector removes the ts attributes
ord = 1:227
freq = (ord-1)/453
plot(freq, per[ord], type="l")  # graph
cbind(freq , per[ord])          # list
```

The El Niño peak is around .024 or approx 1 cycle/42 months (`freq = 0.02428` and local max `per = 1.22`). The annual peak is at `freq = 0.0838` with `per = 9.33`.

2.11 (a) `ts.plot(oil, gas, col=1:2)` will plot the data on the same graph. The data resemble the random walks shown in Figure 1.10. We showed in Chapter 1, Example 1.18, that random walks are not stationary [this is also covered in Problem 1.8]. See Figure 2. Use the following URL for more details on the data set or more concurrent series: `http://tonto.eia.doe.gov/dnav/pet/pet_pri_spt_s1_w.htm` .

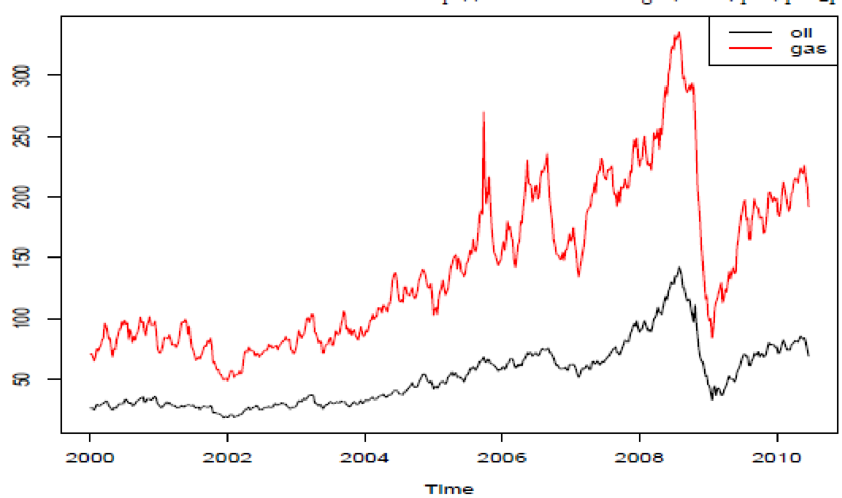

Figure 2: Oil in dollars per barrel; gas in cents per gallon

(c) The code is below and follows the hints. The transformed series look stationary and there is very little autocorrelation left after transformation, so a random walk seems plausible for each series. There are some very extreme values [students are pointed to note the outliers in part (e)]. The two series seem to be moving at the same time [this is stressed in parts (d) and (e)].

```
poil = diff(log(oil))
pgas = diff(log(gas))
ts.plot(poil, pgas, col=1:2)
acf(poil)
acf(pgas)
```

(d) There is strong cross-correlation at the zero lag [.66] and `poil` one week ahead [.18]. There is also feedback for `pgas` three weeks ahead [.17]. As noted in the hint, `ccf(poil, pgas)` has oil leading for negative values of `Lag`; i.e., R computes `corr[poil(t+Lag), pgas(t)]` for `Lag = 0, ±1, ±2, ...`.

(e) As in the hint, `lag.plot2(poil, pgas, 3, smooth=TRUE)` will give the plots. Aside from the aforementioned extreme outliers (10-30% changes in oil/gas prices in one week) the data seem fairly linear. Also, there may be some asymmetry noted in the lag 1 plot, but this appearance could be caused by outliers.

(f) (i) The hint shows how to set up the regression. Note that the interactions are not significant if you include them. The results are

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.006445   0.003464  -1.860  0.06338 .
poil         0.683127   0.058369  11.704  < 2e-16 ***
poilL        0.111927   0.038554   2.903  0.00385 **
indi         0.012368   0.005516   2.242  0.02534 *

Residual standard error: 0.04169 on 539 degrees of freedom
Multiple R-squared: 0.4563,     Adjusted R-squared: 0.4532
F-statistic: 150.8 on 3 and 539 DF,  p-value: < 2.2e-16
```

(ii) Note $-0.006445 + 0.012368 = .006$. Hence, the two models are

$$\widehat{G}_t = -.006 + .68O_t + .11O_{t-1} \quad \text{if } O_t < 0$$

$$\widehat{G}_t = .006 + .68O_t + .11O_{t-1} \quad \text{if } O_t \geq 0.$$

which suggests that there is no asymmetry for these data. The prices are FOB (*Free On Board: A sales transaction in which the seller makes the product available for pick up at a specified port or terminal at a specified price and the buyer pays for the subsequent transportation and insurance.*) so the asymmetry might be found in the gas pump prices.

