Stat 349 Homework #2

February 12, 2020

Problem 1. Suppose $Y_t = \mu + \frac{1}{2}e_t - \frac{1}{2}e_{t-1}$. Find $Var(\overline{Y})$. Note any unusual results. In particular, compare your answer to what would have been obtained if $Y_t = \mu + e_t$. Describe the effect that the autocorrelation in Y has on $Var(\overline{Y})$.

$$\rho_1 = \frac{cov\left\{\frac{e_t - e_{t-1}}{2}, \frac{e_{t-1} - e_{t-2}}{2}\right\}}{\sqrt{Var\left(\frac{e_t - e_{t-1}}{2}\right) var\left(\frac{e_{t-1} - e_{t-2}}{2}\right)}} = \frac{-\frac{1}{4}\sigma^2}{\sigma^2} = -0.25$$

$\rho_k = 0, \quad k > 1$

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

So, we get

$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[1 + 2\left(1 - \frac{1}{n}\right)(-0.25)\right] = \frac{\gamma_0}{n}\left[1 - \frac{1}{2}\left(\frac{n-1}{n}\right)\right] \longrightarrow 0.5\frac{\gamma_0}{n}$$

when n usually occurring in time series.

when $Y_t = \mu + e_t$, we have

$$Var(\bar{Y}) = \frac{\gamma_0}{n}$$

we see that the negative correlation at lag 1 has improved the estimation of mean.

Problem 2. Suppose $Y_t = \mu + e_t + 2e_{t-1}$. Find $Var(\overline{Y})$. Compare your answer to what would have been obtained if $Y_t = \mu + e_t$. Describe the effect that the autocorrelation in Y has on $Var(\overline{Y})$.

$$\rho_1 = \frac{\mathrm{cov}\{e_t+2e_{t-1},\ e_{t-1}+2e_{t-2}\}}{\sqrt{Var(e_t+2e_{t-1})\ Var(e_{t-1}+2e_{t-2})}} = \frac{4\sigma^2}{5\sigma^2} = 0.8$$

$\rho_k = 0, \quad k > 1$

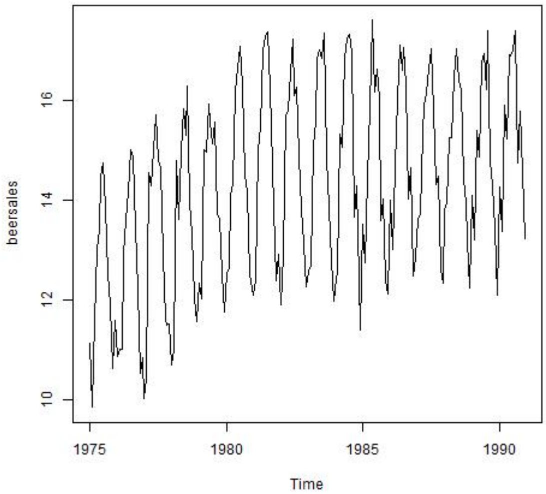
$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[1+2\left(1-\frac{1}{n}\right)0.8\right] = \frac{\gamma_0}{n}\left[1+1.6\left(\frac{n-1}{n}\right)\right]$$

$\longrightarrow 2.6\frac{\gamma_0}{n}$, when n usually occurring in time series.

we see that the positive correlation at lag 1 has reduced the estimation of mean.

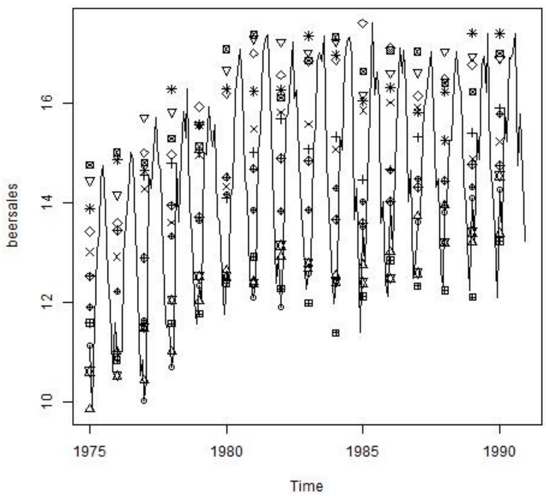
Problem 3

(a)



The plot shows that beer sales have a period of 12 months and have a slow increasing trend.

