HW2 Chunhua Yu 3.2 3.3 3.6 3.12 3.16 3.17

45/45

3.2 Suppose $Y_t = \mu + e_t - e_{t-1}$. Find $Var(\bar{Y})$ Note any unusually results. In particular compare your answer to what would have been obtained if $Y_t = \mu + e_t$ (Hint: avoid Equation 3.2.3.)

5/5

Solution:

$$\sum_{t=1}^{n} Y_t = \sum_{t=1}^{n} (\mu + e_t - e_{t-1}) = n\cdot\mu + \sum_{t=1}^{n} (e_t - e_{t-1})$$

$$= n\cdot\mu + e_1 - e_0 + e_2 - e_1 + \cdots + e_n - e_{n-1}$$

$$= n\mu + e_n - e_0$$

$$\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t = \mu + \frac{1}{n}(e_n - e_0)$$

$$Var(\bar{Y}) = Var(\mu + \frac{1}{n}(e_n - e_0)) = \frac{1}{n^2} Var(e_n - e_0) = \frac{2}{n^2}\sigma_e^2 \checkmark$$

The denominator of n^2 is very unusual. The denominator in the Variance of a sample mean is supposed to be n.

Applying the Ex 2.2 results in this case, $b=1$ $c=-1$

$\rho_1 = -\frac{1}{2}$ & $\rho_k = 0$ for $k>1$

recall: $Y_t = a + bZ_t + cZ_{t-1}$ $t = \{0, \pm1, \pm2, \ldots\}$ with $\{Z_t\} \sim W.N(0, \sigma^2)$

$$\gamma_k = \begin{cases} (b^2+c^2)\sigma^2 & k=0 \\ bc\cdot\sigma^2 & k=\pm1 \\ 0 & k=\pm2,\pm3\cdots \end{cases} \qquad \rho_k = \begin{cases} \frac{bc}{b^2+c^2} & k=\pm1 \\ 0 & k=\pm2,\pm3 \end{cases}$$

Hence, we see that the negative correlation at lag 1 makes it easier to estimate the process mean when compared with estimating the mean of a white noise process.

3.3 Suppose $Y_t = \mu + e_t + e_{t-1}$. Find the $Var(\bar{Y})$. Compare your answer to what would have been obtained if $Y_t = \mu + e_t$. Describe the effect that the autocorrelation in $\{Y_t\}$ has on $Var(\bar{Y})$

5/5

Solution:

$$\sum_{t=1}^{n} Y_t = \sum_{t=1}^{n} (\mu + e_t + e_{t-1}) = n\mu + \sum_{t=1}^{n} (e_t + e_{t-1})$$

$$= n\cdot\mu + e_1 + e_0 + e_2 + e_1 + \cdots + e_n + e_{n-1} = n\mu + e_0 + e_n + 2\sum_{t=1}^{n-1} e_t$$

$$Var(\bar{Y}) = Var(\frac{1}{n}\sum_{t=1}^{n} Y_t) = Var(\mu + \frac{1}{n}(e_0 + e_n + 2\sum_{t=1}^{n-1} e_t))$$

$$= \frac{1}{n^2}(\sigma_e^2 + \sigma_e^2 + 2(n-1)\sigma_e^2) = \frac{2(2n-1)}{n^2}\sigma_e^2 \checkmark$$

if $Y_t = \mu + e_t \Rightarrow Var(\bar{Y}) = \frac{1}{n}\sigma_e^2$

in the case given by condition. $Var(\bar{Y}) \approx \frac{4}{n} \cdot \sigma_e^2$ for large n, approximately four times larger.

Applying Ex 2.2 results in this case $\rho_1 = 1/2$ & $\rho_k = 0$ for $k > 1$, due to the positive autocorrelation at lag 1 makes it more difficult to estimate the mean in the white noise case.