(iii) `plot.ts(resid(fit))` and `acf(resid(fit))` seem to suggest the fit is not bad except for the very extreme outliers.

3.4 (a) Write this as $(1 - .3B)(1 - .5B)x_t = (1 - .3B)w_t$ and reduce to $(1 - .5B)x_t = w_t$. Hence the process is a causal and invertible AR(1): $x_t = .5x_{t-1} + w_t$.

(b) The AR polynomial is $1 - 1z + .5z^2$ which has complex roots $1 \pm i$ outside the unit circle (note $|1 \pm i|^2 = 2$). The MA polynomial is $1 - z$ which has root unity. Thus the process is a causal but not invertible ARMA$(2, 1)$.

3.6 Refer to Example 3.10. The roots of $\phi(z) = 1 - .9z^2$ are $\pm\, i/\sqrt{.9}$. Because the roots are purely imaginary, $\theta = \arg(i/\sqrt{.9}) = \pi/2$ and consequently, $\rho(h) = a\sqrt{.9}^h \cos(\frac{\pi}{2}h + b)$, or $\rho(h)$ makes one cycle every 4 values of h. Because $\rho(0) = 1$ and $\rho(1) = \phi_1/(1-\phi_2) = 0$, it follows that $a = 1$ and $b = 0$ in which case $\rho(h) = \sqrt{.9}^h \cos(\frac{\pi}{2}h)$. Thus $\rho(h) = \{1, 0, -\sqrt{.9}, 0, \sqrt{.9}^5, \dots\}$ for $h = 0, 1, 2, 3, 4, \dots$.

```
u = ARMAacf(ar=c(0,-.9), lag.max=25)
plot(0:25, u, type="h", xlab="Lag", ylab="ACF")
lines(0:25, u, type="p")
abline(h=0)
```

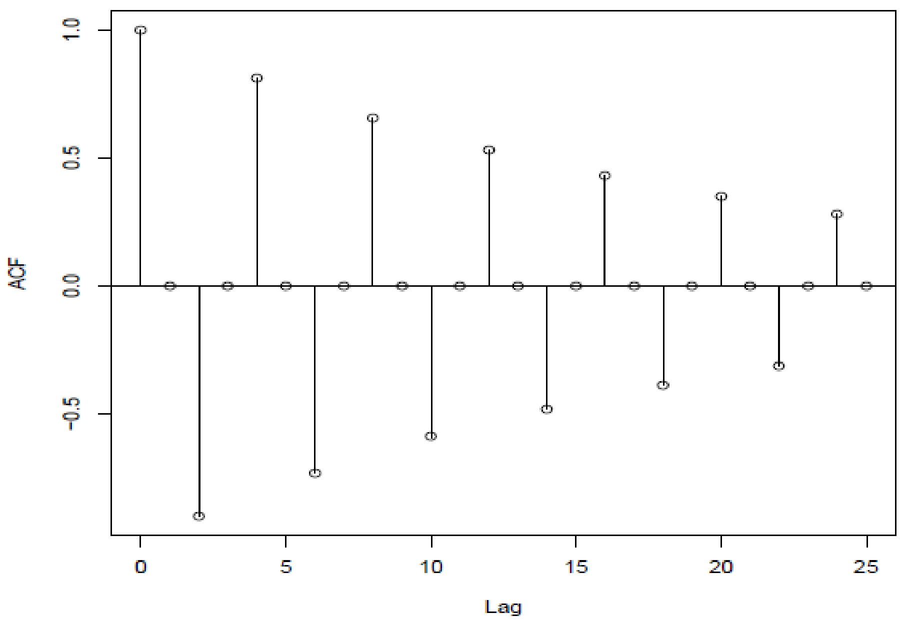

Figure 1: ACF for Problem 3.6

3.7 These problems can be tedious, so we suggest not assigning all three parts. Also, students can check their answers in R using `ARMAacf(ar=c(_,_), lag.max=10)`; fill in the blanks. For the ACFs we have $\rho(0) = 1$, $\rho(1) = \phi_1/(1-\phi_2)$ and

$$\text{distinct roots:}\quad \rho(h) = c_1 z_1^{-h} + c_2 z_2^{-h} \qquad \text{equal roots:}\quad \rho(h) = z_0^{-h}(c_1 + c_2 h)$$

