Time Series Analysis

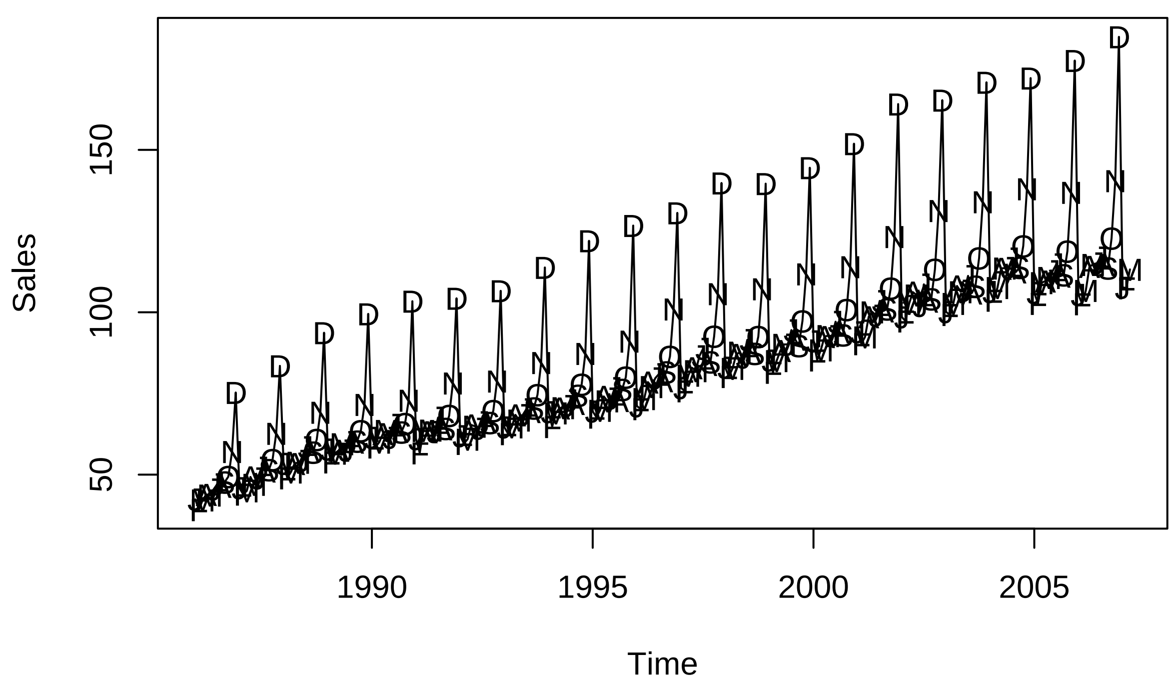
Homework of week 2

Hanbin Liu 11912410

3.8

(a)

```
data(retail)
plot(retail, type = 'l')
points(y = retail, x = time(retail), pch = as.vector(season(retail)))
```



As we would expect with retail sales in the U.K., there is substantial seasonality in the series. However, there is also a general upward "trend" with increasing variation at the higher levels of the series.

(b)

