# STAT4870 HW3

Anton Yang

2024-09-19

## Problem 1a

```
library(astsa)
year<-time(jj) - 1970 #center the time at 1970
quarter<-factor(cycle(jj))
model1<-lm(log(jj) ~ 0 + year + quarter, na.action = NULL)
summary(model1)
```

```
##
## Call:
## lm(formula = log(jj) ~ 0 + year + quarter, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## year     0.167172   0.002259   74.00   <2e-16 ***
## quarter1 1.052793   0.027359   38.48   <2e-16 ***
## quarter2 1.080916   0.027365   39.50   <2e-16 ***
## quarter3 1.151024   0.027383   42.03   <2e-16 ***
## quarter4 0.882266   0.027412   32.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

According to the summary, our model $x_t = 0.167172t + 1.052793Q_1(t) + 1.080916Q_2(t) + 1.151024Q_3(t) + 0.882266Q_4(t) + w_t$, and we can see that our model produces very high $R^2 = 0.9935$ and low residual standard error of 0.1254.

## Problem 1b.

According to the summary, the estimated average annual increase in the logged earnings per share is the sum of each quarter coefficient, so it will be $1.052793 + 1.080916 + 1.151024 + 0.882266 = 4.166999$.

### Problem 1c.

According to the summary, the average logged earnings rate decrease by -0.268758 ($\alpha_4 - \alpha_3 = -0.268758$).
$\frac{\alpha_4 - \alpha_3}{\alpha_3} = \frac{0.882266 - 1.151024}{1.151024} = -0.2334947$
This means that there is a decrease in the average logged earnings from the third quarter to the fourth quarter by 23.34947%.

### Problem 1d.

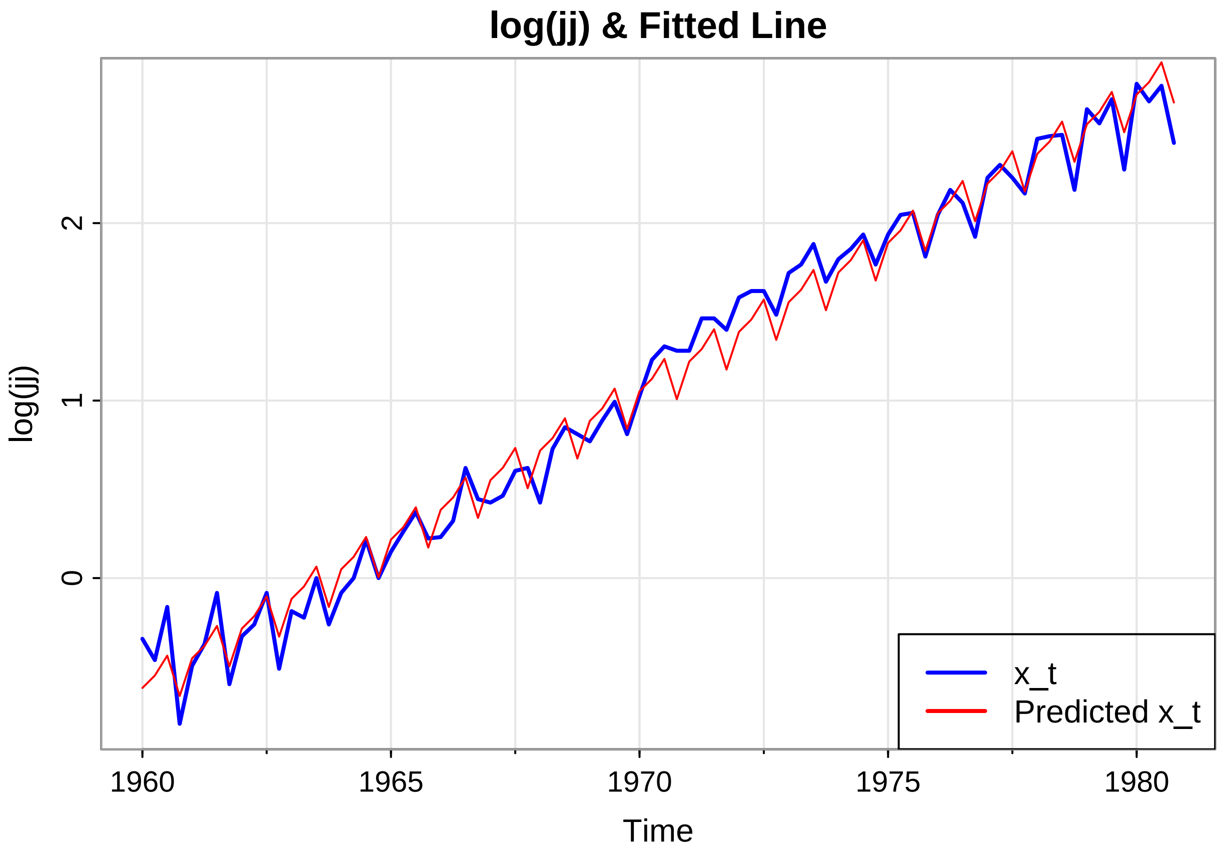
```
model2<-lm(log(jj) ~ year + quarter)
summary(model2)
```

