Assignment 02, Question 1&2

Jeff Nguyen

February 14, 2021

University of Southern California
Marshall School of Business
FBE 543 Forecasting and Risk Analysis

Student Name: Ngoc Son (Jeff) Nguyen

Question 1

a. Estimating a one-factor CAPM for a single-stock portfolio containing only AAPL.

To estimate the β coefficients of the single stock (AAPL) CAPM portfolio, we follow these steps:

1. Download data from Yahoo Finance. The following tickers are downloaded: AAPL, GSPC (representing the S&P 500) and TNX (representing the US Treasury Bills.)
2. Calculate the following statistics of the portfolio using adjusted monthly closing price (converted from daily): monthly returns (AAPL, GSPC and TNX), mean returns.
3. The expected risk premium of the portfolio based on the CAPM model is given as:

$$
\begin{aligned}
E[R_a - R_f] =& \beta * (R_m - R_f) \\
& \text{or} \\
R_a =& R_f + \beta * (R_m - R_f) \\
R_a - R_f =& \alpha_{Jensen} + \beta * (R_m - R_f) \\
& \text{or} \\
Y =& \alpha_{Jensen} + \beta * X + epsilon \\
\text{with} & \\
Y =& R_a - R_f \\
X =& R_m - R_f \\
\beta =& \text{Market risk or systematic risk} \\
\epsilon =& \text{stochastic error term}
\end{aligned} \tag{1}
$$

```
library(quantmod)
```

```
## Loading required package: xts
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: TTR

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

## Version 0.4-0 included new data defaults. See ?getSymbols.
```

```
# Set start date and end date of data
start_date <- "2018-01-01"
end_date <- "2020-12-31"

# Get data
getSymbols("AAPL", src = "yahoo", from = start_date, to = end_date)
```

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.

## [1] "AAPL"
```

```
getSymbols("^GSPC", src = "yahoo", , from = start_date, to = end_date) # S&P 500
```

```
## [1] "^GSPC"
```

```
getSymbols("^TNX", src = "yahoo", from = start_date, to = end_date) # TNX (10-year T-bill)
```

```
## Warning: ^TNX contains missing values. Some functions will not work if objects
## contain missing values in the middle of the series. Consider using na.omit(),
## na.approx(), na.fill(), etc to remove or replace them.

## [1] "^TNX"
```

```r
# Adjusted Prices
adjAAPL <- AAPL$AAPL.Adjusted

# Get adjusted returns data
rAAPL <- diff(log(to.monthly(AAPL)$AAPL.Adjusted))
rGSPC <- diff(log(to.monthly(GSPC)$GSPC.Adjusted))
rTNX <- (to.monthly(TNX)$TNX.Adjusted) / 1200 # Using monthly rate
```

```
## Warning in to.period(x, "months", indexAt = indexAt, name = name, ...): missing
## values removed from data
```

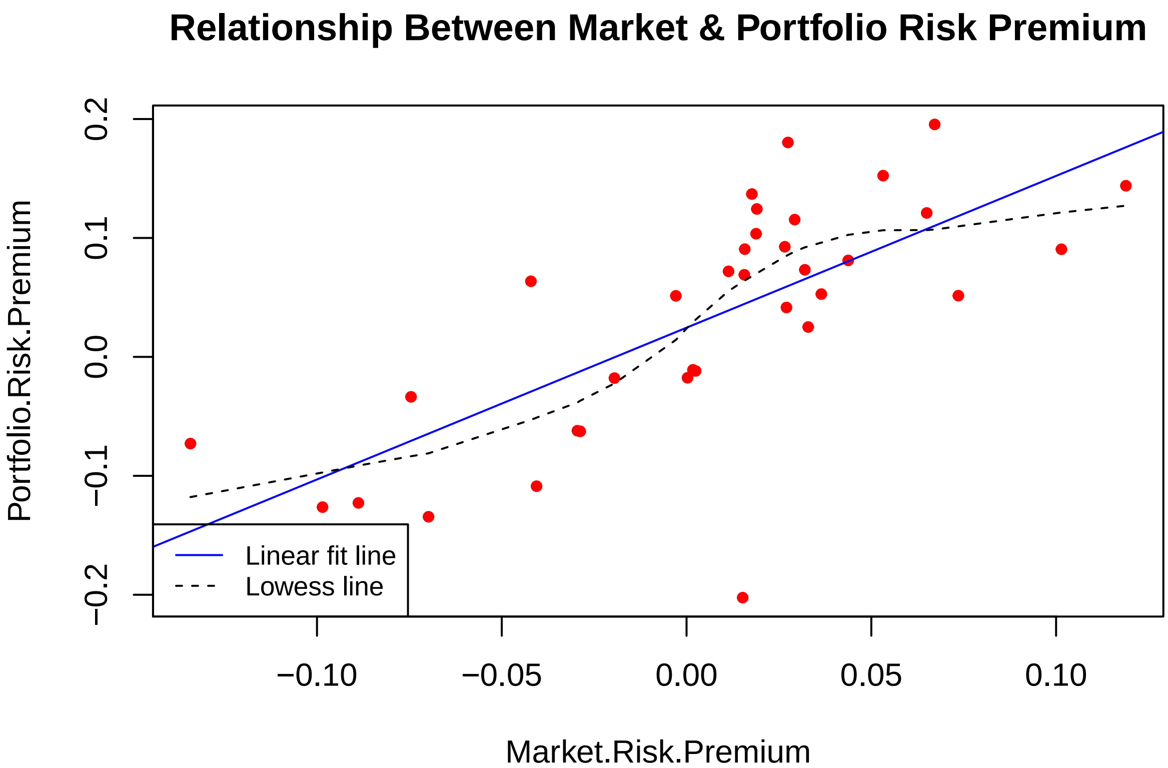
```r
# Calculate Portfolio Risk Premium
Y <- na.omit(rAAPL - rTNX)
names(Y)[1] <- "Portfolio Risk Premium"  # Rename column
Y_bar <- mean(Y)
Y_bar
```

