Homework 1

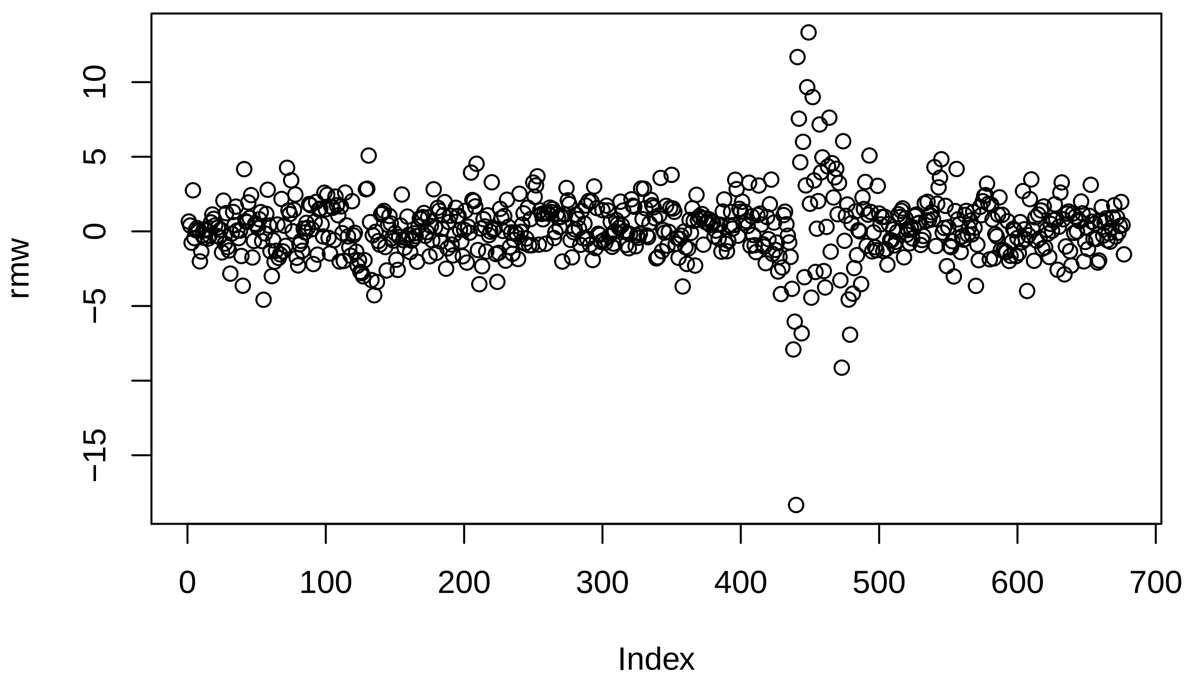
Problem 1

Question 1

```
ff5 <- read_csv("F-F_Research_Data_5_Factors_2x3.CSV", skip = 3)
```

```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   `Mkt-RF` = col_character(),
##   SMB = col_character(),
##   HML = col_character(),
##   RMW = col_character(),
##   CMA = col_character(),
##   RF = col_character()
## )
```

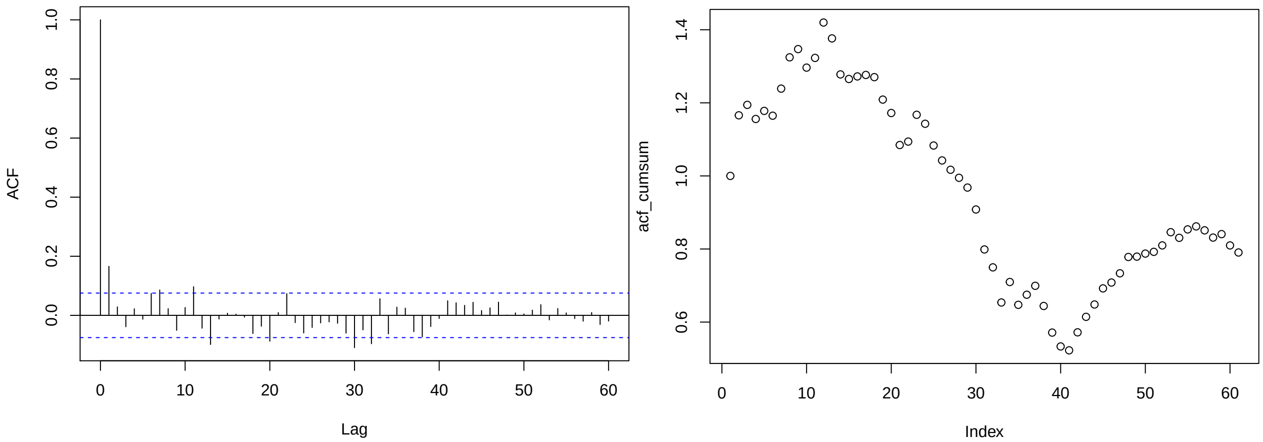
```
rmw <- ff5[1:677, 5]
rmw <- as.numeric(rmw$RMW)
plot(rmw)
```




```
rmw_annual_mean <- mean(rmw)*12
rmw_annual_sd <- sd(rmw)*sqrt(12)
```