3.16 Suppose that a stationary time series $\{Y_t\}$, has an auto correlation

10/10 function of form $\rho_k = \phi^k$ for $k > 0$, where ϕ is a constain in the range $(-1, +1)$

(a) Show that $Var(\bar{Y}) = \frac{\gamma_0}{n}\left[\frac{1+\phi}{1-\phi} - \frac{2\phi}{n}\frac{1-\phi^n}{(1-\phi)^2}\right]$ (hint: Use (3.2.3) ~~the finit~~ geometric sum $\sum_{k=0}^{n} \phi^k = \frac{1-\phi^{n+1}}{1-\phi}$; $\sum_{k=0}^{n} k\phi^{k-1} = \frac{d}{d\phi}\left[\sum_{k=0}^{n}\phi^k\right]$)

Solution: From the equation (3.2.3)

$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[1 + 2\cdot\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho_k\right] = \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\phi^k\right]$$

$$= \frac{\gamma_0}{n}\left[1 + 2\sum_{k=0}^{n}\left(1-\frac{k}{n}\right)\phi^k - 2\left(1-\frac{k}{n}\right)\phi^k\Big|_{k=0} - 2\left(1-\frac{k}{n}\right)\phi^k\Big|_{k=n}\right]$$

$$= \frac{\gamma_0}{n}\left[1 + 2\sum_{k=0}^{n}\left(1-\frac{k}{n}\right)\phi^k - 2 - 0\right]$$

$$= \frac{\gamma_0}{n}\left[-1 + 2\sum_{k=0}^{n}\phi^k - \frac{2\phi}{n}\sum_{k=0}^{n}k\phi^{k-1}\right]$$

since
$$\begin{cases} \sum_{k=0}^{n}\phi^k = \frac{1-\phi^{n+1}}{1-\phi} \\ \sum_{k=0}^{n}k\phi^{k-1} = \frac{d}{d\phi}\left[\sum_{k=0}^{n}\phi^k\right] = \frac{d}{d\phi}\left[\frac{1-\phi^{n+1}}{1-\phi}\right] = \frac{-(n+1)\phi^n(1-\phi) - (-1)(1-\phi^{n+1})}{(1-\phi)^2} \end{cases}$$

$$= \frac{(\phi-1)\phi^n n - \phi^n(1-\phi) + 1 - \phi^{n+1}}{(1-\phi)^2} = \frac{\phi^{n+1} n - \phi^n n - \phi^n + \cancel{\phi^{n+1}} + 1 - \cancel{\phi^{n+1}}}{(1-\phi)^2}$$

$$= \frac{\phi^{n+1}\cdot n - \phi^n\cdot n - \phi^n + 1}{(1-\phi)^2} = \frac{1-\phi^n}{(1-\phi)^2} - \frac{n\phi^n}{(1-\phi)}$$

$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[-1 + 2\cdot\frac{1-\phi^{n+1}}{1-\phi} - \frac{2\phi}{n}\cdot\frac{1-\phi^n}{(1-\phi)^2} + \frac{2\phi}{n}\cdot\frac{n\phi^n}{1-\phi}\right]$$

$$= \frac{\gamma_0}{n}\left[\frac{\phi-1+2-2\phi^{n+1}+2\phi^{n+1}}{1-\phi} - \frac{2\phi}{n}\cdot\frac{1-\phi^n}{(1-\phi)^2}\right]$$

$$= \frac{\gamma_0}{n}\left[\frac{1+\phi}{1-\phi} - \frac{2\phi}{n}\cdot\frac{1-\phi^n}{(1-\phi)^2}\right]$$

Verified.

(b) If n is large, argue that $Var(\bar{Y}) \approx \frac{\gamma_0}{n}\left[\frac{1+\phi}{1-\phi}\right]$

