Assigment 2
Mgmt 237E: Empirical Methods
Ian Laker, PrasanthKumar, Nitish Ramkumar

Problem 1

Lets load the data into R

```
options(warn = -1)
suppressMessages(library(lubridate))
suppressMessages(library(xts))
options(warn = 0)

# Retrieve Data

portfolio.data <- read.csv("49_Industry_Portfolios.csv",header=TRUE,sep = ",",stringsAsFactors = FALSE,s
portfolio.data$X <- as.Date(paste(floor(as.numeric(portfolio.data$X)/100),as.numeric(portfolio.data$X) %
portfolio.data <- xts(portfolio.data[,-1],order.by = portfolio.data$X)

datafactors <- read.csv("F-F_Research_Data_Factors.csv",header=TRUE,sep = ",",stringsAsFactors = FALSE,s
datafactors$X <- as.Date(paste(floor(as.numeric(datafactors$X)/100),as.numeric(datafactors$X) %% 100,"0
datafactors <- xts(datafactors[,-1],order.by = datafactors$X)

#Constraints
#Date
portfolio.data <- portfolio.data[index(portfolio.data)>="1960-01-01" & index(portfolio.data)<="2015-12-3
datafactors <- datafactors[index(datafactors)>="1960-01-01" & index(datafactors)<="2015-12-31",]

#Remove columns with -99.99 value
contains99 <- logical(dim(portfolio.data)[2])
for(col in 1:dim(portfolio.data)[2]){
    contains99[col] <- sum(portfolio.data[,col] %in% -99.99) > 0
}
portfolio.data <- portfolio.data[,!contains99]


#Excess Returns
excessReturns <- function(industryReturns){
  industryReturns - datafactors$RF
}

excess.returns <- matrix(nrow=dim(portfolio.data)[1],ncol=dim(portfolio.data)[2])
for(col in 1:dim(portfolio.data)[2]){
  excess.returns[,col] <- portfolio.data[,col] - datafactors$RF
}
colnames(excess.returns) <- colnames(portfolio.data)
```

For each industry, regress the excess return on market excess return

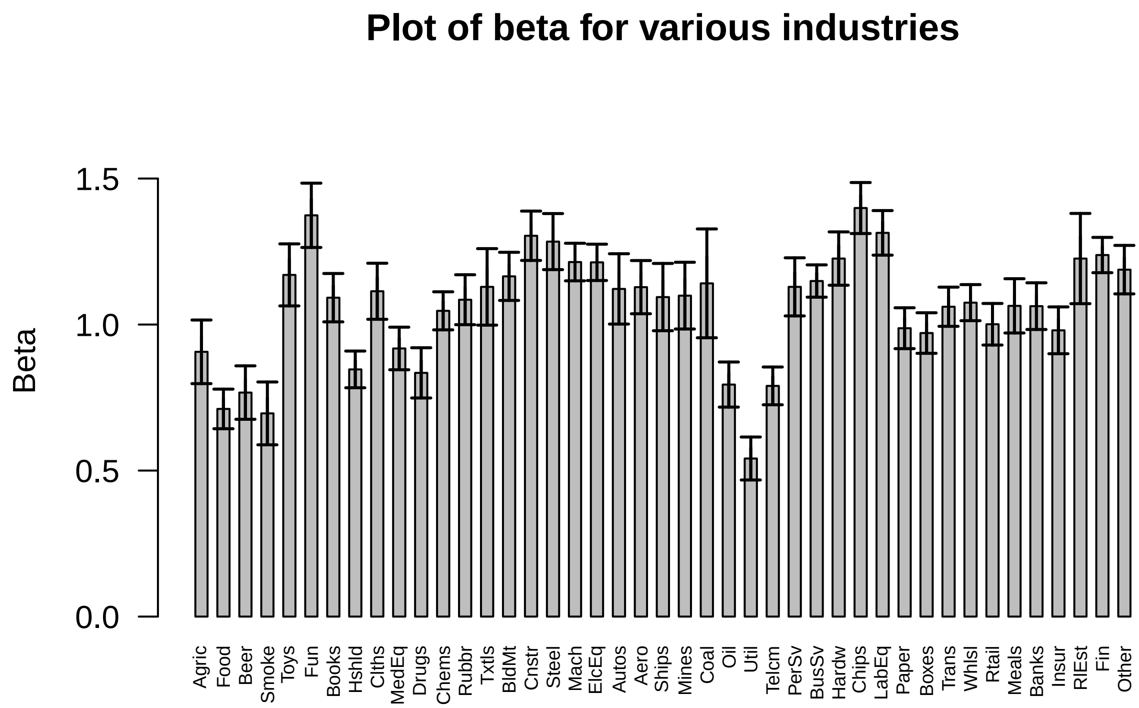
```
options(warn = -1)
suppressWarnings(library(DataAnalytics))
options(warn = 0)
regressionData <- function(excess.industryReturns){
```

```
  reg <- lm(excess.industryReturns~datafactors$Mkt.RF[index(excess.industryReturns)])
  dummy <- capture.output(summ <- lmSumm(reg,HAC = TRUE))
  result <- c(summ$coef.table[2,c(1,2)],summ$coef.table[1,c(1,2)],summary(summ$lmfit)$r.squared)
  t(result)
}

excess.returns <- xts(excess.returns,order.by = as.Date(index(portfolio.data)))
regressionInfo <- apply(excess.returns, 2, regressionData)
```