```
## [1] 0.03269352
```

```r
# Calculate Market Risk Premium
X <- na.omit(rGSPC - rTNX)
X_bar <- mean(X)
names(X)[1] <- "Market Risk Premium"  # Rename column

# Create a data frame of X, Y
data1 <- data.frame(X, Y)
Y_data <- data1$Portfolio.Risk.Premium
X_data <- data1$Market.Risk.Premium
```

We can observe the relationship between the market (S&P 500) risk premium and the portfolio (AAPL) risk premium as follow:

```r
plot(data1, col='red', main="Relationship Between Market & Portfolio Risk Premium"
     , pch=20, cex=1)

# Add fit lines
abline(lm(Y_data~X_data), col="blue")  # Regression line Y ~ X
lines(lowess(X_data,Y_data), col="black", lty=2)  # Lowess line (X,Y)

legend("bottomleft",c("Linear fit line", "Lowess line"), cex=.8, col=c("blue", "black")
       , lty=1:2)
```



Relationship of AAPL with market

Upon inspection, we observe a strong positive correlation between the AAPL portfolio and the market. Therefore, we can conclude that AAPL follow the market relatively closely, but not necessarily linearly.

Hypothesis testing

Next, we attempts to fit an equation of a line: $Y = \alpha_{Jensen} + \beta * X + \epsilon$

Hypothesis for regression:

$$\begin{aligned} H_0 : \alpha &= 0 \\ H_a : \alpha &\neq 0 \\ &\text{and} \\ H_0 : \beta &= 0 \\ H_a : \beta &\neq 0 \end{aligned} \tag{2}$$