```
##
## Call:
## lm(formula = log(jj) ~ year + quarter)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.052793   0.027359  38.480  < 2e-16 ***
## year         0.167172   0.002259  73.999  < 2e-16 ***
## quarter2     0.028123   0.038696   0.727   0.4695
## quarter3     0.098231   0.038708   2.538   0.0131 *
## quarter4    -0.170527   0.038729  -4.403 3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

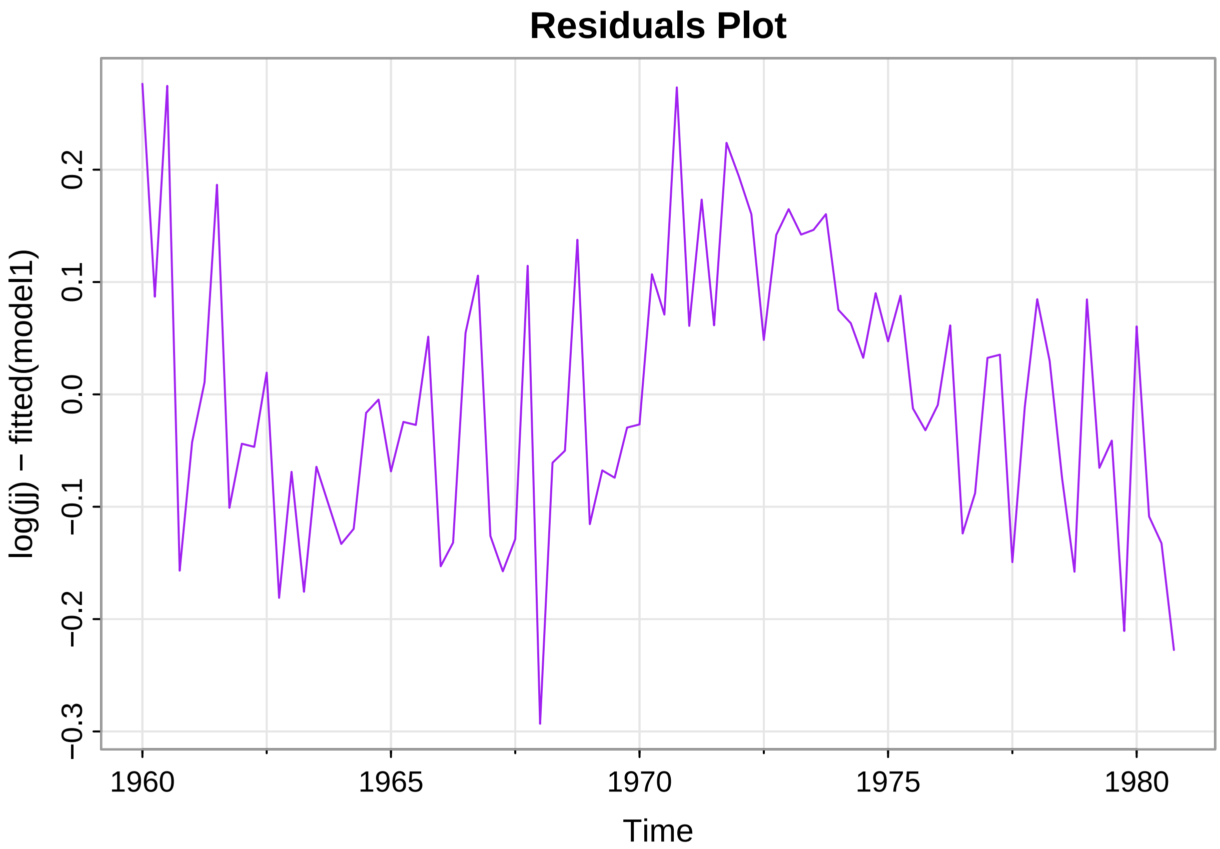
According to the summary, the intercept removed the first quarter and affects the significant and coefficient for the rest of the quarter. This is because our goal is to see the effect of each quarter, and intercept disrupt the interpretability of this model.