a

Lets plot the betas of various industry

```
betaTable <- regressionInfo[c(1,2),]
betaBars <- barplot(height = betaTable[1,],
                  names.arg = colnames(betaTable),
                  beside = true, las = 2,
                  ylim = c(0, 1.75),
                  cex.names = 0.75, xaxt = "n",
                  main = "Plot of beta for various industries",
                  ylab = "Beta",
                  border = "black", axes = TRUE,width=35,space = 0.8)

text(x = betaBars, y = -0.1, srt = 90,
     adj = 1, labels = colnames(betaTable), xpd = TRUE,cex=0.6)

segments(betaBars, betaTable[1,]- betaTable[2,] * 2, betaBars,
         betaTable[1,] + betaTable[2,], lwd = 1.5)

arrows(betaBars, betaTable[1,]- betaTable[2,] * 2, betaBars,
         betaTable[1,] + betaTable[2,]*2, lwd = 1.5, angle = 90,
       code = 3, length = 0.05)
```



b

The range of the betas can be found by using the *range* function

```
range(betaTable[1,])
```

```
## [1] 0.5413 1.3990
```

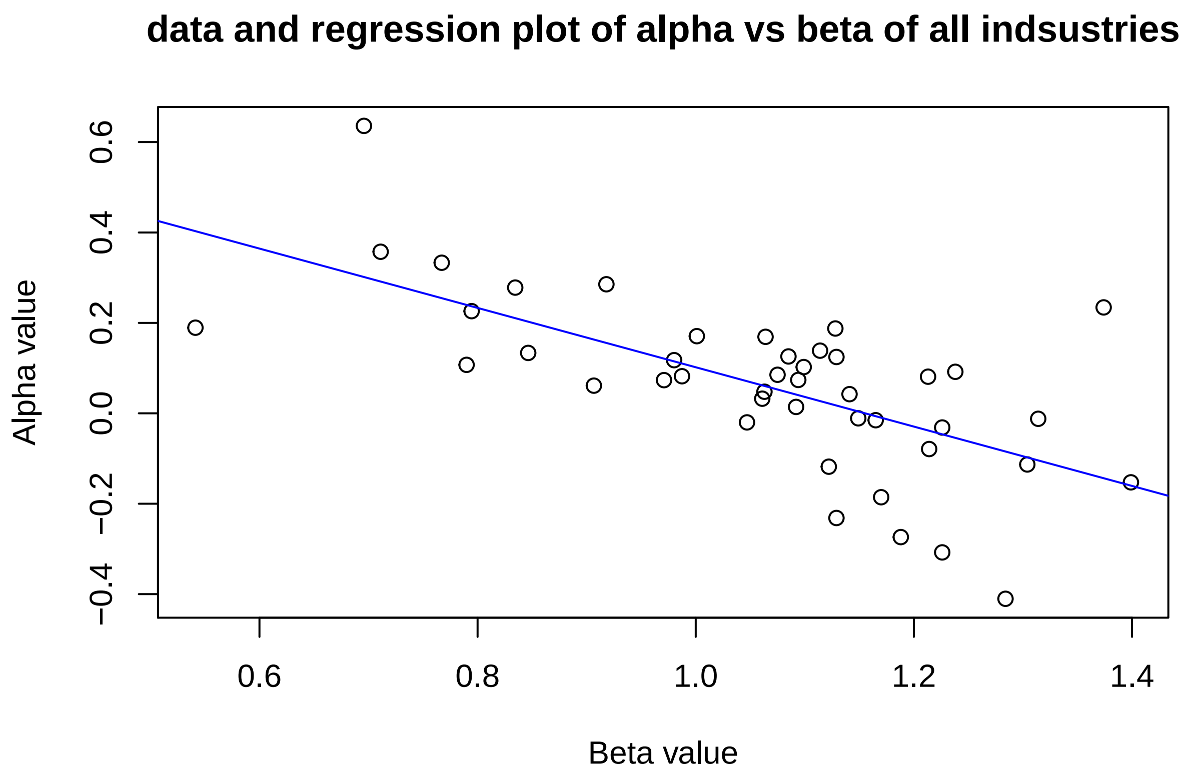
The max, min and mean of the R^2 across industries is as below

```
resultsR2 <- c()
resultsR2[1] <- min(regressionInfo[5,])
resultsR2[2] <- max(regressionInfo[5,])
resultsR2[3] <- mean(regressionInfo[5,])
names(resultsR2) <- c("Min","Max","Mean")
resultsR2
```

```
##       Min       Max      Mean
## 0.2505057 0.8055172 0.5776919
```

c

```r
plot(regressionInfo[1,],regressionInfo[3,],xlab="Beta value",ylab="Alpha value",main = "data and regress
alphalm <- lm(regressionInfo[3,]~regressionInfo[1,])
abline(alphalm$coefficients[1],alphalm$coefficients[2],col = "blue")
```



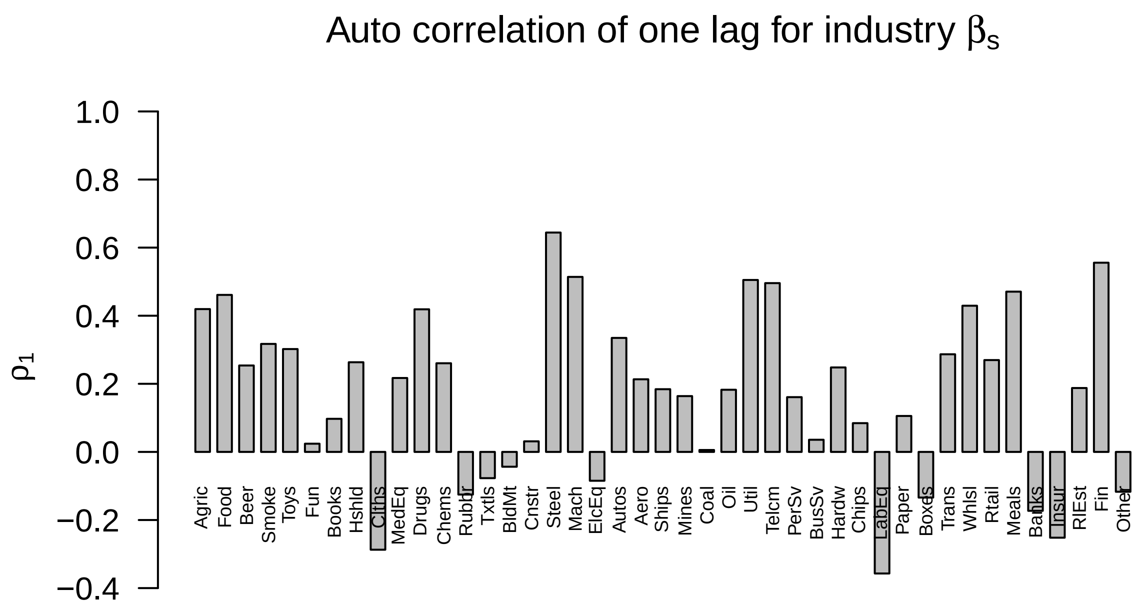
We can see that for lower beta, we mostly get higher alpha.