```r
fit1 <- lm(Y_data~X_data, data=data1)
summary(fit1)
```

```
##
## Call:
## lm(formula = Y_data ~ X_data, data = data1)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.246364 -0.040481  0.000741  0.049734  0.120726
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.02455    0.01195   2.055   0.0479 *
## X_data       1.27630    0.22000   5.801 1.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0702 on 33 degrees of freedom
## Multiple R-squared:  0.5049, Adjusted R-squared:  0.4899
## F-statistic: 33.66 on 1 and 33 DF,  p-value: 1.731e-06
```

The estimated equation is $Y = .02455 - 1.27630 * X$, where the p_{value} for the intercept $.0479 < .05$.
Therefore, we reject the null hypothesis at 95% confidence level that the intercept α_{Jensen} statistically is no different from 0. Thus, we reject the null hypothesis $H_0 : \alpha = 0$ and accept the alternative hypothesis $H_a : \alpha \neq 0$.
The coefficient $\beta = 1.27630$ represents the increase in portfolio risk premium relative to increase in the market portfolio risk premium. The p_{value} for β is $1.73e - 06 < .05$, implying that the coefficient β statistically is significant at 95% or more, and we reject the null hypothesis $H_0 : \beta = 0$ and accept the alternative hypothesis $H_a : \beta \neq 0$.

Goodness of Fit:
Through inspection of the linear regression result, we observe the $R^2 = .5049$ value to not be close to 1 at all. $R^2 = .5049$ implies that 50.49% of the variations in the portfolio risk premium is explained by the market risk premium.

Standard Error of Regression:
We can see that the Standard Error of Regression is $S.E. = .0702$.
From this, we can calculate the forecasting efficiency statistic to be:

$$\begin{aligned} \frac{S.E.}{\overline{Y}} &= \frac{.0702}{0.0326935} \\ &= 214.72\% > 10\% \end{aligned} \tag{3}$$

This statistic implies that the linear model is not a good forecasting model.

Thus, upon exploring the goodness of fit and standard error of regression, we confirm our initial observation that the portfolio risk premium and the market portfolio risk premium has a weak linear relationship.

b. Estimating a one-factor CAPM for a single-stock portfolio containing only AAPL including intercept and slope dummies for the shut-down of the economy in March 16, 2020.

We run regression for the model to see if the March 16, 2020 pandemic shut-down has any effect on Jensen alpha:

$$
\begin{aligned}
&Y = \alpha_{Jensen} + \beta * X + \beta_1 * D + \epsilon \\
&\text{where:} \\
&\quad Y = \text{Portfolio Risk Premium } R_a - R_f \\
&\quad \beta = \text{Estimated } \beta \text{ of the CAPM} \\
&\quad X = \text{Market portfolio risk premium } R_m - R_f \\
&\quad \beta_1 = \text{Coefficient representing the relationship of the portfolio risk premium with the pandemic factor} \\
&\quad D = \text{Dummy variable representing the pandemic factor} \\
&\quad \epsilon = \text{Standard Error of Regression}
\end{aligned} \tag{4}
$$

Hypothesis for regression:

$$
\begin{gathered}
H_0 : \alpha = 0 \\
H_a : \alpha \neq 0 \\
\text{and} \\
H_0 : \beta_1 = 0 \\
H_a : \beta_1 \neq 0
\end{gathered} \tag{5}
$$

```r
# Generating the dummy variable
D <- rep(0, 35)
D[26] <-1
D
```

```
##  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
```

```r
data2 <- data.frame(D, Y)
names(data2)[1] <- "Pandemic Variable"
```

We examine our dataset with the pandemic variable:

```r
data2
```

```
##          Pandemic Variable Portfolio.Risk.Premium
## Feb 2018                 0             0.06357069
## Mar 2018                 0            -0.06208779
## Apr 2018                 0            -0.01758038
## May 2018                 0             0.12439008
## Jun 2018                 0            -0.01183693
## Jul 2018                 0             0.02512882
## Aug 2018                 0             0.18029585
## Sep 2018                 0            -0.01088427
## Oct 2018                 0            -0.03358408
## Nov 2018                 0            -0.20242328
## Dec 2018                 0            -0.12632697
## Jan 2019                 0             0.05149095
## Feb 2019                 0             0.04154389
## Mar 2019                 0             0.09059090
## Apr 2019                 0             0.05280998
```

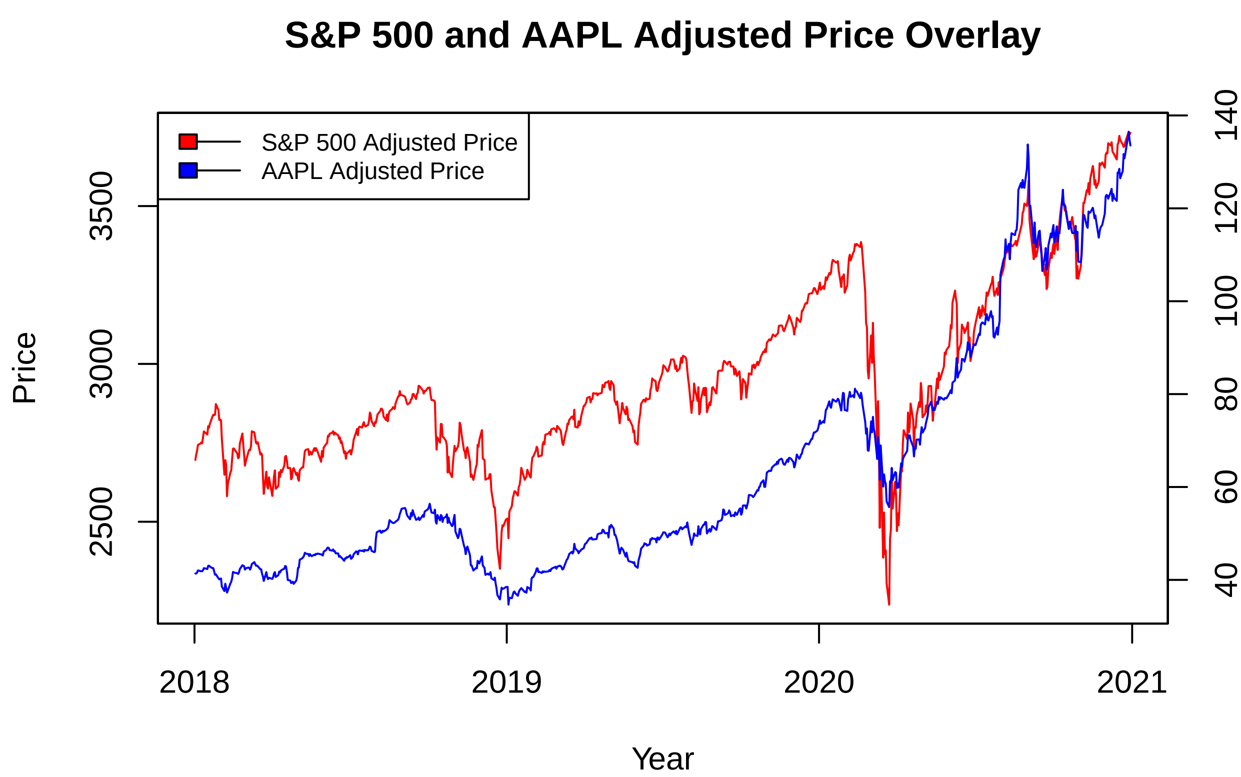
```
## May 2019                  0            -0.13441711
## Jun 2019                  0             0.12101019
## Jul 2019                  0             0.07193279
## Aug 2019                  0            -0.01785310
## Sep 2019                  0             0.06902680
## Oct 2019                  0             0.10356726
## Nov 2019                  0             0.07322116
## Dec 2019                  0             0.09260502
## Jan 2020                  0             0.05133505
## Feb 2020                  0            -0.12276945
## Mar 2020                  1            -0.07289590
## Apr 2020                  0             0.14390556
## May 2020                  0             0.08112683
## Jun 2020                  0             0.13694231
## Jul 2020                  0             0.15238742
## Aug 2020                  0             0.19545737
## Sep 2020                  0            -0.10873568
## Oct 2020                  0            -0.06260496
## Nov 2020                  0             0.09050138
## Dec 2020                  0             0.11543281
```

Initial Observation

We then do a plot of AAPL portfolio stock price:

```r
# Initialize xts objects contain adjusted price for S&P 500 and AAPL and merge
gspc_xts <- as.xts(GSPC[,"GSPC.Adjusted"])
aapl_xts <- as.xts(AAPL[,"AAPL.Adjusted"])
price_compare <- merge.xts(gspc_xts, aapl_xts)

# Graph monthly AAPL and monthly S&P500 on one coordinate system
# Plot S&P 500
plot(as.zoo(price_compare[, "GSPC.Adjusted"]), screens = 1,
     main = "S&P 500 and AAPL Adjusted Price Overlay",
     xlab = "Year", ylab = "Price", col = "Red")

# Keep working on the same plot
par(new = TRUE)

# Plot AAPL and suppress axis value
plot(as.zoo(price_compare[, "AAPL.Adjusted"]),
     screens = 1,
     xaxt = "n", yaxt = "n",
     xlab = "", ylab = "",
     col = "Blue")

# Add right-handed axis to display AAPL price
axis(4)

# Add legend
legend("topleft",
       c("S&P 500 Adjusted Price", "AAPL Adjusted Price"),
       lty = 1:1,
```

```r
       cex = 0.75,
       fill = c("red", "blue"))
```



Upon inspection, we notice there is violent decline in the price of the market portfolio, represented by the S&P500 as well as our single-stock portfolio of AAPL around March 2020.

Regression

Let's assume linearity, we run regression on $Y = \alpha_{Jensen} + \beta * X + \beta_1 * D$ as hypothesized above.