(a) $\phi(z) = 1 + 1.6z + .64z^2 = (1 + .8z)^2$. This is equal roots case with $z_0 = -1/.8$. Thus $\rho(h) = -.8^h(a + bh)$. To solve for a and b, note for $h = 0$ we have $\rho(0) = 1 = a$ and for $h = 1$ we have $\rho(1) = -1.6/(1 + .64) = -.8(1 + b)$ or $b = .22$. Finally, $\rho(h) = -.8^h(1 + .22h)$ for $h = 0, 1, 2, \dots$.

(b) $\phi(z) = 1 - .4z - .45z^2 = (1 - .9z)(1 + .5z)$. This is the unequal roots case with $z_1 = 1/.9$ and $z_2 = -1/.5$. For the ACF, $\rho(h) = a.9^h - b.5^h$ where a and b are found by solving $1 = a + b$ and $.4/(1 - .45) = .9a - .5b$ or $a = .88$ and $b = .12$.

(c) $\phi(z) = 1 - 1.2z + .85z^2$. This is the complex roots case, with inverse roots $.6 \pm .7i$. `Arg(.6+.7i)` and `Mod(.6+.7i)` give $\theta = \arg(.6 + .7i) = .86$ radians and $|.6 + .7i| = .92$. Thus, $\rho(h) = a\ .92^h \cos(.86h + b)$ where a and b are found by solving $1 = a\cos(b)$ $[h = 0]$ and $1.2/(1 + .85) = a\ .92\cos(.86 + b)$ $[h = 1]$. Solving, $b = \pi$ and $a = -1$.

Part (I)

R Code :

```
#Function to return p-value from a linear regression fit
lmp <- function (modelobject) {
  if (class(modelobject) != "lm") stop("Not an object of class 'lm' ")
  f <- summary(modelobject)$fstatistic
  p <- pf(f[1],f[2],f[3],lower.tail=F)
  attributes(p) <- NULL
  return(p)
}

#(i) Simulating independent random walks
xt=rep(0,100)
yt=rep(0,100)
for(i in 2:100){
  xt[i]=xt[i-1]+rnorm(1)
  yt[i]=yt[i-1]+rnorm(1)
}

#(iii) Linear regression fit
lm1=lm(yt~xt)
```

```
#(b) Repeating the experiment 1000 times
count1=0
for(k in 1:1000)
{
  xt=rep(0,100)
  yt=rep(0,100)

  for(i in 2:100){
    xt[i]=xt[i-1]+rnorm(1)
    yt[i]=yt[i-1]+rnorm(1)
  }

  #(i).
  #plot(xt,yt)

  #(ii.) and (iii.)
  lm1=lm(yt~xt)
  #summary(lm1)

  if(lmp(lm1)<0.05){ count1=count1+1 }

}
```

```
print(count1)
```

The null hypothesis(no significant relationship) for the regression model is expected to be accepted as the two random walks are independently simulated and there is no meaningful relationship between them.

(b). A large number of significant outcomes were obtained while repeating the experiment 1000 times because the two series x_t and y_t are random walks and violate the assumptions of a linear model such as constant variance and independence. Thus it is necessary to be careful while performing linear regression when the variables are not stationary.

HW 4 Solutions

(Points : 3.10, 3.15, 3.17, 3.18, 3.20 each carry 20 pts. (Bonus)3.42 carries 10 pts.)

3.10 R code:

```
(reg = ar.ols(cmort, order=2, demean=FALSE, intercept=TRUE))
  Coefficients
       1       2
  0.4286  0.4418
  Intercept: 11.45 (2.394)
  Order selected 2  sigma^2 estimated as  32.32
predict(reg, n.ahead=4)
 $pred
  Time Series:
  Start = c(1979, 41)
  End = c(1979, 44)
  Frequency = 52
  [1] 87.59986 86.76349 87.33714 87.21350
 $se
  Time Series:
  Start = c(1979, 41)
  End = c(1979, 44)
  Frequency = 52
  [1] 5.684848 6.184973 7.134227 7.593357
```

3.15 For an AR(1), equation (3.86) is exact; that is, $E(x_{t+m} - x_{t+m}^t)^2 = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$. For an AR(1), $\psi_j = \phi^j$ and thus $\sigma_w^2 \sum_{j=0}^{m-1} \phi^{2j} = \sigma_w^2(1-\phi^{2m})/(1-\phi^2)$, the desired expression.

