Homework 2 Solution

February 24, 2019

Problem 1. The OLS estimator of β_0 and β_1 is given by minimizing the following equation:

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

Then we calculate the the partial derivatives of $L(\beta_0, \beta_1)$:

$$\begin{aligned}
\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_0} &= \sum_{i=1}^{n} -2(Y_i - \beta_0 - \beta_1), \\
\frac{\partial L(\beta_0, \beta_1)}{\partial \beta_1} &= \sum_{i=1}^{n} -2t_i(Y_i - \beta_0 - \beta_1).
\end{aligned}$$

Make the two partial derivatives to 0 and do some algebra we get the result:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(t_i - \bar{t})}{\sum_{i=1}^{n} (t_i - \bar{t})^2}, \\
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{t}.
\end{aligned}$$

The problem is not done here because we need to check that this is the global minimum. To show this we just calculate the Hessian matrix and show that it is positive definite(Or equivalently, the multivariate second-order test).

Problem 3.2. Denote the variance of e_t as σ_e^2

$$\begin{aligned}
\bar{Y} &= \mu + \frac{1}{n} \sum_{t=1}^{n} (e_t - e_{t-1}) = \mu + \frac{1}{n}(e_n - e_0), \\
\mathrm{Var}(\bar{Y}) &= \frac{1}{n^2} \mathrm{Var}(e_n - e_0) = \frac{2}{n^2} \sigma_e^2.
\end{aligned}$$

The denominator of n^2 is unusual because we normally see a denominator of n in the variance of a sample mean. In this setting, the negative autocorrelation at lag one makes the process mean has a smaller order compared with those process that has independent component.

Problem 3.3.

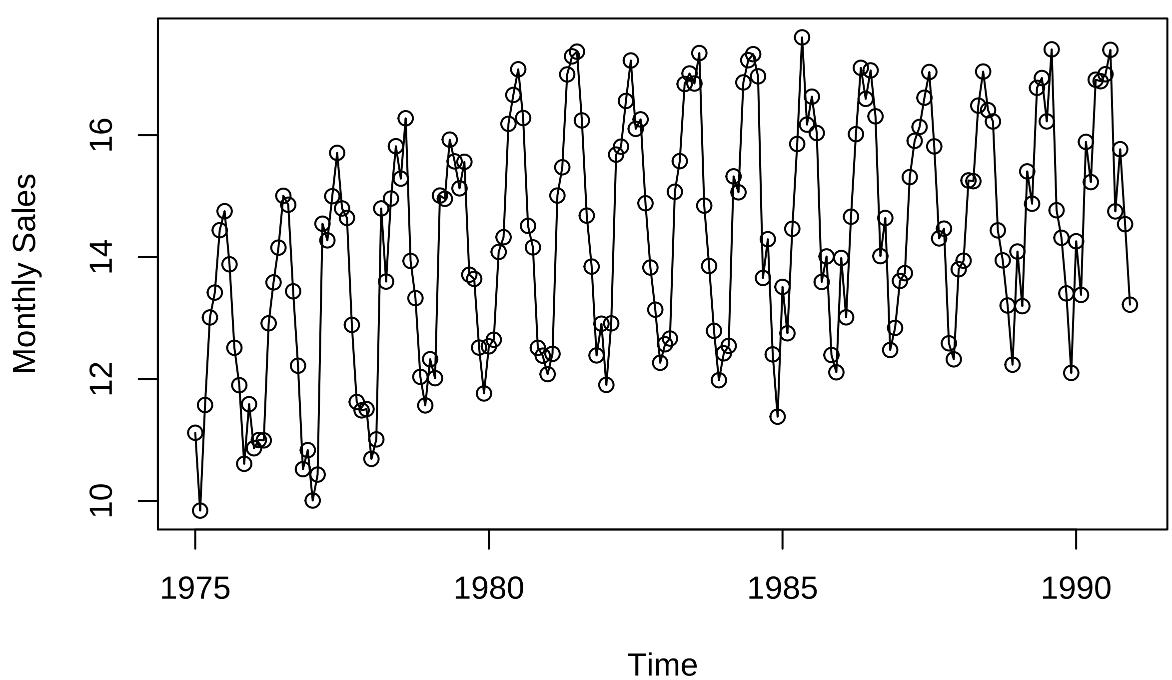
$$\begin{aligned}
\bar{Y} &= \mu + \frac{1}{n}(e_0 + e_n + 2\sum_{t=1}^{n-1} e_t), \\
\mathrm{Var}(\bar{Y}) &= \frac{1}{n^2}[\sigma_e^2 + \sigma_e^2 + 4(n-1)\sigma_e^2], \\
&= \frac{(4n-2)\sigma_e^2}{n^2}.
\end{aligned}$$