```r
fit2 <- lm(Y~X+D)
summary(fit2)
```

```
##
## Call:
## lm(formula = Y ~ X + D)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24480 -0.04023 -0.00145  0.04889  0.12068
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.02100    0.01223   1.717   0.0957 .
```

```
## X             1.40762    0.24431   5.762 2.17e-06 ***
## D             0.09508    0.07910   1.202   0.2382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06973 on 32 degrees of freedom
## Multiple R-squared:  0.5263, Adjusted R-squared:  0.4967
## F-statistic: 17.78 on 2 and 32 DF,  p-value: 6.424e-06
```

The estimated equation is $Y = .02100 + 1.40762 * X - .09508 * D$, where the p_{value} for the intercept $.0957 > .05$. Therefore, we failed reject the null hypothesis at 95% confidence level that the intercept α_{Jensen} statistically is no different from zero. Thus, we failed to reject the null hypothesis $H_0 : \alpha = 0$.

Pandemic Factor

In addition, the coefficient $\beta_1 = -.16285$ represents the increase in portfolio risk premium relative to increase in the pandemic variable. The p_{value} for β_1 is $2.17e - 06 < .05$, implying that the coefficient β_1 statistically is significant at 95% or more, and we reject the null hypothesis $H_0 : \beta_1 = 0$ and accept the alternative hypothesis $H_a : \beta_1 \neq 0$.

Goodness of Fit:
Through inspection, we observe the $R^2 = .5263$ value to not be close to 1 at all. $R^2 = .5263$ implies that 52.63% of the variations in the portfolio risk premium is explained by the pandemic variable.

Standard Error of Regression:
We can see that the Standard Error of Regression is $S.E. = .06973$.
From this, we can calculate the forecasting efficiency statistic to be:

$$\begin{aligned} \frac{S.E.}{\overline{Y}} &= \frac{.06973}{0.0326935} \\ &= 213.28\% > 10\% \end{aligned} \tag{6}$$

This statistic implies that this is not a good forecasting model.

Remarks

Based on the above result of the regression, we confirm our initial observation that there is evidence that the pandemic shut-down did have an effect on the AAPL portfolio's portfolio risk premium and consequently, its valuation as well.

c. Estimating a three-factor CAPM for a single-stock portfolio containing only AAPL.

The FF CAPM 3-factor model

The Fama-French three-factor is expressed as:

$$
\begin{aligned}
R_a - R_f = &\quad \alpha + \beta_1(R_m - R_f) + \beta_2\text{SMB} + \beta_3\text{HML} + \epsilon \\
\text{where} & \\
R_a - R = f = &\quad \text{Portfolio Risk Premium} \\
R_m - R = f = &\quad \text{Market Factor} \\
R_f = &\quad \text{Risk-free rate} \\
\beta_1 = &\quad \text{analogous to the original CAPM beta} \\
\text{SMB} = &\quad \text{Size Factor} \\
\text{HML} = &\quad \text{Value Factor} \\
\epsilon = &\quad \text{stochastic error term}
\end{aligned} \tag{7}
$$

Importing data

Importing Fama-French CAPM 3-factor data to our data frame and extract the data

```
# Importing the FF 3-factor CAPM
ff_data <- read.table("ffdata.csv", header=TRUE, sep=",")
ff_data
```

```
##      Date Mkt.RF   SMB    HML   RF        AAPL
## 1  201802  -3.65  0.30  -1.01 0.11  0.06357069
## 2  201803  -2.35  4.05  -0.12 0.12 -0.06208779
## 3  201804   0.29  1.13   0.50 0.14 -0.01758038
## 4  201805   2.65  5.28  -3.15 0.14  0.12439008
## 5  201806   0.48  1.19  -2.31 0.14 -0.01183693
## 6  201807   3.19 -2.21   0.57 0.16  0.02512882
## 7  201808   3.44  1.15  -3.92 0.16  0.18029585
## 8  201809   0.06 -2.28  -1.63 0.15 -0.01088427
## 9  201810  -7.68 -4.78   3.46 0.19 -0.03358408
## 10 201811   1.69 -0.71   0.34 0.18 -0.20242328
## 11 201812  -9.55 -2.48  -1.89 0.19 -0.12632697
## 12 201901   8.41  2.89  -0.46 0.21  0.05149095
## 13 201902   3.40  2.12  -2.71 0.18  0.04154389
## 14 201903   1.10 -3.05  -4.19 0.19  0.09059090
## 15 201904   3.96 -1.75   2.02 0.21  0.05280998
## 16 201905  -6.94 -1.18  -2.28 0.21 -0.13441711
## 17 201906   6.93  0.22  -0.79 0.18  0.12101019
## 18 201907   1.19 -2.08   0.34 0.19  0.07193279
## 19 201908  -2.58 -2.40  -4.85 0.16 -0.01785310
## 20 201909   1.43 -1.05   6.77 0.18  0.06902680
## 21 201910   2.06  0.24  -1.88 0.15  0.10356726
## 22 201911   3.87  0.91  -2.05 0.12  0.07322116
## 23 201912   2.77  0.67   1.91 0.14  0.09260502
## 24 202001  -0.11 -3.08  -6.27 0.13  0.05133505
## 25 202002  -8.13  1.02  -3.92 0.12 -0.12276945
## 26 202003 -13.38 -5.03 -13.96 0.12 -0.07289590
## 27 202004  13.65  2.75  -1.39 0.00  0.14390556
## 28 202005   5.58  2.49  -5.05 0.01  0.08112683
## 29 202006   2.46  2.71  -2.35 0.01  0.13694231
## 30 202007   5.77 -2.18  -1.39 0.01  0.15238742
## 31 202008   7.63 -0.25  -2.94 0.01  0.19545737
```