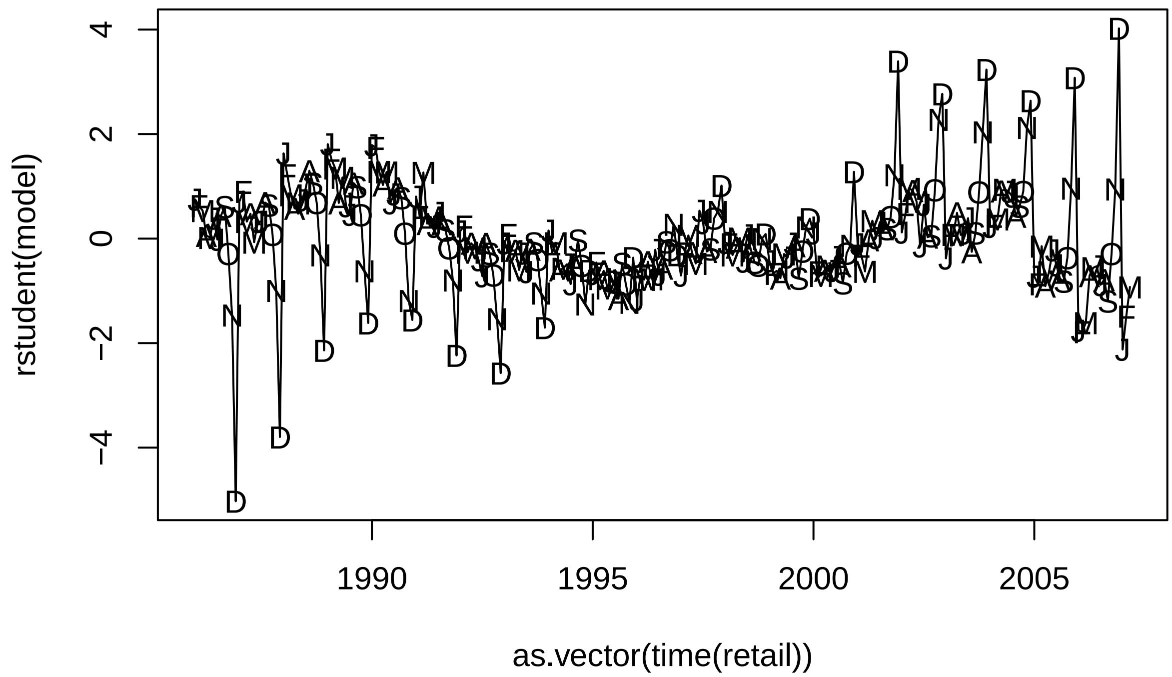
```r
month_ <- season(retail)
model <- lm(retail ~ month_ - 1 + time(retail))
summary(model)
```

```
##
## Call:
## lm(formula = retail ~ month_ - 1 + time(retail))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.8950  -2.4440  -0.3518   2.1971  16.2045
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## month_January   -7.249e+03  8.724e+01  -83.10   <2e-16 ***
## month_February  -7.252e+03  8.724e+01  -83.13   <2e-16 ***
## month_March     -7.249e+03  8.725e+01  -83.09   <2e-16 ***
## month_April     -7.246e+03  8.723e+01  -83.07   <2e-16 ***
## month_May       -7.246e+03  8.723e+01  -83.07   <2e-16 ***
## month_June      -7.246e+03  8.723e+01  -83.07   <2e-16 ***
## month_July      -7.243e+03  8.724e+01  -83.03   <2e-16 ***
## month_August    -7.246e+03  8.724e+01  -83.06   <2e-16 ***
## month_September -7.246e+03  8.725e+01  -83.05   <2e-16 ***
## month_October   -7.241e+03  8.725e+01  -82.99   <2e-16 ***
## month_November  -7.229e+03  8.725e+01  -82.85   <2e-16 ***
## month_December  -7.197e+03  8.726e+01  -82.48   <2e-16 ***
## time(retail)     3.670e+00  4.369e-02   84.00   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.278 on 242 degrees of freedom
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9978
## F-statistic:  8791 on 13 and 242 DF,  p-value: < 2.2e-16
```

About 99% of the variation in the retail series is explained by the regression model.

(c)

```r
plot(y = rstudent(model), x = as.vector(time(retail)), type = 'l')
points(y = rstudent(model), x = as.vector(time(retail)), pch = as.vector(season(retail)))
```



This residual plot is the best we have seen for models of this series but perhaps there are better models to be explored later.

3.14

(a)

```
month_ <- season(retail)
model <- lm(retail ~ month_ - 1 + time(retail))
```