### Problem 1e.

```
tsplot(log(jj), main = "log(jj) & Fitted Line", col = "blue", lwd = 2)
lines(fitted(model1), col = "red")
legend("bottomright", legend = c("x_t", "Predicted x_t"), col = c("blue", "red"), lwd = 2)
```




```
tsplot(log(jj) - fitted(model1), main = "Residuals Plot", col = "purple")
```




According to the Residuals Plot, we can see that the noise seems to be random and the residuals looks white and the fit and the model fits well.

### Problem 2a.

```
temp <- tempr - mean(tempr)
temp2 <- temp^2
trend <- time(cmort)

n <- length(tempr)

model1 <- lm(cmort ~ trend + temp + temp2 + part, na.action = NULL)
model2 <- lm(cmort[5:n] ~ trend[5:n] + temp[5:n] + temp2[5:n] + part[5:n] + part[1:(n-4)], na.action =

summary(model1)
```

```
##
## Call:
## lm(formula = cmort ~ trend + temp + temp2 + part, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.0760  -4.2153  -0.4878   3.7435  29.2448
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.831e+03  1.996e+02   14.19  < 2e-16 ***
## trend       -1.396e+00  1.010e-01  -13.82  < 2e-16 ***
## temp        -4.725e-01  3.162e-02  -14.94  < 2e-16 ***
## temp2        2.259e-02  2.827e-03    7.99 9.26e-15 ***
## part         2.554e-01  1.886e-02   13.54  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.385 on 503 degrees of freedom
## Multiple R-squared:  0.5954, Adjusted R-squared:  0.5922
## F-statistic:   185 on 4 and 503 DF,  p-value: < 2.2e-16
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = cmort[5:n] ~ trend[5:n] + temp[5:n] + temp2[5:n] +
##     part[5:n] + part[1:(n - 4)], na.action = NULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.228  -4.314  -0.614   3.713  27.800
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.808e+03  1.989e+02  14.123  < 2e-16 ***
## trend[5:n]      -1.385e+00  1.006e-01 -13.765  < 2e-16 ***
## temp[5:n]       -4.058e-01  3.528e-02 -11.503  < 2e-16 ***
## temp2[5:n]       2.155e-02  2.803e-03   7.688 8.02e-14 ***
## part[5:n]        2.029e-01  2.266e-02   8.954  < 2e-16 ***
## part[1:(n - 4)]  1.030e-01  2.485e-02   4.147 3.96e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.287 on 498 degrees of freedom
## Multiple R-squared:  0.608,  Adjusted R-squared:  0.6041
## F-statistic: 154.5 on 5 and 498 DF,  p-value: < 2.2e-16
```

According to the summary, the predictors are all statistically significant. We can see that the adjusted $R^2$ for the new model is higher the original model. Therefore, the extra predictor contribute significantly to the model. Our model is: $\hat{M} = 2809 + 1.385t - 0.4058(T_t - T) + 0.02155(T_t - T)^2 + 0.2029P_t + 0.01030P_{t-4} + w_t$