```
## 32 202009  -3.63  0.06  -2.51 0.01 -0.10873568
## 33 202010  -2.10  4.44   4.03 0.01 -0.06260496
## 34 202011  12.47  5.48   2.11 0.01  0.09050139
## 35 202012   4.63  4.91  -1.46 0.01  0.11543280
```

```r
# Extract column data
rmrf <- ff_data[,2]
smb <- ff_data[,3]
hml <- ff_data[,4]
rf <- ff_data[,5]
fund <- ff_data[,6]

# Calculate excess return for AAPL portfolio
fundExcess <- fund - rf
```

Regression Analysis

Hypothesis for regression:

$$
\begin{aligned}
H_0 : \alpha &= 0 \\
H_a : \alpha &\neq 0 \\
&\text{and} \\
H_0 : \beta_1 &= \beta_2 = \beta_3 = 0 \\
H_a : \text{At least one of} \quad \beta_1, \beta_2, \beta_3 &\neq 0
\end{aligned} \tag{8}
$$

Next up, we run the linear regression on the original FF CAPM 3-factor using imported data:

```r
ffRegression <- lm(fundExcess ~ rmrf + smb + hml)
summary(ffRegression)
```

```
##
## Call:
## lm(formula = fundExcess ~ rmrf + smb + hml)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.276625 -0.042814 -0.005902  0.043436  0.175208
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.124303   0.019316  -6.435 3.58e-07 ***
## rmrf         0.016149   0.003490   4.627 6.24e-05 ***
## smb          0.007402   0.007182   1.031   0.3107
## hml         -0.010387   0.005193  -2.000   0.0543 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09977 on 31 degrees of freedom
## Multiple R-squared:  0.5252, Adjusted R-squared:  0.4793
## F-statistic: 11.43 on 3 and 31 DF,  p-value: 3.273e-05
```

The estimated equation is $R_a - R_f = -.124303 + .016149(R_m - R_f) + .007402 * SMB - .010387 * HML$, where the p_{value} for the intercept $3.58e - 07 < .05$.
Therefore, we reject the null hypothesis at 95% confidence level that the intercept α_{Jensen} statistically is no different from 0. Thus, we reject the null hypothesis $H_0 : \alpha = 0$ and accept the alternative hypothesis $H_a : \alpha \neq 0$.

Market factor The coefficient $\beta_1 = .016149$ represents the increase in portfolio risk premium relative to increase in the market portfolio risk premium. The p_{value} for β_1 is $6.24e - 05 < .05$, implying that the coefficient β_1 statistically is significant at 95% or more, and we reject the null hypothesis $H_0 : \beta_1 = 0$ and accept the alternative hypothesis $H_a : \beta_1 \neq 0$.

Size factor The coefficient $\beta_2 = ..007402$ represents the increase in portfolio risk premium relative to increase firm size. The p_{value} for β_1 is $.3107 > .05$, implying that the coefficient β_2 statistically is insignificant at 95% or more, and we accept the null hypothesis $H_0 : \beta_2 = 0$ and reject the alternative hypothesis $H_a : \beta_2 \neq 0$.

Value factor The coefficient $\beta_3 = .010387$ represents the increase in portfolio risk premium relative to increase in the market portfolio risk premium. The p_{value} for β_1 is $.0543 > .05$, implying that the coefficient β_3 statistically is insignificant at 95% or more, and we accept the null hypothesis $H_0 : \beta_3 = 0$ and reject the alternative hypothesis $H_a : \beta_3 \neq 0$.

Goodness of Fit:

Through inspection of the linear regression result, we observe the $R^2 = .5252$ value to not be close to 1 at all. $R^2 = .5252$ implies that 52.52% of the variations in the portfolio risk premium is explained by the market risk premium.