Question 2

```r
par(mar=c(4,4,0.2,0.1))
acf_ac <- acf(rmw, lag.max = 60)
acf_cumsum <- cumsum(acf_ac$acf)
plot(acf_cumsum)
```



The plots DO hint some predictabilty of the factor returns. Specifically, the ACF plot shows there is some 1st order autocorrelation; and the cumulative ACF plot shows the autocorrelations from the 1st to 12th order are mainly positive, while the autocorrelations from the 13th to 40th order are mainly negative. Our interpretation is that, at shorter horizons of less than 12 months, factor returns tend to be weakly positively autocorrelated, while at horizons of more than 12 months, stock returns tend to be weakly negatively autocorrelated.

Question 3

```r
Box.test(rmw, lag = 6, type = 'Ljung-Box')
```

```
##
##  Box-Ljung test
##
## data:  rmw
## X-squared = 24.488, df = 6, p-value = 0.0004246
```

The null hypothesis is $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$, and the alternative hypothesis is at least one of the β is not equal to 0. The p-value of this Ljung-Box test equals 0.0004246, which is less than 0.01, means we can reject the null hypothesis that the fist 6 autocorrelations jointly equals zero at 1% significance level.

Question 4

```r
ar1 <- lm(rmw ~ lag(rmw, 1))
summary(ar1)
```

```
##
## Call:
## lm(formula = rmw ~ lag(rmw, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.5408  -1.0991   0.0134   0.9963  14.5078
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.21522    0.08254   2.608  0.00932 **
## lag(rmw, 1)  0.16601    0.03800   4.368 1.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.13 on 674 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02753,    Adjusted R-squared:  0.02609
## F-statistic: 19.08 on 1 and 674 DF,  p-value: 1.449e-05
```

Because the 1st order autocorrelation is significant than other orders based on the ACF plot, so we choose the ar(1) model and run the regression. The fitted model we get is $rmw_t = 0.21522 + 0.16601 * rmw_{t-1}$; as a result, the prediction model should be $rmw_{t+1} = 0.21522 + 0.16601 * rmw_t$.

From the fitted model, we learn that stocks with higher RMW factor in last period should have higher RMW factor in this period as well. Although RMW factors of the last 12 periods cumulatively are positively correlated with RMW factor of this period due to short-term momemtum in stock returns, and RMW factors of the last 13-40 periods cumulatively are negatively correlated with RMW factor of this period due to lont-term reversal in stock returns, but the factor individually is not statistically significant enough, except for the neariest period.

Question 5

```r
white_se <- sqrt(diag(vcovHC(ar1)))
white_se
```

```
## (Intercept) lag(rmw, 1)
##  0.09736613  0.12163561
```

```r
ols_se <- sqrt(diag(vcov(ar1)))
ols_se
```

```
## (Intercept) lag(rmw, 1)
##  0.08253926  0.03800321
```

Problem 2

Question 1

```r
par(mar=c(4,4,0.2,0.1))

t = 600
N = 10000
beta <- double(N)
se <- double(N)
for (i in 1:N){
  r1 = 0.005 + 0.04 * rnorm(600)
  r2 = 0.005 + 0.04 * rnorm(600)
  fit <- lm(r1 ~ r2)
  beta[i] <- fit$coefficients[2]
  se[i] <- summary(fit)$sigma
}

mean(beta)
```

```
## [1] 0.0001901631
```

```r
sd(beta)
```

```
## [1] 0.04101806
```

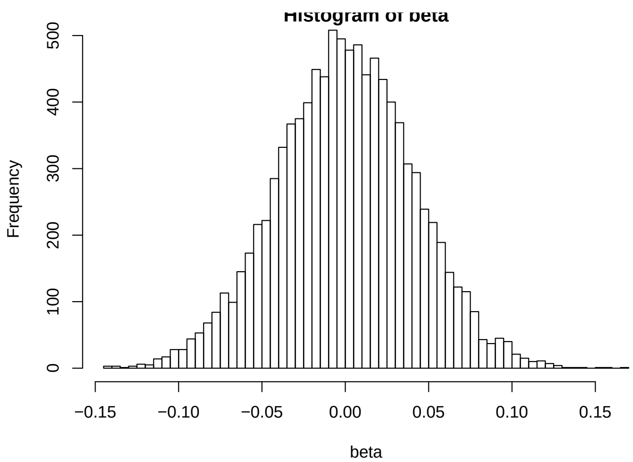
```r
hist(beta, breaks=50)
mean(beta)/(sd(beta)/sqrt(N))
```

