FINM 33150: Homework 1

Michael Beven - 455613

April 2, 2016

3.

First we load modules:

```
# load key packages
library(MASS)
library(Quandl)
library(ggplot2)
library(stringr)
library(R.cache)
library(plotly)
```

Then we add memoization and load the data:

```
# add memoization
reload.data <- TRUE
QLoad <- R.cache::addMemoization(Quandl::Quandl)

# load data
if (reload.data == TRUE) {
  Quandl.api_key('v21snmSix9KyXBWc1RkF')
  cat('Data Reload/n')
  msft.raw <- QLoad('YAHOO/MSFT')
  bac.raw <- QLoad('WIKI/BAC')
}
```

Next we set the data to a 'birth month' and then clean the data:

```
# take desired dates
msft <- subset(msft.raw, msft.raw$Date >= as.Date('1990-10-15') &
```

```
                msft.raw$Date <= as.Date('1990-11-09'))
bac <- subset(bac.raw, bac.raw$Date >= as.Date('1990-10-15') &
                bac.raw$Date <= as.Date('1990-11-09'))

# add index code to data, so we don't confuse them
names(msft) <- paste('YAHOO.MSFT -',names(msft))
names(bac) <- paste('WIKI.BAC -',names(bac))
# fix renamed date column
names(msft)[1] <- 'Date'
names(bac)[1] <- 'Date'
# merge data
df <- merge(msft,bac)

# function to clean Quandl names
clean.quandl.name <- function(x) {
  cleaned <- x
  if (x=='Date') {

  } else {
    tryCatch({
      parts <- stringr::str_split(x, " - ",n=2)
      first.parts <- stringr::str_split(parts[[1]][[1]], "\\.",n=2)
      cleaned <- paste(first.parts[[1]][[2]], parts[[1]][[2]], sep=".")
      cleaned <- str_replace_all(cleaned," ","")
    },
    error = function(e) {cat(paste0("Err on",x,"\n"))}
    )
  }
  stringr::str_trim(cleaned)
}

# use clean.quandl.name
fixed.names <- lapply(names(df),clean.quandl.name)
cat(paste(fixed.names,sep="\n"))
df.renamed <- df
names(df.renamed) <- fixed.names
```

We only need a subset of the data, which is the close prices:

```
# only need close data
df.renamed.close <- subset(df.renamed, select=c('MSFT.Close','BAC.Close'))
```

Now, we can run a linear regression on the close prices:

```
# linear regression on close prices of MSFT vs. BAC
regress <- lm(df.renamed.close$MSFT.Close ~ df.renamed.close$BAC.Close)
print(summary(regress))
```

```
Call:
lm(formula = df.renamed.close$MSFT.Close ~ df.renamed.close$BAC.Close)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8500 -1.0604  0.6566  2.0960  3.7273

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                 76.6265     9.5847   7.995 2.47e-07 ***
df.renamed.close$BAC.Close  -0.7174     0.4907  -1.462    0.161
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.041 on 18 degrees of freedom
Multiple R-squared:  0.1061,Adjusted R-squared:  0.05647
F-statistic: 2.137 on 1 and 18 DF,  p-value: 0.161
```

Finally, we create log returns from the data subset and run a regression on that:

```
# create log returns data
N <- dim(df.renamed.close)[1]
df.logrets <- log(df.renamed.close[2:N,]) - log(df.renamed.close[1:N-1,])
regress.logrets <- lm(df.logrets$MSFT.Close ~ df.logrets$BAC.Close)
print(summary(regress.logrets))

Call:
lm(formula = df.logrets$MSFT.Close ~ df.logrets$BAC.Close)

Residuals:
      Min        1Q    Median        3Q       Max
-0.041179 -0.015816 -0.004455  0.013457  0.069918

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.007217   0.006040   1.195    0.249
df.logrets$BAC.Close -0.035116   0.172365  -0.204    0.841

Residual standard error: 0.02631 on 17 degrees of freedom
Multiple R-squared:  0.002436,Adjusted R-squared:  -0.05624
F-statistic: 0.04151 on 1 and 17 DF,  p-value: 0.841
```