Standard Error of Regression:

We can see that the Standard Error of Regression is $S.E. = .09977$.
From this, we can calculate the forecasting efficiency statistic to be:

$$\begin{aligned} \frac{S.E.}{\overline{Y}} &= \frac{.09977}{0.0326935} \\ &= 305.17\% > 10\% \end{aligned} \tag{9}$$

This statistic implies that the linear model is not a good forecasting model.

Remarks

The 3-factor model differs from the single-factor CAPM in that its hypothesis is that the market valuation of a portfolio is based on three different risks factors: market, size and value compared to just market in the single-factor model.

The addition of the two factors add flexibility to the original CAPM model. It also adjusts for the historical tendency of small-cap stocks outperforming big-cap stock. Thus, it does not penalize predominantly big-cap portfolio and consequently, provids a better tool to evaluate the portfolio performance.

d. Estimating a three-factor CAPM for a single-stock portfolio containing only AAPL including intercept and slope dummies for the shut-down of the economy in March 16, 2020.

The FF CAPM 3-factor model with dummy pandemic variable

The model is expressed as follow:

$$
\begin{aligned}
R_a - R_f = & \quad \alpha + \beta_1(R_m - R_f) + \beta_2 \text{SMB} + \beta_3 \text{HML} + \beta_4 * D\epsilon \\
\text{where} & \\
R_a - R = f = & \quad \text{Portfolio Risk Premium} \\
R_m - R = f = & \quad \text{Market Factor} \\
R_f = & \quad \text{Risk-free rate} \\
\beta_1 = & \quad \text{analogous to the original CAPM beta} \\
\text{SMB} = & \quad \text{Size Factor} \\
\text{HML} = & \quad \text{Value Factor} \\
\beta_4 = & \quad \text{Coefficient representing the relationship of the portfolio risk premium with the pandemic factor} \\
D = & \quad \text{Dummy variable representing the pandemic factor} \\
\epsilon = & \quad \text{stochastic error term}
\end{aligned}
\tag{10}
$$

Regression Analysis

Hypothesis for regression:

$$
\begin{aligned}
H_0 : \alpha &= 0 \\
H_a : \alpha &\neq 0 \\
\text{and} & \\
H_0 : \beta_4 &= 0 \\
H_a : \beta_4 &\neq 0
\end{aligned}
\tag{11}
$$

We examine our dataset with the pandemic variable:

```
# Generating the data frame with dummy variable
ff_data$Pandemic <- D
ff_data
```

```
##      Date Mkt.RF   SMB   HML   RF        AAPL Pandemic
## 1  201802  -3.65  0.30 -1.01 0.11  0.06357069        0
## 2  201803  -2.35  4.05 -0.12 0.12 -0.06208779        0
## 3  201804   0.29  1.13  0.50 0.14 -0.01758038        0
## 4  201805   2.65  5.28 -3.15 0.14  0.12439008        0
## 5  201806   0.48  1.19 -2.31 0.14 -0.01183693        0
## 6  201807   3.19 -2.21  0.57 0.16  0.02512882        0
## 7  201808   3.44  1.15 -3.92 0.16  0.18029585        0
## 8  201809   0.06 -2.28 -1.63 0.15 -0.01088427        0
## 9  201810  -7.68 -4.78  3.46 0.19 -0.03358408        0
## 10 201811   1.69 -0.71  0.34 0.18 -0.20242328        0
## 11 201812  -9.55 -2.48 -1.89 0.19 -0.12632697        0
## 12 201901   8.41  2.89 -0.46 0.21  0.05149095        0
## 13 201902   3.40  2.12 -2.71 0.18  0.04154389        0
## 14 201903   1.10 -3.05 -4.19 0.19  0.09059090        0
```