## Problem 2b.

```r
merge(AIC(model1, model2), BIC(model1, model2), by = 'row.names', all = TRUE)
```

```
## Warning in AIC.default(model1, model2): models are not all fitted to the same
## number of observations
```

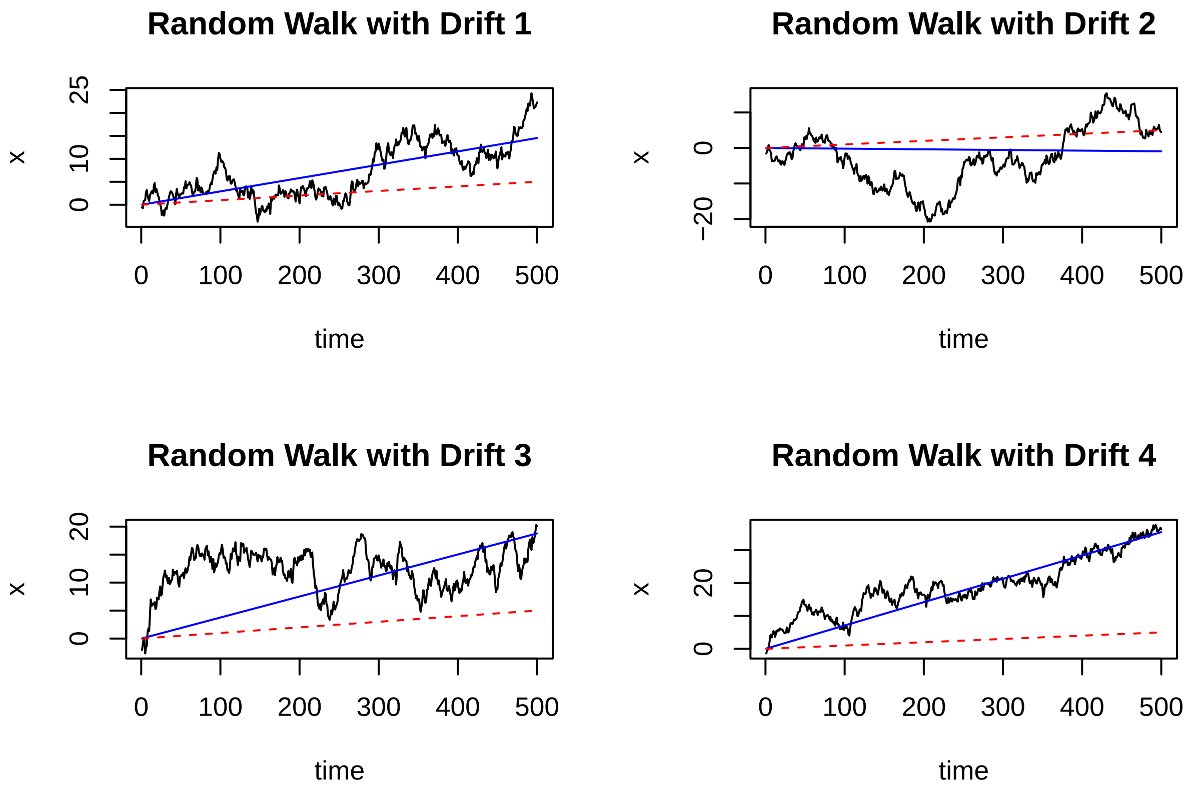
```
## Warning in BIC.default(model1, model2): models are not all fitted to the same
## number of observations
```

```
##   Row.names df.x      AIC df.y      BIC
## 1    model1    6 3332.282    6 3357.664
## 2    model2    7 3291.520    7 3321.078
```

We can see that the new model has a lower AIC and BIC than the original model. Therefore, model2 is an improvement over the original model.

