Homework 2

Benjamin Noland

1. (Throughout, assume that $n > 1$, so that the expressions to be derived are well-defined). The ordinary least squares (OLS) estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1, respectively, are defined to minimize the loss

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} [Y_{t_i} - (\beta_0 + \beta_1 t_i)]^2.$$

First, we compute the stationary points of the loss:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta_0} Q(\beta_0, \beta_1) = \sum_{i=1}^{n} -2[Y_{t_i} - (\beta_0 + \beta_1 t_i)] \\ 0 &= \frac{\partial}{\partial \beta_1} Q(\beta_0, \beta_1) = \sum_{i=1}^{n} -2t_i[Y_{t_i} - (\beta_0 + \beta_1 t_i)]. \end{aligned}$$

Rearranging the first equation, we find that $\hat{\beta}_0$ satisfies

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} Y_{t_i} - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} t_i = \bar{Y} - \hat{\beta}_1 \bar{t},$$

where $\hat{\beta}_1$ is the OLS estimator of β_1, to be computed next. Rearranging the second equation, we find that

$$\begin{aligned} \sum_{i=1}^{n} t_i Y_{t_i} &= \sum_{i=1}^{n} t_i \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} t_i^2 \\ &= \sum_{i=1}^{n} t_i(\bar{Y} - \hat{\beta}_1 \bar{t}) + \hat{\beta}_1 \sum_{i=1}^{n} t_i^2 \\ &= \bar{Y}\sum_{i=1}^{n} t_i - \hat{\beta}_1 \bar{t} \sum_{i=1}^{n} t_i + \hat{\beta}_1 \sum_{i=1}^{n} t_i^2 \\ &= n\bar{Y}\bar{t} - n\hat{\beta}_1 \bar{t}^2 + \hat{\beta}_1 \sum_{i=1}^{n} t_i^2 \\ &= n\bar{Y}\bar{t} + \hat{\beta}_1 \left(\sum_{i=1}^{n} t_i^2 - n\bar{t}^2 \right), \end{aligned}$$

where we substituted the expression for $\hat{\beta}_0$ computed above. Therefore,

$$\sum_{i=1}^{n} t_i Y_{t_i} - n\bar{Y}\bar{t} = \hat{\beta}_1 \left(\sum_{i=1}^{n} t_i^2 - n\bar{t}^2 \right).$$

This equation can be rewritten

$$n\sum_{i=1}^{n} (Y_{t_i} - \bar{Y})(t_i - \bar{t}) = \hat{\beta}_1 n \sum_{i=1}^{n} (t_i - \bar{t})^2.$$

Thus, solving for $\hat{\beta}_1$, we get

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_{t_i} - \bar{Y})(t_i - \bar{t})}{\sum_{i=1}^n (t_i - \bar{t})^2}.$$

Next, we need to verify that the point $(\hat{\beta}_0, \hat{\beta}_1)$ is indeed the unique minimizer of the loss $Q(\beta_0, \beta_1)$. We will apply the second derivative test. The Hessian $H(\beta_0, \beta_1)$ of $Q(\beta_0, \beta_1)$, is given by

$$H(\beta_0, \beta_1) = \begin{pmatrix} \frac{\partial^2}{\partial \beta_0^2} Q(\beta_0, \beta_1) & \frac{\partial^2}{\partial \beta_0 \beta_1} Q(\beta_0, \beta_1) \\ \frac{\partial^2}{\partial \beta_1 \beta_0} Q(\beta_0, \beta_1) & \frac{\partial^2}{\partial \beta_1^2} Q(\beta_0, \beta_1) \end{pmatrix} = \begin{pmatrix} 2n & 2n\bar{t} \\ 2n\bar{t} & 2\sum_{i=1}^n t_i^2 \end{pmatrix}.$$