For a MA(1) process $Y_t = \mu + e_t$, the variance of the mean is $\frac{1}{n}\sigma_e^2$. In the present case, $\mathrm{Var}(\bar{Y}) \approx \frac{4}{n}\sigma_e^2$, and this is because of the positive autocorrelation at lag one.

Homework 2 Solution

(a)

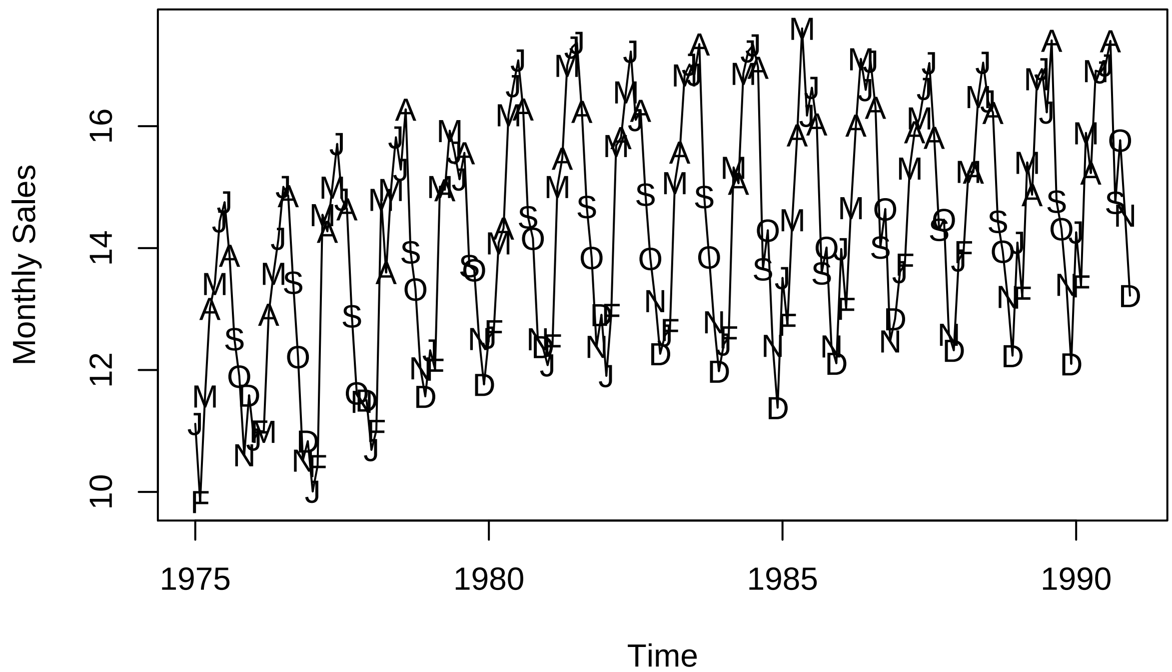
```
library(TSA)
data(beersales)
plot(beersales,ylab="Monthly Sales",type="o")
```



In addition to a possible seasonality in the series,there is a general upward "trend" in the first part of the series. However, this effect "levels off" in the latter years.

(b)

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



Now the seasonality is quite clear with higher sales in the summer months and lower sales in the winter.

(c)

```
month.=season(beersales)
beersales.lm=lm(beersales~month.)
summary(beersales.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
```

This model leaves out the January term so all of the other effects are in comparison to January. The multiple R-squared is rather large at 71% and all the terms except November, December, and February are significant different from January.

(e)

```r
beersales.lm2=lm(beersales~month.+time(beersales)+I(time(beersales)^2))
summary(beersales.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
```

This model seems to do a better job than the seasonal means alone but we should reserve judgement until we look at the residuals.