3.17
$$E(x_{n+m} - \widetilde{x}_{n+m})(x_{n+m+k} - \widetilde{x}_{n+m+k}) = E\Big(\sum_{j=0}^{m-1} \psi_j w_{n+m-j}\Big)\Big(\sum_{\ell=0}^{m+k-1} \psi_\ell w_{n+m+k-\ell}\Big)$$
$$= E\Big(\sum_{j=0}^{m-1} \psi_j w_{n+m-j}\Big)\Big(\sum_{\ell=k}^{m+k-1} \psi_\ell w_{n+m+k-\ell}\Big) = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j \psi_{j+k}$$

3.18 (a)–(b) Below `reg1` is least squares and `reg2` is Yule-Walker. The standard errors for each case are also evaluated; the Yule-Walker run uses Property 3.8. The two methods produce similar results.

```
(reg1 = ar.ols(cmort, order=2))  # coefs: [1] 0.4286 [2] 0.4418; sigma^2 estimated as  32.32
(reg2 = ar.yw(cmort, order=2))   # coefs: [1] 0.4339 [2] 0.4376 ; sigma^2 estimated as  32.84
(reg1$asy.se.coef)               #    se: [1] 0.0397 [2] 0.0397
(sqrt(diag(reg2$asy.var.coef)))  #    se: [1] 0.0400 [2] 0.0400
```

3.20 Although the results will vary, the data should behave like observations from a white noise process and each run should yield different parameter estimates where one is approximately the negative of the other. Students may also get warnings such as non-convergence or the SEs being `NaN`; it may also happen that the `ar` estimate is negative one.

```
x = arima.sim(list(order=c(1,0,1), ar=.9, ma=-.9), n=500)
plot(x)
acf2(x)
sarima(x, 1, 0 ,1)
```

3.42 (a) From the projection theorem, $x_{n+1}^n = \sum_{k=1}^n \alpha_k x_k$, where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)'$ satisfies

$$\begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix}.$$

Solving recursively we get $\alpha_2 = 2\alpha_1$, $\alpha_3 = 3\alpha_1$, and in general, $\alpha_k = k\alpha_1$ for $k = 1, ..., n$. This fact and the last equation gives $\alpha_1 = -\frac{1}{n+1}$ and the result follows.

(b) MSE$= \gamma(0) - a_1\gamma(-n) - a_2\gamma(-n+1) - \cdots - a_n\gamma(-1) = \sigma_w^2[2 - n/(n+1)] = \frac{(n+2)}{(n+1)}\sigma_w^2$.

HW 5 Solutions

(Points : 3.27, 3.35, 4.10 each carry 20 pts and 3.31, 3.32, 3.36, 3.39 each carry 10pts)

3.27 Write $\nabla^k x_t = (1-B)^k x_t = \sum_{j=0}^k c_j x_{t-j}$ where c_j is the coefficient of B^j in the binomial expansion of $(1-B)^k$. Because x_t is stationary, $E(\nabla^k x_t) = \mu_x \sum_{j=0}^k c_j$ independent of t, and (for $h \geq 0$) $\text{cov}(\nabla^k x_{t+h}, \nabla^k x_t) = \text{cov}(\sum_{j=0}^k c_j x_{t+h-j}, \sum_{j=0}^k c_j x_{t-j}) = \sum_{j=0}^{h+k} d_j \gamma_x(j)$, that is, the covariance is a time independent (linear) function of $\gamma_x(0), \dots, \gamma_x(h+k)$. Thus $\nabla^k x_t$ is stationary for any k.

Write $y_t = m_t + x_t$ where m_t is the given q-th order polynomial. Because $\nabla^k x_t$ is stationary for any k, we concentrate on m_t. Note that $\nabla m_t = m_t - m_{t-1} = c_q[t^q - (t-1)^q] + \sum_{j=0}^{q-1} c_j[t^j - (t-1)^j]$; from this it follows that the coefficient of t^q is zero. Now assume the result is true for $\nabla^k m_t$ and show it is true for $\nabla^{k+1} m_t$ [that is, for $k < q$, if $\nabla^k m_t$ is a polynomial of degree $q-k$ then $\nabla^{k+1} m_t$ is a polynomial of degree $q-(k+1)$]. The result holds by induction.