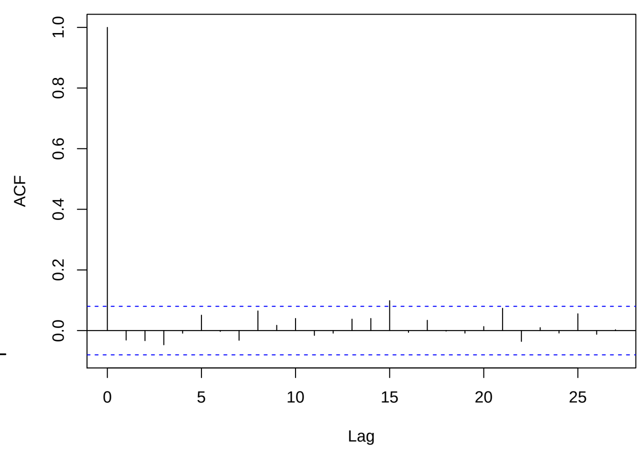
```
## [1] 0.4636081
```

```r
mean(se)
```

```
## [1] 0.03996674
```

```r
acf(fit$residuals)
```





Distribution of beta is cluster around 0, and the t-stat for beta is small, which means the result corresponds to the null hypothesis that $\beta = 0$. Besides, the mean of regress standard errors is very small, and the ACF plot does not indicate obvious autocorrelation, which means they look ok.

Question 2

```r
par(mar=c(4,4,0.2,0.1))

p1 = rep(0,600)
p2 = rep(0,600)
beta2 <- double(N)
se2 <- double(N)
for (j in 1:N){
  r1 <- 0.005 + 0.04*rnorm(600)
  r2 <- 0.005 + 0.04*rnorm(600)
  p1[1] <- r1[1]
  p2[1] <- r2[1]
  for (i in 2:t){
    p1[i] <- p1[i-1] + r1[i]
    p2[i] <- p2[i-1] + r2[i]
  }
  fit2 <- lm(p1 ~ p2)
  beta2[j] <- fit2$coefficients[2]
  se2[j] <- summary(fit2)$sigma
}
```

```r
mean(beta2)
```

```
## [1] 0.9746834
```

```r
sd(beta2)
```

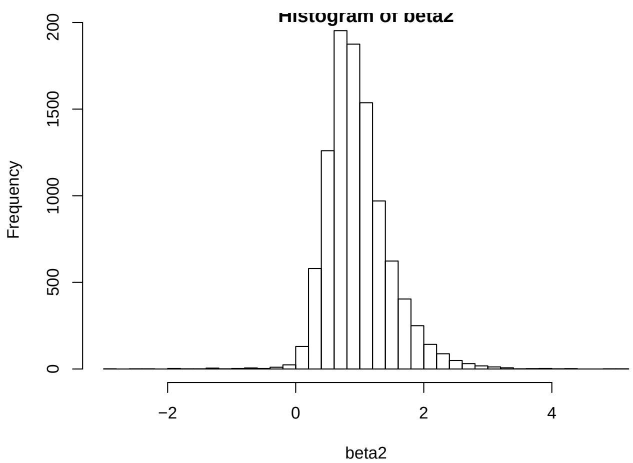
```
## [1] 0.4984527
```

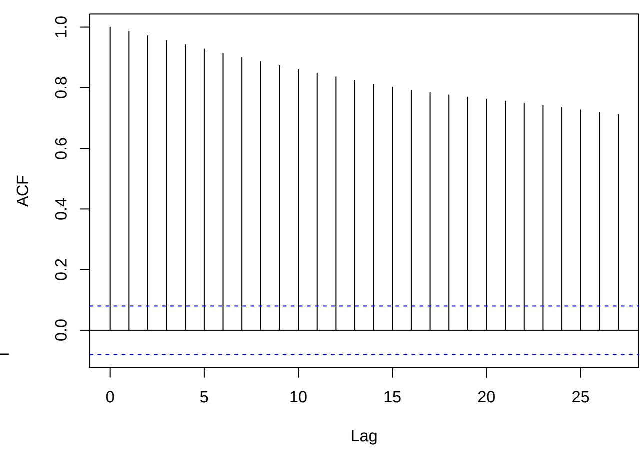
```r
hist(beta2, breaks=50)

mean(se2)
```

```
## [1] 0.3687204
```

```r
acf(fit2$residuals)
```





Distribution of this new beta is cluster around 1, which means the result does not correspond to the null hypothesis that $\beta = 0$. Besides, the mean of regress standard errors is much larger, and the ACF plot indicates that there are high autocorrelation among the residuals, which does not look ok.

The difference between the previous beta and this beta is mainly because, the previous betas are regressed from random numbers that measure returns. Since returns have constant conditional mean and variances and are stationaryt, so the previous betas cluster around 0. While betas in this question are regressed from level of prices, which more likely follow a random walk and are non-stationary, so betas would cluster around 1.