Thus $H(\beta_0, \beta_1)$ is constant in (β_0, β_1). To verify that $H(\beta_0, \beta_1)$ is positive definite, we note that

$$\frac{\partial^2}{\partial \beta_0^2} Q(\beta_0, \beta_1) = 2n > 0$$

and that

$$\det H(\beta_0, \beta_1) = 4n \sum_{i=1}^n t_i^2 - 4n^2 \bar{t} = 4n \sum_{i=1}^n (t_i - \bar{t})^2 > 0.$$

In particular, $H(\hat{\beta}_0, \hat{\beta}_1)$ is positive definite. Therefore $(\hat{\beta}_0, \hat{\beta}_1)$ is a local minimum of the loss $Q(\beta_0, \beta_1)$. However, since $H(\beta_0, \beta_1)$ is everywhere positive definite, $Q(\beta_0, \beta_1)$ is strictly convex in (β_0, β_1), and thus $(\hat{\beta}_0, \hat{\beta}_1)$ is the *unique* global minimizer of $Q(\beta_0, \beta_1)$.

2. (Cryer & Chan, Exercise 3.2) Let $\{e_t\}_{t=0}^\infty$ be a white noise process, and define the process $\{Y_t\}_{t=1}^\infty$ by $Y_t = \mu + e_t - e_{t-1}$. Let $Y_1, \dots, Y_n$ denote the observed time series. Then the sample mean is given by

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

Therefore,

$$\mathrm{Var}(\bar{Y}) = \mathrm{Var}(\mu + \frac{1}{n} e_n) = \frac{1}{n^2} \mathrm{Var}(e_n) = \frac{\sigma_e^2}{n^2},$$

where σ_e^2 denotes the variance of the white noise process. Note that under the model $Y_t = \mu + e_t$, the sample mean variance decreases linearly in n, i.e., $\mathrm{Var}(\bar{Y}) = \sigma_e^2 / n$ (see page 28). However, in this case the sample mean variance decreases quadratically in n.

3. (Cryer & Chan, Exercise 3.3) Let $\{e_t\}_{t=0}^\infty$ be a white noise process, and define the process $\{X_t\}_{t=1}^\infty$ by $X_t = (e_t + e_{t-1})/2$. Then $\{X_t\}$ is a moving average process, and let γ_k denote the autocovariance function for this process. Define the process $\{Y_t\}_{t=1}^\infty$ by $Y_t = \mu + e_t + e_{t-1} = \mu + 2X_t$. Let $Y_1, \dots, Y_n$ denote the observed time series. Then the sample mean is given by

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t = \frac{1}{n} \sum_{t=1}^n (\mu + 2X_t) = \mu + \frac{2}{n} \sum_{t=1}^n X_t.$$

Therefore, recalling the autocovariance structure of $\{X_t\}$ (see page 15), we have

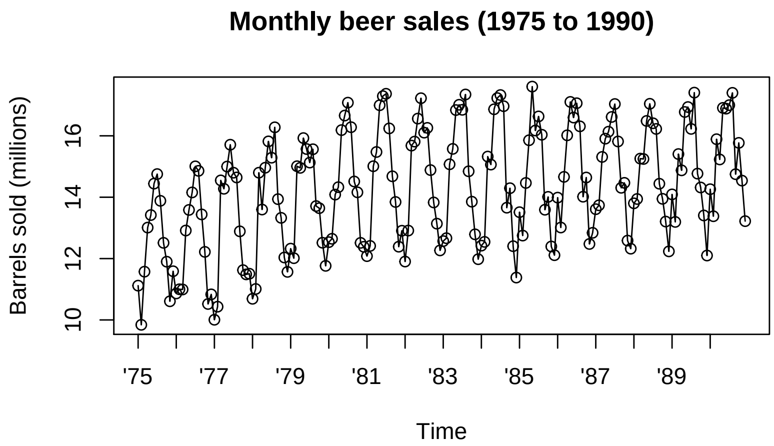
$$
\begin{aligned}
\operatorname{Var}(\bar{Y}) &= \operatorname{Var}\left(\mu + \frac{2}{n}\sum_{t=1}^{n} X_t\right) \\
&= \frac{4}{n^2}\operatorname{Var}\left(\sum_{t=1}^{n} X_t\right) \\
&= \frac{4}{n^2}\left(\sum_{t=1}^{n}\operatorname{Var}(X_t) + 2\sum_{s=2}^{n}\sum_{t=1}^{s-1}\operatorname{Cov}(X_s, X_t)\right) \\
&= \frac{4}{n^2}\left(\sum_{t=1}^{n}\gamma_0 + 2\sum_{s=2}^{n}\sum_{t=1}^{s-1}\gamma_{|t-s|}\right) \\
&= \frac{4}{n^2}\left(n\frac{1}{2}\sigma_e^2 + 2\sum_{s=2}^{n}\frac{1}{4}\sigma_e^2\right) \\
&= \frac{4}{n^2}\left(\frac{n\sigma_e^2}{2} + \frac{(n-1)\sigma_e^2}{2}\right) \\
&= \frac{4\sigma_e^2}{n} - \frac{2\sigma_e^2}{n^2},
\end{aligned}
$$