3.31 Follow the steps of Examples 3.38 and 3.39, `sarima(gnpgr, 1, 0, 0)` will produce the diagnostics. The results should be similar to those in Example 3.39.

3.32 The ACF of the returns reveals only small amounts of autocorrelation. The most appropriate models seem to be ARMA(1,1) or ARMA(0,3). BIC prefers the ARMA(1,1) whereas AIC prefers the ARMA(0,3). The diagnostics are ok but there are some major outliers that may be affecting the results. R code below:

```
poil = diff(log(oil))
acf2(poil)
sarima(poil, 1, 0, 1)   # BIC favors
sarima(poil, 0, 0, 3)   # AIC favors
```

3.35 (a) The model is $\text{ARIMA}(0,0,2) \times (0,0,0)_s$ (s can be anything) or $\text{ARIMA}(0,0,0) \times (0,0,1)_2$.

(b) The MA polynomial is $\theta(z) = 1 + \Theta z^2$ with roots $z = \pm i/\sqrt{\Theta}$ outside the unit circle (because $|\Theta| < 1$). To find the invertible representation, note that $1/[1-(-\Theta z^2)] = \sum_{j=0}^\infty (-\Theta z^2)^j$ from which we conclude that $\pi_{2j} = (-\Theta)^j$ and $\pi_{2j+1} = 0$ for $j = 0, 1, 2, \dots$. Consequently

$$w_t = \sum_{k=0}^{\infty} (-\Theta)^k x_{t-2k}.$$

(c) Write $x_{n+m} = -\sum_{k=1}^\infty (-\Theta)^k x_{n+m-2k} + w_n$ from which we deduce that

$$\widetilde{x}_{n+m} = -\sum_{k=1}^{\infty} (-\Theta)^k \widetilde{x}_{n+m-2k}$$

where $\widetilde{x}_t = x_t$ for $t \leq n$. For the prediction error, note that $\psi_0 = 1$, $\psi_2 = \Theta$ and $\psi_j = 0$ otherwise. Thus, $P_{n+m}^n = \sigma_w^2$ for $m = 1, 2$; when $m > 2$ we have $P_{n+m}^n = \sigma_w^2(1+\Theta^2)$.

3.36 Use the code from Example 3.44 with `ma=.5` instead of `ma=-.5`.

3.39 Because of the increasing variability, the data, jj_t, should be logged prior to any further analysis. A plot of the logged data, say $y_t = \ln jj_t$, shows trend, and one should notice the differences in the behavior of the series at the beginning, middle, and end of the data (as if there are 3 different regimes). *Because of these inconsistencies* *(nonstationarities), it is difficult to discover an ARMA model and one should expect students to come up with various models. In fact, assigning this problem may decrease your student evaluations substantially.*

Next, apply a first difference and seasonal difference to the logged data: $x_t = \nabla_4 \nabla y_t$. The PACF of x_t reveals a large correlation at the seasonal lag 4, so an SAR(1) seems appropriate. The ACF and PACF of the residuals reveals an ARMA(1,1) correlation structure for the within the seasons. This seems to be a reasonable fit. Hence, a reasonable model is an SARIMA$(1,1,0) \times (1,1,0)_4$ on the logged data. Below is R code for this problem.

```
plot(diff(diff(log(jj)),4))
acf2(diff(diff(log(jj)),4))
sarima(log(jj),1,1,0,1,1,0,4)
sarima.for(log(jj),4,1,1,0,1,1,0,4)
```

4.10 (a) Write the model in the notation of Chapter 2 as $x_t = \boldsymbol{\beta}'\boldsymbol{z}_t + w_t$, where $\boldsymbol{z}_t = (\cos(2\pi\omega_k t), \sin(2\pi\omega_k t))'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Then

$$\sum_{t=1}^{n} \boldsymbol{z}_t \boldsymbol{z}_t' = \begin{pmatrix} \sum_{t=1}^{n} \cos^2(2\pi\omega_k t) & \sum_{t=1}^{n} \cos(2\pi\omega_k t)\sin(2\pi\omega_k t) \\ \sum_{t=1}^{n} \cos(2\pi\omega_k t)\sin(2\pi\omega_k t) & \sum_{t=1}^{n} \sin^2(2\pi\omega_k t) \end{pmatrix} = \begin{pmatrix} n/2 & 0 \\ 0 & n/2 \end{pmatrix}$$