(b)

```
runs(rstudent(model))
```

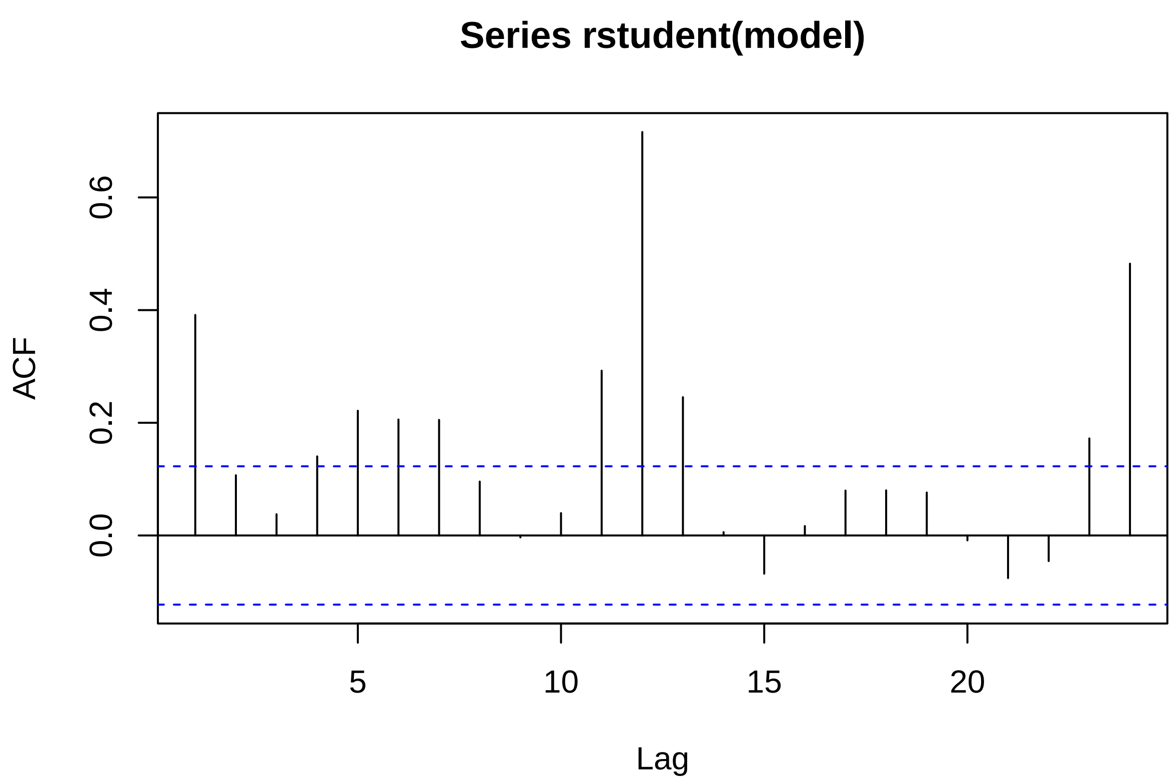
```
## $pvalue
## [1] 9.19e-23
##
## $observed.runs
## [1] 52
##
## $expected.runs
## [1] 127.9333
```

```
##
## $n1
## [1] 136
##
## $n2
## [1] 119
##
## $k
## [1] 0
```

These results suggest strongly that the error terms are not independent.

(c)