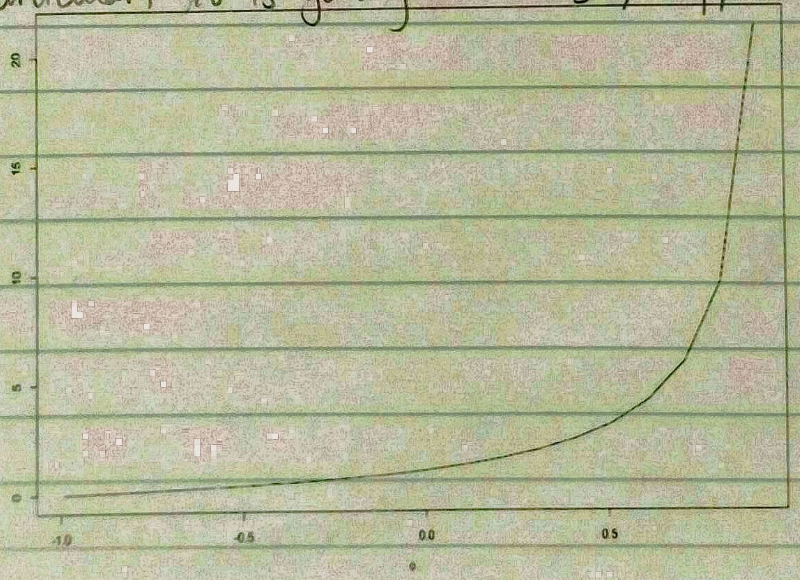
since $\phi \in (-1, +1)$ $\lim_{n\to\infty} \phi^n = 0 \Rightarrow \frac{2\phi}{n}\cdot\frac{1-\phi^n}{(1-\phi)^2} \to 0$ for large n

hence $Var(\bar{Y}) \approx \frac{\gamma_0}{n}\left[\frac{1+\phi}{1-\phi}\right]$ for large n.

(c) Plot $(1+\phi)/(1-\phi)$ for $\phi \in (-1, 1)$, Interpret the plot in terms of the precision in estimating the process mean.

Finding: Negative values of ϕ imply better estimates of the mean compared with white noise.

However, positive values of ϕ imply worse estimates. In particular, it is getting worse as ϕ approaches +1.



6/6

3.17 Verify Equation (3.2.6) on Page 29 (Hint: $\sum_{k=0}^{\infty} \phi^k = \frac{1}{1-\phi}$ for $-1<\phi<1$)

Solution: $Var(\bar{Y}) \approx \frac{\gamma_0}{n}\left[\sum_{k=-\infty}^{\infty} \rho_k\right] = \frac{\gamma_0}{n}\left[\sum_{k=-\infty}^{\infty} \phi^{|k|}\right]$ where $\rho_k = \phi^{|k|}$

$$= \frac{\gamma_0}{n}\left[\sum_{k=0}^{\infty} \phi^k + \sum_{-\infty}^{k=0} \phi^{-k} - \phi^{|k|}\Big|_{k=0}\right]$$

$$= \frac{\gamma_0}{n}\left[2 \cdot \sum_{k=0}^{\infty} \phi^k - 1\right]$$

$$= \frac{\gamma_0}{n}\left[2 \cdot \frac{1}{1-\phi} - 1\right] = \frac{\gamma_0}{n} \cdot \frac{2-(1-\phi)}{1-\phi}$$

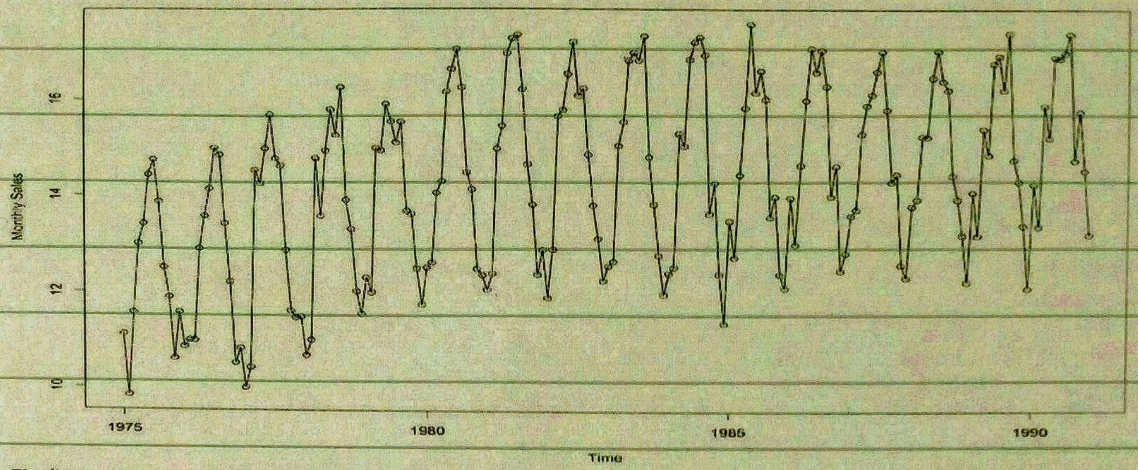
$$= \frac{1+\phi}{1-\phi} \cdot \frac{\gamma_0}{n}$$

Verified.