(b)



The overall trend is the same.

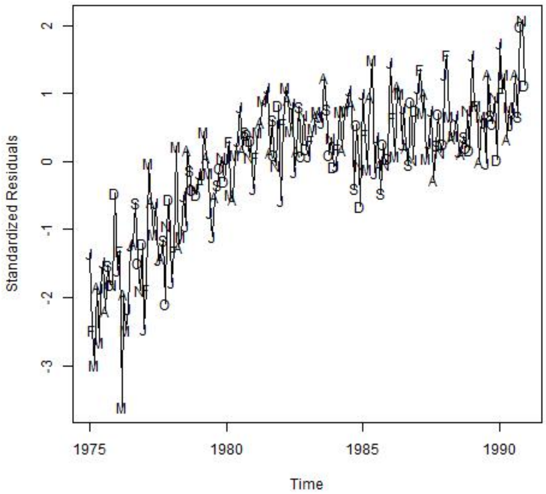
(c)

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
12.4857	12.3431	14.5679	14.8833	16.0846	16.3354

$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$
16.2543	16.0945	14.0585	13.7401	12.4377	12.0626

$\hat{\beta}$ represents month.

(d)



From the plot we know there is an increasing trend.

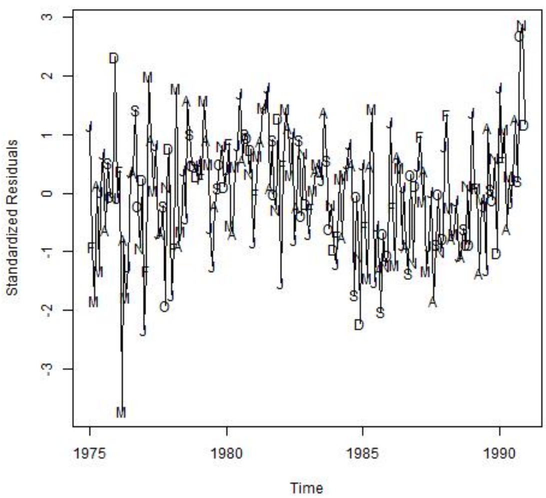
(e)

```
Coefficients:
                                Estimate
season(beersales)January      −7.150e+04
season(beersales)February     −7.150e+04
season(beersales)March        −7.150e+04
season(beersales)April        −7.150e+04
season(beersales)May          −7.149e+04
season(beersales)June         −7.149e+04
season(beersales)July         −7.149e+04
season(beersales)August       −7.149e+04
season(beersales)September    −7.150e+04
season(beersales)October      −7.150e+04
```

```
season(beersales)November    −7.150e+04
season(beersales)December    −7.150e+04
time(beersales)               7.196e+01
I(time(beersales)^2)         −1.810e−02
```

From the output, we can see the estimate of December seasonal mean becomes significant under 99% significant level after adding quadratic time trend. So the model is better now. However, one thing should be noticed that the parameter of quadratic term is negative, as the quadratic term grows faster, so the sales will be less than 0 in the future.

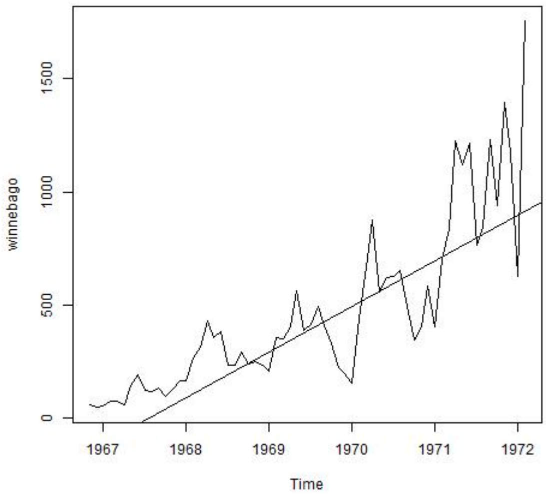
(f)



The standardized residuals graph has no clear pattern with time and does not remain seasonality, which is better than previous model.