from the orthogonality properties of the sines and cosines. For example,

$$\begin{aligned} \sum_{t=1}^{n} \cos^2(2\pi\omega_k t) &= \frac{1}{4}\sum_{t=1}^{n} \left(e^{2\pi i\omega_k t} + e^{-2\pi i\omega_k t}\right)\left(e^{2\pi i\omega_k t} + e^{-2\pi i\omega_k t}\right) \\ &= \frac{1}{4}\sum_{t=1}^{n} \left(e^{4\pi i\omega_k t} + 1 + 1 + e^{-4\pi i\omega_k t}\right) = \frac{n}{2}, \end{aligned}$$

because, for example,

$$\sum_{t=1}^{n} e^{4\pi i\omega_k t} = \frac{e^{4\pi ik/n}\left(1 - e^{4\pi ik}\right)}{1 - e^{4\pi ik/n}} = 0$$

Substituting,

$$\begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} = \frac{2}{n} \begin{pmatrix} \sum_{t=1}^{n} x_t \cos(2\pi\omega_k t) \\ \sum_{t=1}^{n} x_t \sin(2\pi\omega_k t) \end{pmatrix} = 2n^{-1/2} \begin{pmatrix} d_c(\omega_k) \\ d_s(\omega_k) \end{pmatrix}.$$

(b) Now,

$$\begin{aligned} SSE &= \boldsymbol{x}'\boldsymbol{x} - 2n^{-1/2} \left(d_c(\omega_k), \quad d_s(\omega_k) \right) \begin{pmatrix} \sum_{t=1}^{n} x_t \cos(2\pi\omega_k t) \\ \sum_{t=1}^{n} x_t \sin(2\pi\omega_k t) \end{pmatrix} \\ &= \boldsymbol{x}'\boldsymbol{x} - 2\left[d_c^2(\omega_k) + d_s^2(\omega_k)\right] = \boldsymbol{x}'\boldsymbol{x} - 2I_x(\omega_k). \end{aligned}$$

(c) The reduced model is given by $x_t = w_t$, so that $RSS_1 = \sum_{t=1}^{n} x_t^2 = \boldsymbol{x}'\boldsymbol{x}$ and RSS is given in part (b). For the F-test we have $q = 2, q_1 = 0$, so that

$$F_{2,n-2} = \frac{2I_x(\omega_k)}{\boldsymbol{x}'\boldsymbol{x} - 2I_x(\omega_k)} \frac{n-2}{2}$$

is monotone in I_x.

HW 6 Solutions

2.8 (a) The variance in the second half of the varve series is obviously larger than that in the first half. Dividing the data in half gives $\hat{\gamma}_x(0) = 133, 594$ for the first and second parts respectively and the variance is about 4.5 times as large in the second half. The transformed series $y_t = \ln x_t$ has $\hat{\gamma}_y(0) = .27, .45$ for the two halves, respectively and the variance of the second half is only about 1.7 times as large. Histograms, computed for the two series in Figure 1 indicate that the transformation improves the normal approximation.

```
varv1 = varve[1:317]
varv2 = varve[318:634]
var(varv1)        # = 133.4574
var(varv2)        # = 594.4904
var(log(varv1))   # = 0.2707217
var(log(varv2))   # = 0.451371
par(mfrow=c(1,2))
hist(varve)
hist(log(varve))
plot(log(varve))     # for part (b)
acf(log(varve))      # for part (c)
plot(diff(log(varve))) # for part (d)
acf(diff(log(varve)))  # for part (d)
```

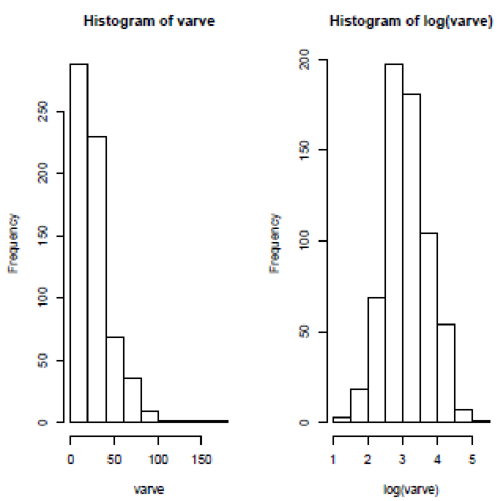

Figure 1: Histograms for varve series x_t and $y_t = \ln x_t$.