3.6 The data file beersales contains monthly U.S. beer sales (in millions of barrels) for the period January 1975 through December 1990.

(a) Display and interpret the plot the time series plot for these data.

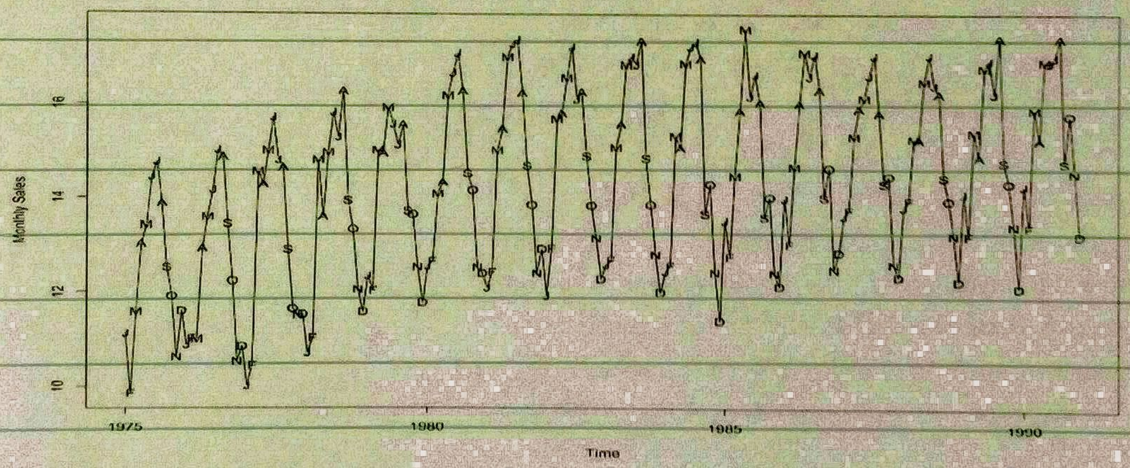
```
> library('TSA')
> data(beersales)
> plot(beersales,ylab='Monthly Sales',type='o')
```



Finding: *The plot is showing a typically seasonality trend. Beyond in the first five years, there is an apparently increasing trend. Afterward, the trend is close to flat. In the last five years, the trend were going down a little bit.*

(b) Now construct a time series plot that uses separate plotting symbols for the various months. Does your interpretation change from that in part (a)?

```
> plot(beersales,ylab='Monthly Sales',type='l')
> points(y=beersales,x=time(beersales),pch=as.vector(season(beersales)))
```



Finding:

From the plot with month symbols, it is pretty clear to find that the highest sales in the summer and lower sales in the winter.

(c) Use least squares to fit a seasonal-means trend to this time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.

> month.=season(beersales)

> Bsales.lm=lm(beersales~month.)

> summary(Bsales.lm)