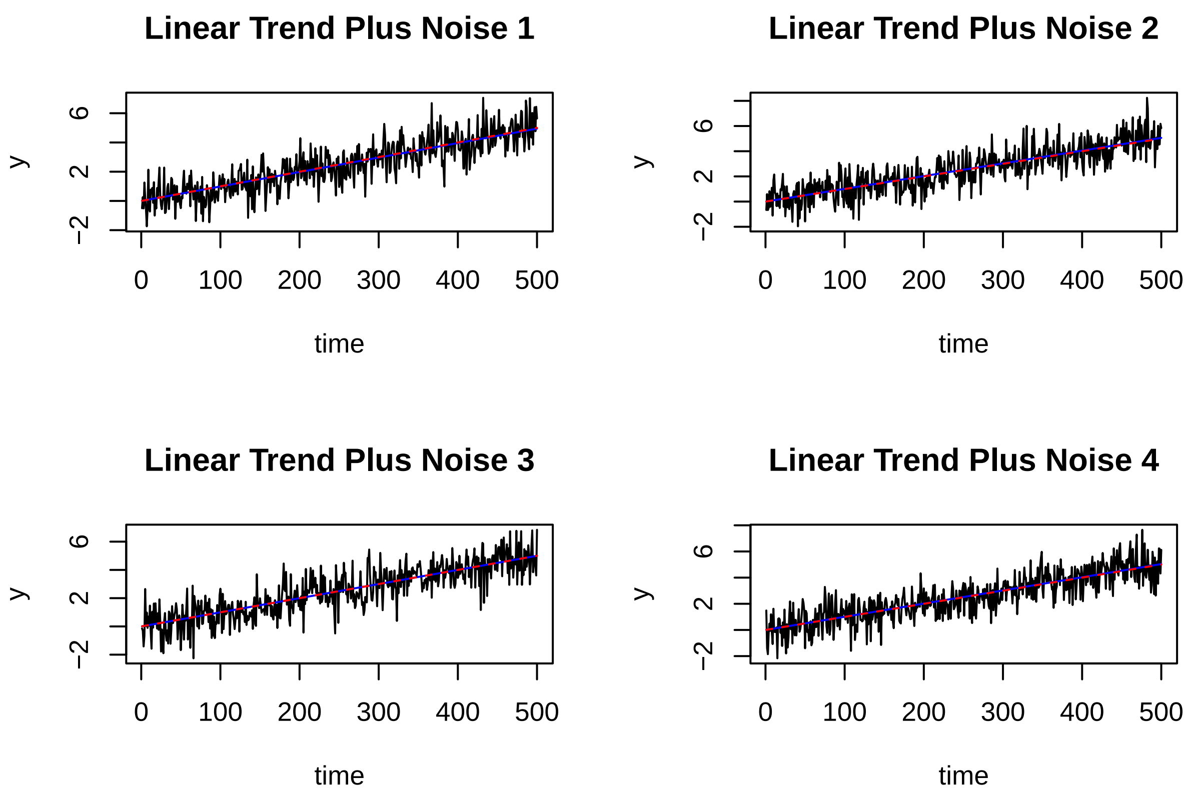
## Problem 3a.

```
set.seed(123)
n<-500
drift<-0.01
time<-1:n

par(mfrow = c(2,2))

for(i in 1:4){
  w<-rnorm(n)
  x<-c()
  for(t in time){
    x[t] = drift * t + sum(w[i:t])
  }
  mean_function<-drift*time
  model1<-lm(x ~ 0 + time, na.action = NULL)
  plot(time, x, type = "l", main = paste("Random Walk with Drift",i))
  lines(time, fitted(model1), col = "blue")
  lines(time, mean_function, col = "red", lty =2)
}
```




**Problem 3b.**

```
par(mfrow = c(2,2))
for (i in 1:4){
  w2<-rnorm(n)
  y<-drift*time + w2
  mean_function2<-drift*time
  model2<-lm(y ~ 0 + time)
  plot(time, y, type = "l", main = paste("Linear Trend Plus Noise", i))
  lines(time, fitted(model2), col = "blue")
  lines(time, mean_function2, col = "red", lty = 2)
}
```




## Problem 3c.

There's a clear difference in distance between the fitted line and the true mean in (b). We can see that (b) is significantly closer compared to (a). This is due to the errors in $y_t$ which are independent and is one of the assumptions of the linear regression, and the error in $x_t$ are correlated due to the accumulation of the white noise.