Problem 4

(a)

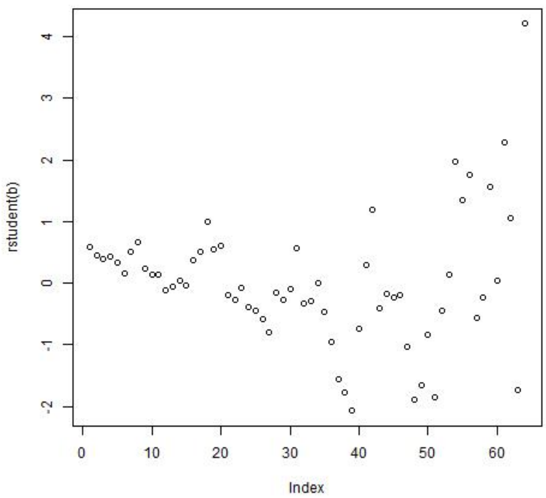


The plot shows that it has an exponential trend.

(b) $\hat{\beta}_1 = 200.74$, $\hat{\beta}_0 = -394885.68$

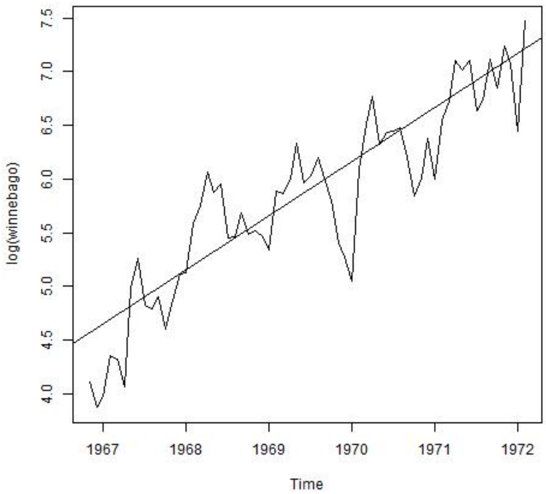
$\hat{\beta}_1$ is the slope of the line, which means when time lag is 1, it will be 200 more than previous.

$\hat{\beta}_0$ is the intersection of the line, which mean when time is 0, it is -394885 ideally.



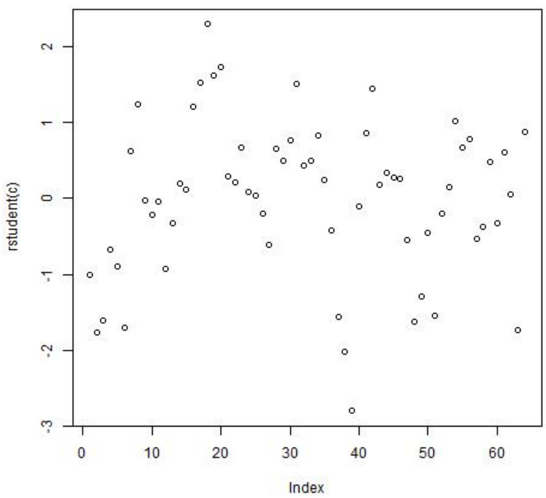
The plot shows that it has an increasing trend.

(c)



The graph shows that it seems have a linear increasing and cyclical trends.

(d)



The standardized residuals plot shows that it distributes near 0 randomly, which implies that linear regression may fit the logarithms well.

(e)

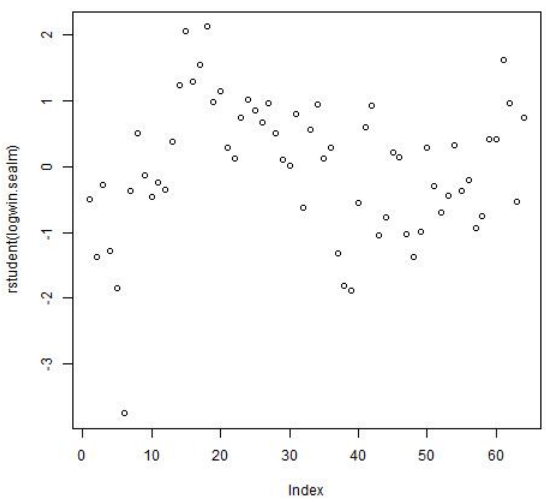
Coefficients:

	Estimate	**t** value	Pr(>|**t**|)	
season(winnebago)January	−997.33061	−19.70	<2e−16	***

season(winnebago)February	−996.70616	−19.68	<2e−16	***
season(winnebago)March	−996.64842	−19.68	<2e−16	***
season(winnebago)April	−996.52102	−19.68	<2e−16	***
season(winnebago)May	−996.46109	−19.68	<2e−16	***
season(winnebago)June	−996.46753	−19.68	<2e−16	***
season(winnebago)July	−996.77669	−19.68	<2e−16	***
season(winnebago)August	−996.76073	−19.68	<2e−16	***
season(winnebago)September	−996.75490	−19.68	<2e−16	***
season(winnebago)October	−997.06713	−19.69	<2e−16	***
season(winnebago)November	−997.04379	−19.69	<2e−16	***
season(winnebago)December	−997.08259	−19.69	<2e−16	***
time(**log**(winnebago))	0.50909	19.80	<2e−16	***

The regression coefficients are significant under 99.9% or 99% significant level.

(f)



This plot of residuals is better than results in (b) and (d), the variance is more equal compared to the former ones except some individual points. Also, there is not much seasonality remaining in the residuals.

Problem 5

(a)

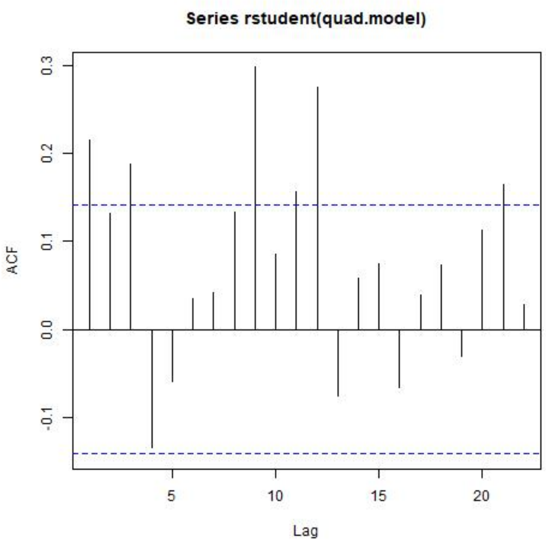
```
>rstudent(quad.model)
```

(b)

pvalue	observed.runs	expected.runs	n1	n2	k
0.0127	79	96.625	90	102	0

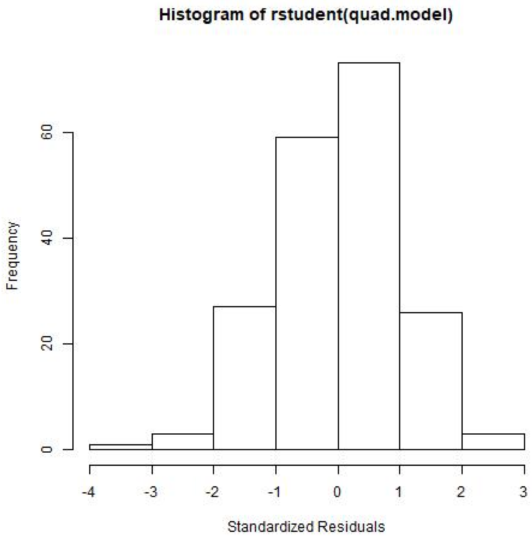
The p-value is $0.01 < 0.05$, so under 95% significance level, we can reject hypothesis that the error term is independent.