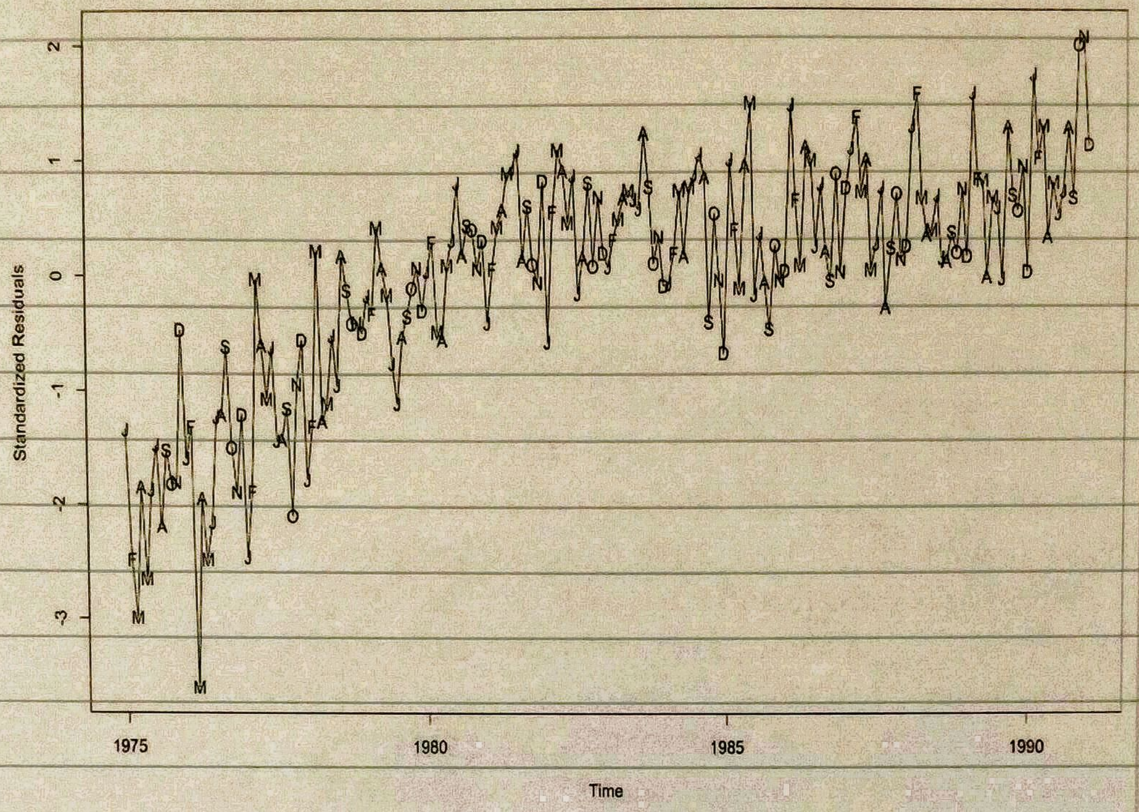
```
Call:
lm(formula = beersales ~ month.)
Residuals:
Min     1Q  Median    3Q    Max
-3.5745 -0.4772  0.1759  0.7312  2.1023
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)     12.48568   0.26392  47.309  < 2e-16 ***
month.February  -0.14259   0.37324  -0.382 0.702879
month.March      2.08219   0.37324   5.579 8.77e-08 ***
month.April      2.39760   0.37324   6.424 1.15e-09 ***
month.May        3.59896   0.37324   9.643  < 2e-16 ***
month.June       3.84976   0.37324  10.314  < 2e-16 ***
month.July       3.76866   0.37324  10.097  < 2e-16 ***
month.August     3.60877   0.37324   9.669  < 2e-16 ***
month.September  1.57282   0.37324   4.214 3.96e-05 ***
month.October    1.25444   0.37324   3.361 0.000948 ***
month.November  -0.04797   0.37324  -0.129 0.897881
month.December  -0.42309   0.37324  -1.134 0.258487
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.056 on 180 degrees of freedom
Multiple R-squared:  0.7103,     Adjusted R-squared:  0.6926
F-statistic: 40.12 on 11 and 180 DF,  p-value: < 2.2e-16
```

<u>Finding:</u>

The model with intercept term omits the January coefficient, all the other coefficients represent the different effect from January. Based on the p-values, we can conclude that January effect is significant, and March, April, May, June, July, August, September, and October all have the significant difference from January. Also, the Multiple R-squared is 71&.

(d) Construct and interpret the time series plot of the standardized residuals from part (c). Be sure to use proper plotting symbols to check on seasonality in the standardized residuals.

```
> plot(y=rstudent(Bsales.lm),x=as.vector(time(beersales)),type = 'l',xlab='Time',ylab = 'Standardized Residuals')
> points(y=rstudent(Bsales.lm),x=as.vector(time(beersales)),pch=as.vector(season(beersales)))
```



Finding:

From the plot, the residuals have a increasing trend rather than random. I don't think the least square model fit it well.