where σ_e^2 denotes the variance of the white noise process. Note that under the model $Y_t = \mu + e_t$, the sample mean variance decreases linearly in n, i.e., $\operatorname{Var}(\bar{Y}) = \sigma_e^2/n$ (see page 28). However, from the above calculation we see that the autocovariance structure of the model in question introduces a term that decays quadratically in n.

4. (Cryer & Chan, Exercise 3.6)

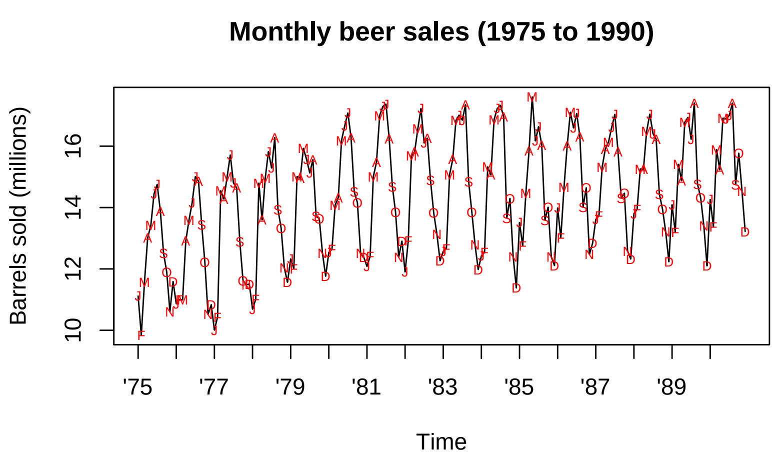
   a. The time series plot (displayed below) seems to indicate a seasonal trend in monthly beer sales. In addition, there appears to be an upward trend in monthly beer sales averaged over a complete year from approximately 1975 to 1980.

```r
plot(
  x = beersales,
  type = "o",
  main = "Monthly beer sales (1975 to 1990)",
  ylab = "Barrels sold (millions)",
  xaxt = "n"  # Suppress default x axis.
)
axis(1, at = 1975:1990, labels = paste0("'", 75:90))
```



b. Monthly beer sales appear to be lower in the colder months than in the warmer months, with the lowest sales in the middle of the winter and the highest in the middle of the summer. Thus, as mentioned in part (a), there *does* appear to be a seasonal trend in the data.

```
plot(
  x = beersales,
  type = "l",
  main = "Monthly beer sales (1975 to 1990)",
  ylab = "Barrels sold (millions)",
  xaxt = "n"   # Suppress default x axis.
)
points(
  y = beersales,
  x = time(beersales),
  pch = as.vector(season(beersales)),
  cex = 0.6,
  col = "red"
)
axis(1, at = 1975:1990, labels = paste0("'", 75:90))
```