(b) The data between 300 and 450 show a positive trend that is similar to the global temperature data. (Presumably, this is due to a difference in the Earth's rotation.)

(c) The ACF of the y_t is positive for a large number of lags and decreases in a linear fashion.

(d) The plot of u_t and its ACF seem to indicate stationarity. The ACF has one significant value at lag 1 (with a value of $-.3974$). Because u_t can be written in the form

$$u_t = \log\left(\frac{x_t}{x_{t-1}}\right) = \log\left(1 + \frac{x_t - x_{t-1}}{x_{t-1}}\right) \approx \frac{x_t - x_{t-1}}{x_{t-1}},$$

it can be interpreted as the proportion of annual change.

(e)-(f) Note that

$$\gamma_u(0) = E[u_t - \mu_u)^2] = E[w_t^2] + \theta^2 E[w_{t-1}]^2 = \sigma_w^2(1+\theta^2)$$

and

$$\gamma_u(1) = E[(w_{t+1} - \theta w_t)(w_t - \theta_{t-1})] = -\theta E[w_t^2] = -\theta\sigma_w^2,$$

with $\gamma_u(h) = 0$ for $|h| > 1$. The ACF is

$$\rho(1) = \frac{-\theta}{1+\theta^2}$$

or

$$\rho(1)\theta^2 + \theta + \rho(1) = 0$$

and we may solve for

$$\theta = \frac{-1 \pm \sqrt{1 - 4\rho^2(1)}}{2\rho(1)}$$

using the quadratic formula. Hence, for $\hat{\rho}(1) = -.3974$

$$\theta = \frac{-1 \pm \sqrt{1 - 4(-.3974)^2}}{-2(.3974)},$$

yielding the roots $\hat{\theta} = -.4946, 2.0217$. We take the root $\theta = -.4946$ (this is the invertible root, see Chapter 3). Then,

$$\sigma_w^2 = \frac{\hat{\gamma}_u(0)}{1+\theta^2} = \frac{.3317}{1+(-.4946)^2} = .2665$$

4.16 (a) Since the means are both zero and the ACF's and CCF's

$$\gamma_x(h) = \begin{cases} 2 & h=0 \\ -1 & h=\pm 1 \\ 0 & |h| \geq 2 \end{cases} \quad \gamma_y(h) = \begin{cases} 1/2 & h=0 \\ 1/4 & h=\pm 1 \\ 0 & |h| \geq 2 \end{cases} \quad \gamma_{xy}(h) = \begin{cases} 0 & h=0 \\ -1/2 & h=1 \\ 1/2 & h=-1 \\ 0 & |h| \geq 2 \end{cases}$$

do not depend on the time index, the series are jointly stationary.

(b)

$$f_x(\omega) = |1 - \mathrm{e}^{-2\pi i\omega}|^2 = 2(1-\cos(2\pi\omega)) \quad \text{and} \quad f_y(\omega) = \frac{1}{4}|1+e^{-2\pi i\omega}|^2 = \frac{1}{2}(1+\cos(2\pi\omega))$$

As ω goes from $0 \to \frac{1}{2}$, $f_x(\omega)$ increases, whereas $f_y(\omega)$ decreases. This means x_t has more high frequency behavior and y_t has more low frequency behavior.

(c)

$$P\left\{\frac{2La}{f_y(.10)} \leq \frac{2L\bar{f}_y(.10)}{f_y(.10)} \leq \frac{2Lb}{f_y(.10)}\right\} = P\left\{\frac{2La}{f_y(.10)} \leq \chi^2_{2L} \leq \frac{2Lb}{f_y(.10)}\right\}$$

We can make the probability equal to .90 by setting

$$\frac{2La}{f_y(.10)} = \chi^2_{2L}(.95) \quad \text{and} \quad \frac{2Lb}{f_y(.10)} = \chi^2_{2L}(.05)$$

Setting $L = 3, \chi^2_6(.95) = 1.635, \chi^2_6(.05) = 12.592, f_y(.10) = .9045$ and solving for a and b yields $a = .25, b = 1.90$.