(e) Use least squares to fit a seasonal-means plus quadratic time trend to the beer sales time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.

> Bsales.lm2=lm(beersales~month.+time(beersales)+I(time(beersales)^2))

> summary(Bsales.lm2)

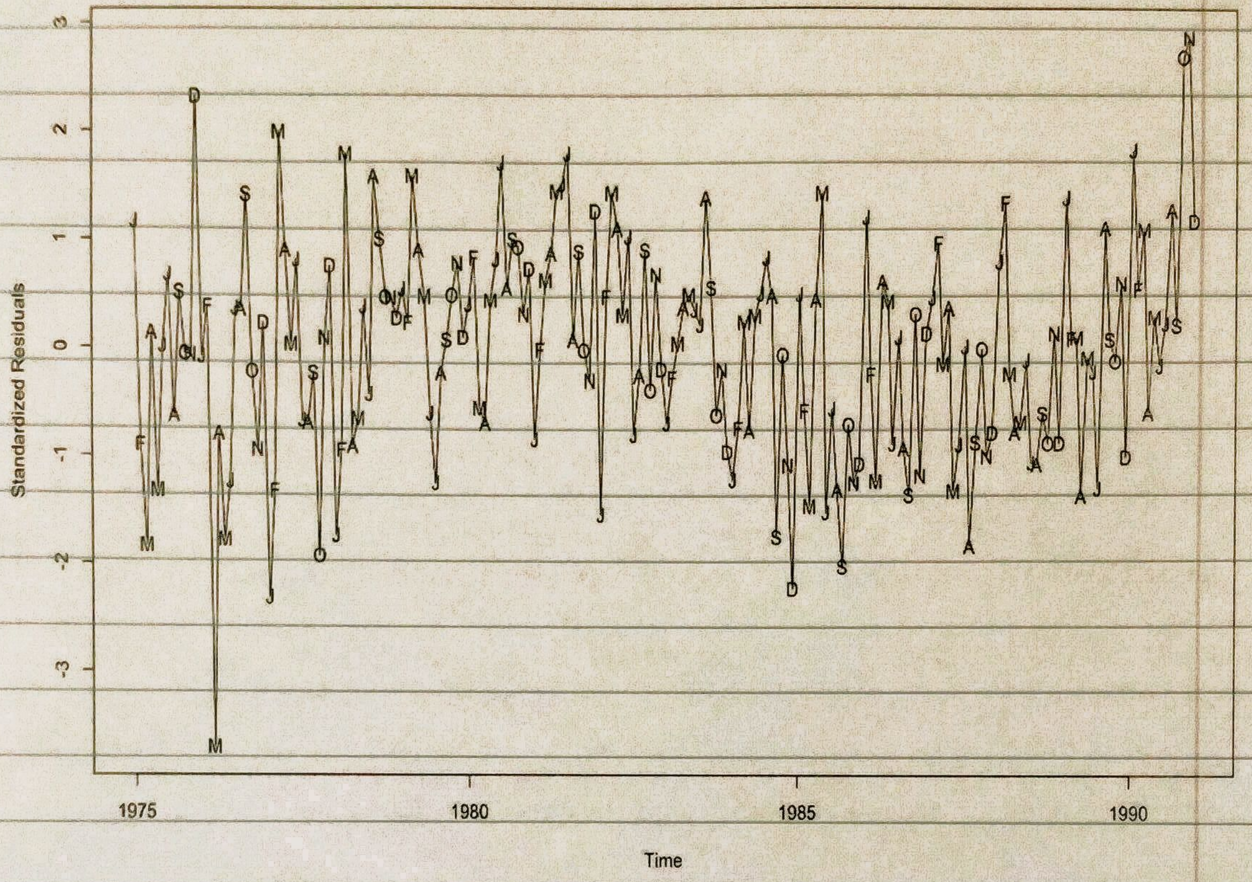
```
Call:
lm(formula = beersales ~ month. + time(beersales) + I(time(beersales)^2))
Residuals:
     Min       1Q   Median       3Q      Max
-2.03203 -0.43118  0.04977  0.34509  1.57572
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -7.150e+04  8.791e+03  -8.133 6.93e-14 ***
month.February       -1.579e-01  2.090e-01  -0.755  0.45099
month.March           2.052e+00  2.090e-01   9.818  < 2e-16 ***
month.April           2.353e+00  2.090e-01  11.256  < 2e-16 ***
month.May             3.539e+00  2.090e-01  16.934  < 2e-16 ***
month.June            3.776e+00  2.090e-01  18.065  < 2e-16 ***
month.July            3.681e+00  2.090e-01  17.608  < 2e-16 ***
month.August          3.507e+00  2.091e-01  16.776  < 2e-16 ***
month.September       1.458e+00  2.091e-01   6.972 5.89e-11 ***
month.October         1.126e+00  2.091e-01   5.385 2.27e-07 ***
month.November       -1.894e-01  2.091e-01  -0.906  0.36622
month.December       -5.773e-01  2.092e-01  -2.760  0.00638 **
time(beersales)       7.196e+01  8.867e+00   8.115 7.70e-14 ***
I(time(beersales)^2) -1.810e-02  2.236e-03  -8.096 8.63e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5911 on 178 degrees of freedom
Multiple R-squared:  0.9102,  Adjusted R-squared:  0.9036
F-statistic: 138.8 on 13 and 178 DF,  p-value: < 2.2e-16
```