c. The following R code fits a seasonal-means model to the data, i.e., a linear model with a separate dummy variable for each month:

```r
month. <- season(beersales)
model <- lm(beersales ~ month. - 1)
summary(model)
```

```
##
## Call:
## lm(formula = beersales ~ month. - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5745 -0.4772  0.1759  0.7312  2.1023
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## month.January     12.4857     0.2639   47.31   <2e-16 ***
## month.February    12.3431     0.2639   46.77   <2e-16 ***
## month.March       14.5679     0.2639   55.20   <2e-16 ***
## month.April       14.8833     0.2639   56.39   <2e-16 ***
## month.May         16.0846     0.2639   60.95   <2e-16 ***
## month.June        16.3354     0.2639   61.90   <2e-16 ***
## month.July        16.2543     0.2639   61.59   <2e-16 ***
## month.August      16.0945     0.2639   60.98   <2e-16 ***
## month.September   14.0585     0.2639   53.27   <2e-16 ***
## month.October     13.7401     0.2639   52.06   <2e-16 ***
## month.November    12.4377     0.2639   47.13   <2e-16 ***
## month.December    12.0626     0.2639   45.71   <2e-16 ***
## ---
## 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:  0.995,  Adjusted R-squared:  0.9946
## F-statistic:  2964 on 12 and 180 DF,  p-value: < 2.2e-16
```

Each of the estimated coefficients in the model is the estimated mean beer sales for the corresponding month (over all the years in the data set). As expected given the plot from part (b), the coefficients for the warmer months are generally higher than those for the colder months, so that according to the model, average monthly beer sales tend to be higher in the warmer months than in the colder months. Furthermore, we have $R^2 = 0.995$, so that the model explains about 99.5% of the variation in the data.

The following R code extracts the Studentized residuals from the model:

```r
resid <- rstudent(model)
head(resid)
```

```
##         1         2         3         4         5         6
## -1.341099 -2.482410 -2.993870 -1.845174 -2.652005 -1.865308
```

e. The following R code fits a seasonal-means plus quadratic time trend to the data, i.e., a linear model with a separate dummy variable for each month, and a quadratic time term:

```r
month. <- season(beersales)
time <- time(beersales)
model <- lm(beersales ~ month. + I(time^2) - 1)
summary(model)
```

```
##
## Call:
## lm(formula = beersales ~ month. + I(time^2) - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.48235 -0.51086 -0.02017  0.51451  1.36884
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## month.January   -1.542e+02  1.071e+01  -14.40   <2e-16 ***
## month.February  -1.543e+02  1.071e+01  -14.42   <2e-16 ***
## month.March     -1.521e+02  1.071e+01  -14.21   <2e-16 ***
## month.April     -1.518e+02  1.071e+01  -14.18   <2e-16 ***
## month.May       -1.506e+02  1.071e+01  -14.07   <2e-16 ***
## month.June      -1.504e+02  1.071e+01  -14.04   <2e-16 ***
## month.July      -1.505e+02  1.071e+01  -14.05   <2e-16 ***
## month.August    -1.507e+02  1.071e+01  -14.07   <2e-16 ***
## month.September -1.527e+02  1.071e+01  -14.26   <2e-16 ***
## month.October   -1.531e+02  1.071e+01  -14.29   <2e-16 ***
## month.November  -1.544e+02  1.071e+01  -14.41   <2e-16 ***
## month.December  -1.548e+02  1.072e+01  -14.44   <2e-16 ***
## I(time^2)        4.241e-05  2.723e-06   15.57   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6899 on 179 degrees of freedom
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9977
## F-statistic:  6425 on 13 and 179 DF,  p-value: < 2.2e-16
```

For a fixed month, the quadratic time term in the model describes the relationship between year and beer sales, with the coefficient of the dummy variable for the corresponding month playing the role of the intercept term. Thus we see that according to the model, for a fixed month, monthly beer sales increase (slightly) with year. Furthermore, we have $R^2 = 0.9979$, so that the model explains about 99.8% of the variation in the data.

The following R code extracts the Studentized residuals from the model:

```
resid <- rstudent(model)
head(resid)
```

```
##          1          2          3          4          5          6
## -0.1627371 -1.8866332 -2.6605871 -0.9253646 -2.1425396 -0.9555924
```