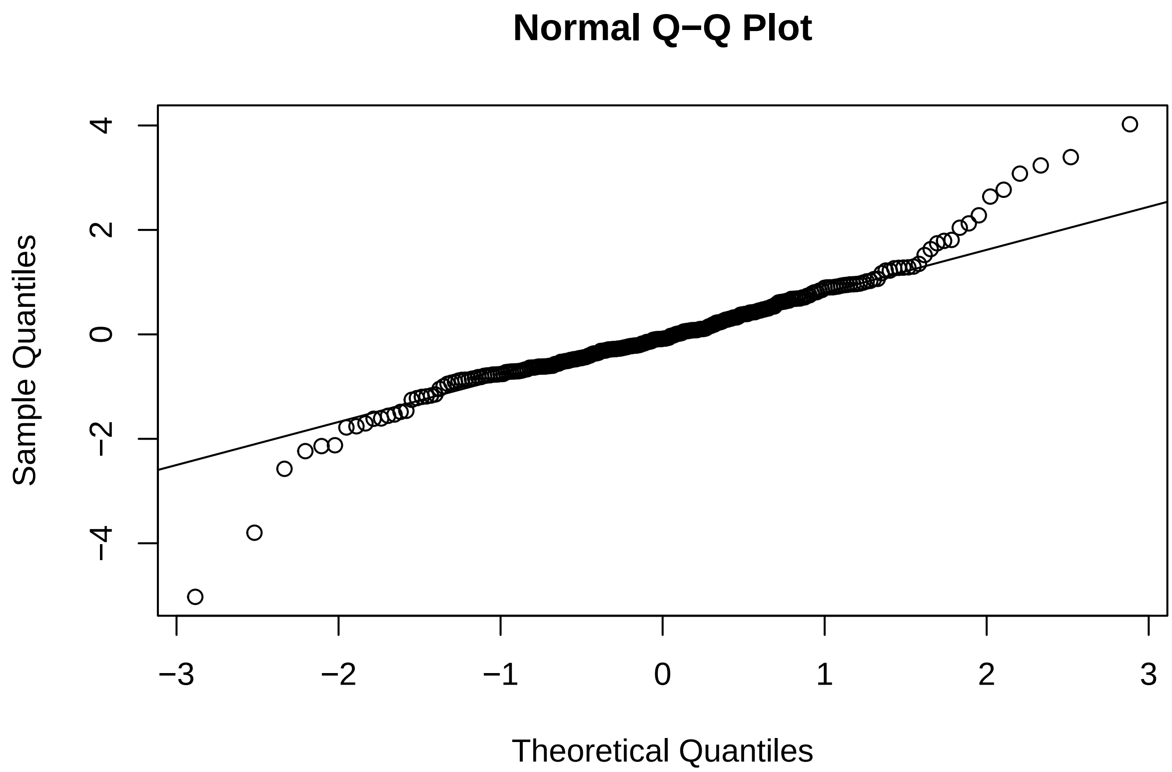
```
acf(rstudent(model))
```



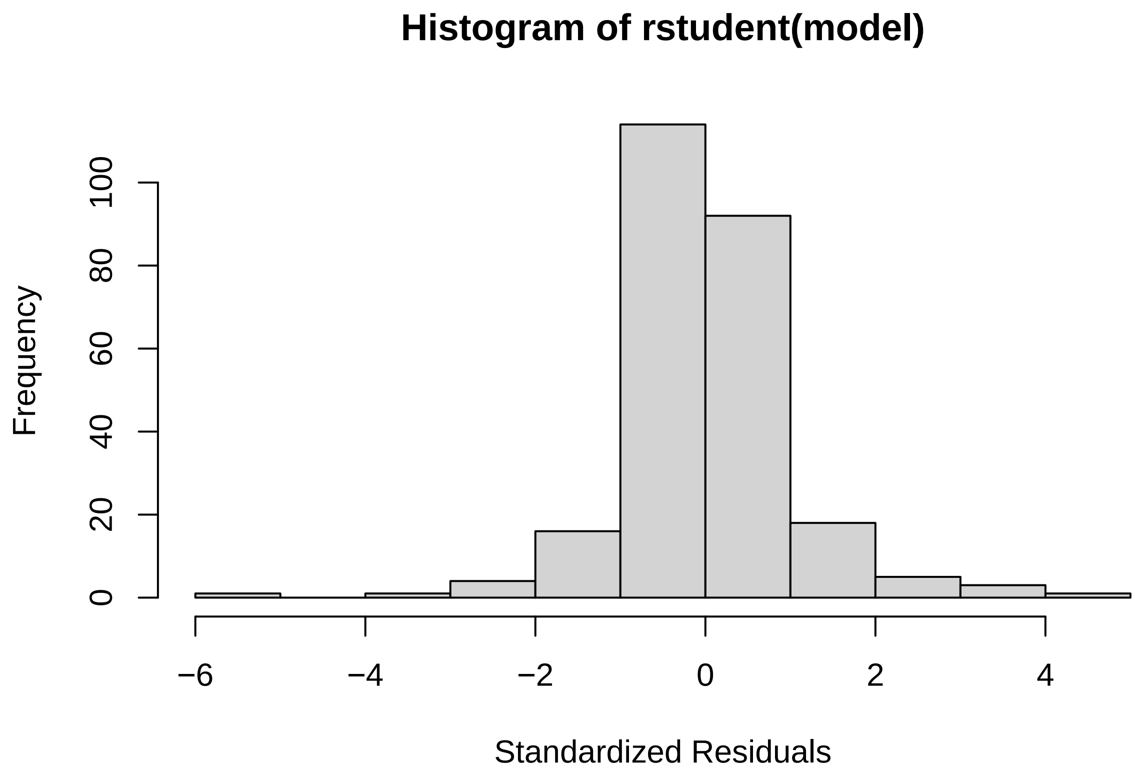
More evidence that the error terms are not independent. Especially at the lag 12.