Finding:

The Multiple R-squared is rather large 91%. All the terms except February and November are significantly different from January. It looks much better than the model in (c)

(f) Construct and interpret the time series plot of the standardized residuals from part (e). Again use proper plotting symbols to check for any remaining seasonality in the residuals.

```
> plot(y=rstudent(Bsales.lm2),x=as.vector(time(beersales)),type = 'l',xlab='Time',ylab = 'Standardized Residuals')
> points(y=rstudent(Bsales.lm2),x=as.vector(time(beersales)),pch=as.vector(season(beersales)))
```



Finding:

This model fits much better than the model in (c). Since the coefficient of quadratic term is significant, there is sufficient evidence that quadratic has significant effect on the model. Due to the coefficient is negative, it can be predicted quadratic trend will be decreasing.

3.12 Consider the time series in the data file beersales.

(a) Obtain the residuals from the least squares fit of the seasonal-means plus quadratic time trend model.

```
> Bsales.lm2=lm(beersales~month.+time(beersales)+I(time(beersales)^2))
> resid.lm2=rstudent(Bsales.lm2)
> round(resid.lm2,3)     # round residuals up to three decimal places
```

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
1.165 -0.899 -1.829 0.139 -1.314 0.023 0.680 -0.623 0.518 -0.045 -0.062 2.323 -0.072 0.386 -3.714 -0.787 -1.768

18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
-1.230 0.379 0.367 1.422 -0.205 -0.935 0.240 -2.311 -1.317 1.994 0.915 0.048 0.826 -0.659 -0.680 -0.217 -1.921

35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
0.107 0.773 -1.731 -0.935 1.793 -0.896 -0.638 0.400 -0.411 1.586 1.023 0.487 0.482 0.295 0.564 0.251 1.590

52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68
0.919 0.496 -0.584 -1.237 -0.213 0.091 0.509 0.799 0.119 0.420 0.848 -0.537 -0.679 0.458 0.836 1.712 0.567

69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85
1.023 0.950 0.331 0.749 -0.827 -0.003 0.640 0.893 1.454 1.528 1.802 0.086 0.910 -0.005 -0.284 1.279 -1.530

86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102
0.490 1.447 1.116 0.320 1.043 -0.784 -0.234 0.920 -0.371 0.695 -0.176 -0.680 -0.262 0.061 0.390 0.505 0.370

103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119
0.227 1.390 0.569 -0.605 -0.183 -0.944 -1.203 -0.727 0.252 -0.749 0.311 0.518 0.841 0.492 -1.736 -0.048 -1.069