5.6 An AR(1) or MA(2) can be fit to the growth rate as detailed in Chapter 3, Example 3.38 (and the subsequent examples). Analysis of the residuals, the plot and the [P]ACF, suggest some ARCH behavior and the ARCH parameter is significant (although small). The following code is used to fit an ARCH(1) to the AR(1) residuals of the GNP growth rate.

```
gnpgr = diff(log(gnp))
sarima(gnpgr, 1, 0, 0)
acf2(innov^2) # get (p)acf of the squared residuals
library(fGarch)
summary(fit<-garchFit(~arma(1,0)+garch(1,0), data=gnpgr))
```

5.11 The data can be found in `sales.dat` and `lead.dat`. R code is below. After fitting the regression, the ACF and PACF indicate an AR(1) for the residuals, which fits well.

```
u = ts.intersect(diff(sales), lag(diff(lead),-3))
ds = u[,1]
dl3 = u[,2]
(fit = lm(ds~dl3))  # beta1hat is highly significant
acf2(fit$resid)     # => an AR(1) for the residuals
(fit = arima(ds, order=c(1,0,0), xreg=dl3))  # reg with ar1 errors
plot(resid(fit))
acf2(resid(fit))
```

5.13 (a) We transform inflow and take the seasonal difference $y_t = \ln i_t - \ln i_{t-12}$ which is proportional to the percentage yearly increase in flow. Monthly precipitation has some zero values and we use the square root transformation to stabilize this variable. Fitting to two series separately leads to the two ARIMA models

$$x_t = \nabla_{12} P_t = (1 - .812(.029)B^{12})w_t$$

and

$$y_t = (1 - .764(.033)B^{12})z_t,$$

with $\hat{\sigma}_w^2 = 32.503$ and $\hat{\sigma}_z^2 = .225$.

5.15 Below is the R code. Note that there is still zero-order correlation among the series, so a model that contains zero-order regressors may have done better (this is entirely reasonable because the data are quarterly, and hence the values evolve over months).

```
x = log(econ5[,1:3])
library(vars)
VARselect(x, lag.max=10, type="both")    # suggests an order 2 or 3 model
summary(fit <- VAR(x, p=2, type="both"))
(fit.pr = predict(fit, n.ahead = 24, ci = 0.95)) # 4 weeks ahead
fanchart(fit.pr) # plot prediction + error
```

Part II

(a) R Code

```
xt1=arima.sim(n=100,list(ar=c(0.5)))
xt2=arima.sim(n=100,list(ar=c(0.9)))
w=rnorm(100,0,1)

yt1=rep(0,100)
yt2=rep(0,100)
for(i in 1:100){
 yt1[i]=xt1[i]+sum(w[1:i])
 yt2[i]=xt2[i]+5*sum(w[1:i])
}
zt=5*yt1-yt2
par(mfrow=c(3,1))
plot.ts(yt1)
plot.ts(yt2)
plot.ts(zt)
```

(b). $\mathrm{Cov}(y_{t+h,1}, y_{t,1}) = \mathrm{Cov}(x_{t+h,1} + \sum_{j=1}^{t+h} w_j, x_{t,1} + \sum_{j=1}^{t} w_j) = \gamma_{x_1}(h) + t$.

The dependence of the covariance function on t makes $y_{t,1}$ nonstationary. Similarly, $y_{t,2}$ is nonstationary. Note that $x_{t,1}$ and $x_{t,2}$ are causal stationary $\mathrm{AR}(1)$ processes and their covariance functions are denoted as $\gamma_{x_1}(.)$ and $\gamma_{x_2}(.)$ respectively.

(c). $\mathrm{Cov}(z_{t+h}, z_t) = \mathrm{Cov}(5x_{t+h,1} - x_{t+h,2}\ , 5x_{t,1} - x_{t,2}) = 25\gamma_{x_1}(h) + \gamma_{x_2}(h)$.

Also, $E(z_t) = E(y_{t,1}) = E(y_{t,2}) = 0$. Hence z_t is weakly stationary.

(e). The cross covariance function $\gamma_{12}(s,t) = \mathrm{Cov}(y_{s,1}, y_{t,2}) = \mathrm{Cov}(\sum_{j=1}^{s} w_j,\ 5\sum_{k=1}^{t} w_k) = 5\ min(s,t)$.

Then using (1.16) and expression for the cross correlation function can be obtained. Since the series $y_{t,1}$ and $y_{t,2}$ are not each stationary they will not be jointly stationary.