(d)

```
qqnorm(rstudent(model))
qqline(rstudent(model))
```




```
hist(rstudent(model), xlab = 'Standardized Residuals')
```




```
shapiro.test(rstudent(model))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rstudent(model)
## W = 0.93897, p-value = 8.534e-09
```

There are many outliers on the low end and high end in Q-Q plot. It does not look like a normal distribution from the histogram. Further, the p-value < 0.05, thus we should reject the null hypothesis that the stochastic component of this model is normally distributed.

Additional problem

(1)

Let S denote $\sum_{t=1}^{T}(y_t - \beta_0 - \beta_t x_t)^2$. Since we want to minimize S, it then follows that

$$\begin{aligned}\frac{\partial S}{\partial \beta_0} &= 2\sum_{t=1}^{T}(y_t - \beta_0 - \beta_t x_t) \times (-1) &= 0\\ \frac{\partial S}{\partial \beta_1} &= 2\sum_{t=1}^{T}(y_t - \beta_0 - \beta_t x_t) \times (-x_t) &= 0\end{aligned}$$

Solving these two equations yileds that $\sum_{t=1}^T \hat{u}_t = 0$, $\sum_{t=1}^T x_t\hat{u}_t = 0$. Besides, $\bar{\hat{y}} = \sum_{t=1}^T \hat{y}_t/T = \sum_{t=1}^T (y_t - \hat{u}_t)/T = \sum_{t=1}^T y_t/T - \sum_{t=1}^T \hat{u}_t/T = \bar{y}$.

(2)

By using the fact $\sum_{t=1}^T \hat{u}_t = 0$ and $\sum_{t=1}^T x_t\hat{u}_t = 0$ in the conclusion of (1), we obtain $\mathrm{cov}(\hat{u}_t, \hat{y}_t) \approx 0$. So, $\mathrm{cov}(y_t, \hat{y}_t) = \mathrm{cov}(\hat{y}_t + \hat{u}_t, \hat{y}_t) \approx \mathrm{var}(\hat{y}_t)$. Therefore, $\rho^2(y_t, \hat{y}_t) = \frac{\mathrm{cov}^2(y_t,\hat{y}_t)}{\mathrm{var}(y_t)\mathrm{var}(\hat{y}_t)} \approx \frac{\mathrm{var}(\hat{y}_t)}{\mathrm{var}(y_t)}$. It then follows that

$$R^2 = \frac{(\sum_{t=1}^T (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}}))^2}{\sum_{t=1}^T (y_t - \bar{y})^2 \sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2} \approx \rho^2(y_t, \hat{y}_t) \approx \frac{\mathrm{var}(\hat{y}_t)}{\mathrm{var}(y_t)}.$$

Also, we have $R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} \approx \frac{\mathrm{var}(\hat{y}_t)}{\mathrm{var}(y_t)}$. Therefore, the two definition of R^2 are asymtotically equal.

*** Another method**

From the second definition of R^2, we have

$$\begin{aligned} R^2 &= \frac{[\sum_{t=1}^T (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})]^2}{\sum_{t=1}^T (y_t - \bar{y})^2 \cdot \sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2} \\ &\triangleq \frac{\mathrm{A}^2}{\mathrm{SST} \cdot \mathrm{SSR}}. \end{aligned}$$

Also, we have $R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = \frac{\mathrm{SSR}^2}{\mathrm{SST}\cdot\mathrm{SSR}}$. So it is sufficient to show that $\mathrm{SSR} = \mathrm{A}$. We consider their difference.

$$\begin{aligned} \mathrm{A} - \mathrm{SSR} &= \sum_{t=1}^T (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}}) - \sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2 \\ &= \sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})(y_t - \bar{y} - \hat{y}_t + \bar{\hat{y}}) \\ &= \sum_{t=1}^T (\hat{y}_t - \bar{y})(y_t - \hat{y}_t) \\ &= \sum_{t=1}^T (\hat{y}_t - \bar{y})\hat{u}_t \\ &= \sum_{t=1}^T (\hat{\beta}_0 + \hat{\beta}_1 x_t - \bar{y})\hat{u}_t \\ &= (\hat{\beta}_0 - \bar{y}) \cdot \sum_{t=1}^T \hat{u}_t + \hat{\beta}_1 \cdot \sum_{t=1}^T x_t\hat{u}_t \\ &= 0 + 0 = 0. \end{aligned}$$

Therefore, the two definition of R^2 are exactly equal.