In CAPM, we expect low beta to have low risk, which means low returns. But in this case, there is high alpha. So there is a chance that the total returns is higher for low beta stocks. This goes against CAPM.

Problem 2

```r
acfBeta <- function(indsustryBeta){
  acf_Val <- acf(indsustryBeta,plot = FALSE)
  acf_Val$acf[2]
}


fiveYearBeta <- matrix(nrow=((2010-1960)/5)+1,ncol=dim(excess.returns)[2])
fiveYearBeta.Error <- matrix(nrow=((2010-1960)/5)+1,ncol=dim(excess.returns)[2])
fiveyears.Endpoints <- seq(1960,2010,by=5)
for(fiveYearStart in fiveyears.Endpoints){
  excess.5year.returns <- excess.returns[paste0(fiveYearStart,"/",fiveYearStart+4)]
  results <- sapply(excess.5year.returns,regressionData)
  fiveYearBeta[((fiveYearStart-1960)/5)+1,] <- results[1,]
```

```r
  fiveYearBeta.Error[((fiveYearStart-1960)/5)+1,] <- results[2,]

}

fiveYearBeta <- xts(fiveYearBeta,order.by= as.Date(paste(fiveyears.Endpoints,"01","01",sep = "-")))
fiveYearBeta.Error <- xts(fiveYearBeta.Error,order.by= as.Date(paste(fiveyears.Endpoints,"01","01",sep =
colnames(fiveYearBeta) <- colnames(excess.returns)
colnames(fiveYearBeta.Error) <- colnames(excess.returns)

industry.acf <- apply(fiveYearBeta,2,acfBeta)
beta.5yearsCor <- barplot(height = industry.acf,
                  names.arg = colnames(industry.acf),
                  beside = true, las = 2,
                  ylim = c(-0.5, 1),
                  cex.names = 0.75, xaxt = "n",
                  main = expression(paste("Auto correlation of one lag for industry ",beta[s])),
                  ylab = expression(rho[1]),
                  border = "black", axes = TRUE,width=35,space = 0.5)
text(x = beta.5yearsCor, y = -0.1, srt = 90,
     adj = 1, labels =colnames(excess.returns), xpd = TRUE,cex=0.6)
```



Some of the industries are almost 0 correlation (like Coal, Cnstr, BusSv). For these industries, the Beta will be totally unstable.

The difference across the 5-year periods can explained either due to changes in the true Beta in these industries, or due to sampling error.

For example, if we take Coal, the rolling betas and its standard errors can be found below.

```
results <- cbind(fiveYearBeta[,"Coal"],fiveYearBeta.Error[,"Coal"])
results <- c("Coal beta","Coal beta error")
```

As you can see, in some time periods, the error is very high. For example, in the 5 year period beginning in 2000, the beta is 0.6773, and the error is 0.421. This is quite significant and can lead to improper correlations.