120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136
-2.218 0.511 -0.558 -1.445 0.457 1.434 -1.509 -0.543 -1.299 -2.014 -0.692 -1.231 -1.056 1.211 -0.228 -1.218 0.621

137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153
0.441 -0.870 0.108 -0.912 -1.348 0.326 -1.167 0.144 0.479 0.975 -0.138 0.375 -1.308 -0.880 0.023 -1.822 -0.863

154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
-0.001 -0.992 -0.771 0.808 1.339 -0.229 -0.775 -0.678 -0.110 -1.052 -1.067 -0.598 -0.873 0.141 -0.876 1.384 0.088

171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187
0.098 -1.360 -0.089 -0.213 -1.294 1.111 0.080 -0.120 0.603 -1.004 1.821 0.542 1.090 -0.599 0.289 -0.159 0.228

188 189 190 191 192
1.267 0.213 2.684 2.866 1.167
```

(b) Perform a runs test on the standardized residuals and interpret the results.

> runs(resid.lm2)   # resid.lm2=rstudent(Bsales.lm2)

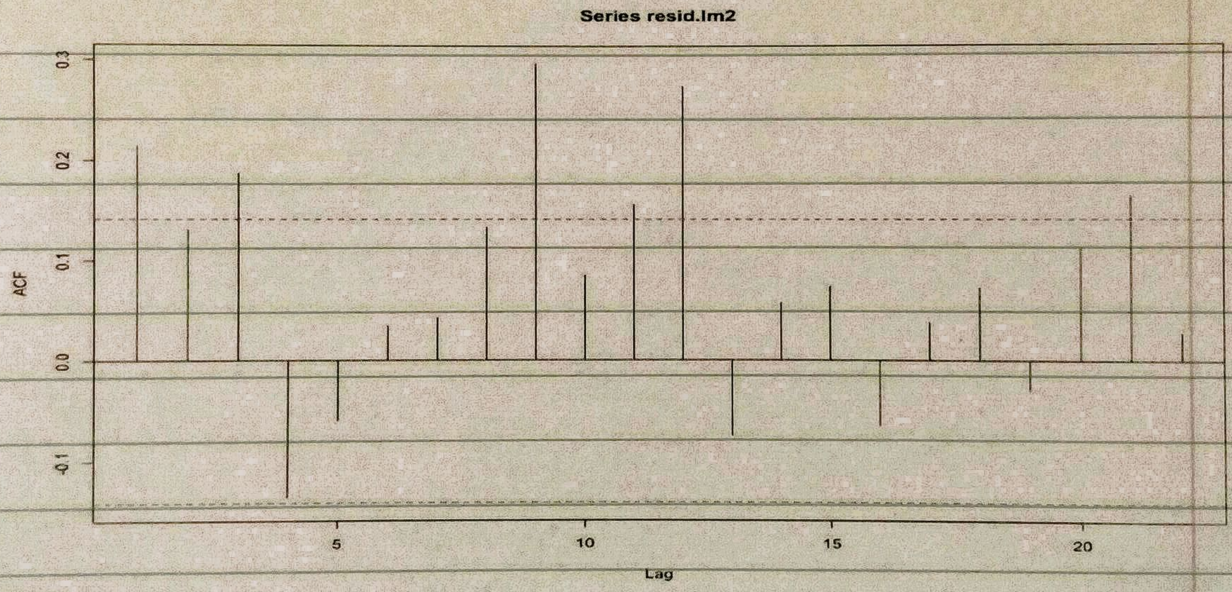
```
$pvalue
[1] 0.0127
$observed.runs
[1] 79
$expected.runs
[1] 96.625
$n1
[1] 90
$n2
[1] 102
$k
[1] 0
```

Finding:

Due to p value 0.0127 in runs test, we reject independence of the error terms, and conclude that neighboring residuals are positively dependent and tend to "hang together" over time

(c) Calculate and interpret the sample autocorrelations for the standardized residuals.

> acf(resid.lm2)



Finding:

Some of autocorrelations exceed two standard errors that also show the lack of independence in the error terms of this model.