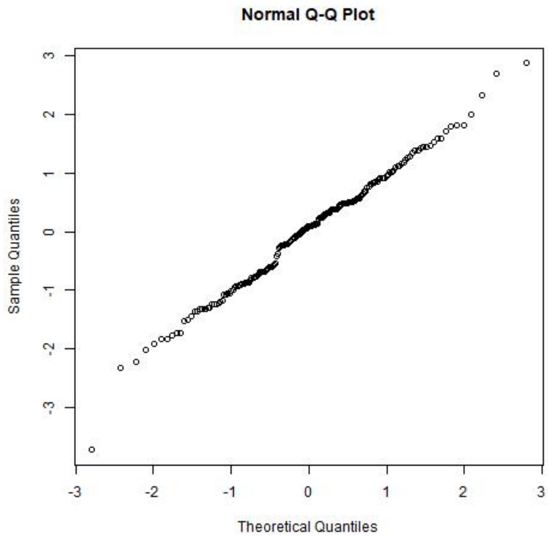
(c)



We can see autocorrelations are not 0, indicating the error term is dependent.

(d)





It is clear that the residuals are under the normal distribution.

Problem 6

(a)

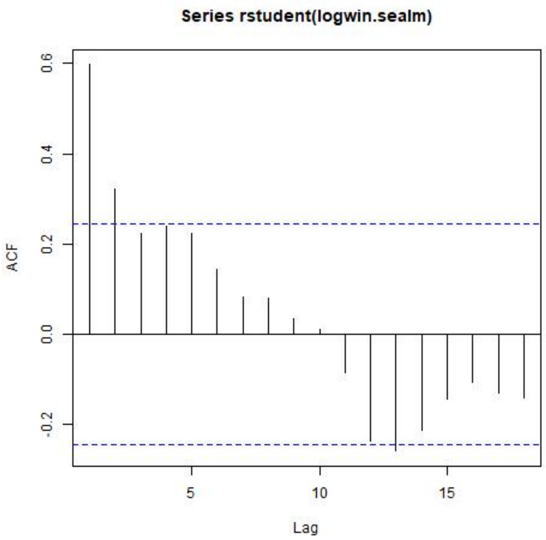
```
>rstandard(logwin.sealm)
```

(b)

pvalue	observed.runs	expected.runs	n1	n2	k
0.000243	18	32.71875	29	35	0

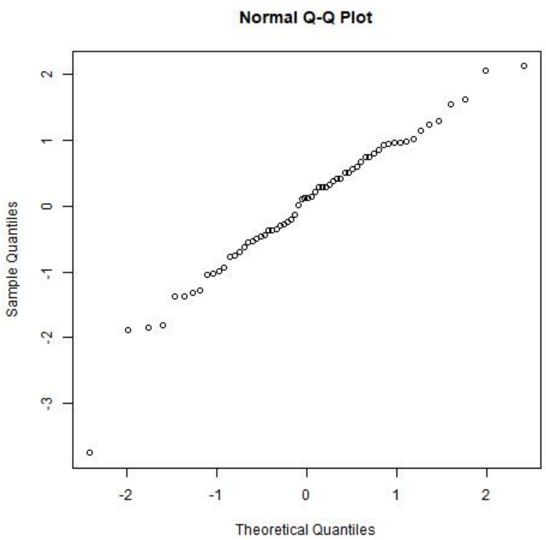
The p-value is $0.0002 < 0.05$, so under 95% significance level, we can reject hypothesis that the error term is independent.

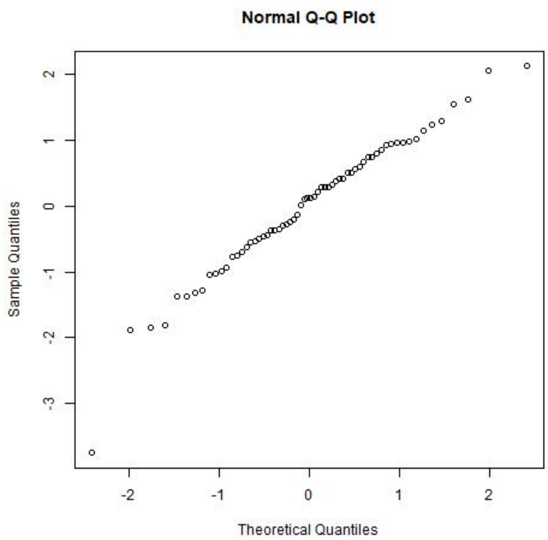
(c)



We can see autocorrelations are not 0, indicating the error term is dependent.

(d)





It is clear that the residuals are under the normal distribution.