```
## 15 201904   3.96 -1.75   2.02 0.21  0.05280998        0
## 16 201905  -6.94 -1.18  -2.28 0.21 -0.13441711        0
## 17 201906   6.93  0.22  -0.79 0.18  0.12101019        0
## 18 201907   1.19 -2.08   0.34 0.19  0.07193279        0
## 19 201908  -2.58 -2.40  -4.85 0.16 -0.01785310        0
## 20 201909   1.43 -1.05   6.77 0.18  0.06902680        0
## 21 201910   2.06  0.24  -1.88 0.15  0.10356726        0
## 22 201911   3.87  0.91  -2.05 0.12  0.07322116        0
## 23 201912   2.77  0.67   1.91 0.14  0.09260502        0
## 24 202001  -0.11 -3.08  -6.27 0.13  0.05133505        0
## 25 202002  -8.13  1.02  -3.92 0.12 -0.12276945        0
## 26 202003 -13.38 -5.03 -13.96 0.12 -0.07289590        1
## 27 202004  13.65  2.75  -1.39 0.00  0.14390556        0
## 28 202005   5.58  2.49  -5.05 0.01  0.08112683        0
## 29 202006   2.46  2.71  -2.35 0.01  0.13694231        0
## 30 202007   5.77 -2.18  -1.39 0.01  0.15238742        0
## 31 202008   7.63 -0.25  -2.94 0.01  0.19545737        0
## 32 202009  -3.63  0.06  -2.51 0.01 -0.10873568        0
## 33 202010  -2.10  4.44   4.03 0.01 -0.06260496        0
## 34 202011  12.47  5.48   2.11 0.01  0.09050139        0
## 35 202012   4.63  4.91  -1.46 0.01  0.11543280        0
```

Next up, we run the linear regression on the original FF CAPM 3-factor using imported data:

```r
ffRegression_wDummy <- lm(fundExcess ~ rmrf + smb + hml + D)
summary(ffRegression_wDummy)
```

```
##
## Call:
## lm(formula = fundExcess ~ rmrf + smb + hml + D)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.277943 -0.042272 -0.003439  0.044782  0.175915
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.124194   0.019540  -6.356 5.18e-07 ***
## rmrf         0.016652   0.003649   4.564 7.97e-05 ***
## smb          0.007844   0.007310   1.073    0.292
## hml         -0.008407   0.006387  -1.316    0.198
## D            0.076188   0.139785   0.545    0.590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1009 on 30 degrees of freedom
## Multiple R-squared:  0.5299, Adjusted R-squared:  0.4672
## F-statistic: 8.454 on 4 and 30 DF,  p-value: 0.0001083
```

The estimated equation is $R_a - R_f = -.124303 + .016149(R_m - R_f) + .007402 * SMB - .010387 * HML + .076188 * D$, where the p_{value} for the intercept $3.58e - 07 < .05$.
Therefore, we reject the null hypothesis at 95% confidence level that the intercept α_{Jensen} statistically is no different from 0. Thus, we reject the null hypothesis $H_0 : \alpha = 0$ and accept the alternative hypothesis $H_a : \alpha \neq 0$.

Pandemic factor In addition, the coefficient $\beta_4 = .076188$ represents the increase in portfolio risk premium relative to increase in the pandemic variable. The p_{value} for β_4 is $.590 > .05$, implying that the coefficient β_4 statistically is insignificant at 95% or more, and we failed to reject the null hypothesis $H_0 : \beta_4 = 0$ and reject the alternative hypothesis $H_a : \beta_4 \neq 0$.

Goodness of Fit:

Through inspection of the linear regression result, we observe the $R^2 = .5299$ value to not be close to 1 at all. $R^2 = .5299$ implies that 52.99% of the variations in the portfolio risk premium is explained by the market risk premium.

Standard Error of Regression:

We can see that the Standard Error of Regression is $S.E. = .1009$.
From this, we can calculate the forecasting efficiency statistic to be:

$$\begin{aligned} \frac{S.E.}{\overline{Y}} =& \frac{.1009}{0.0326935} \\ =& 308.62\% > 10\% \end{aligned} \tag{12}$$

This statistic implies that the linear model is not a good forecasting model.

Remarks

After including the dummy pandemic variable, the linear regression analysis shows that the pandemic has little effect on AAPL returns, adjusting for size and value. This contradicts with our findings in the original single-factor CAPM regression analysis.

Question 2

a. Test the hypothesis that the mean returns of all four portfolios are not statistically different from each other

We make the following assumptions:

1. The observations are obtained independently and randomly from the population defined by the factor levels.
2. The data of each factor level are normally distributed.
3. These normal populations have a common variance.

Data Cleaning

First, we generate the data set as below:

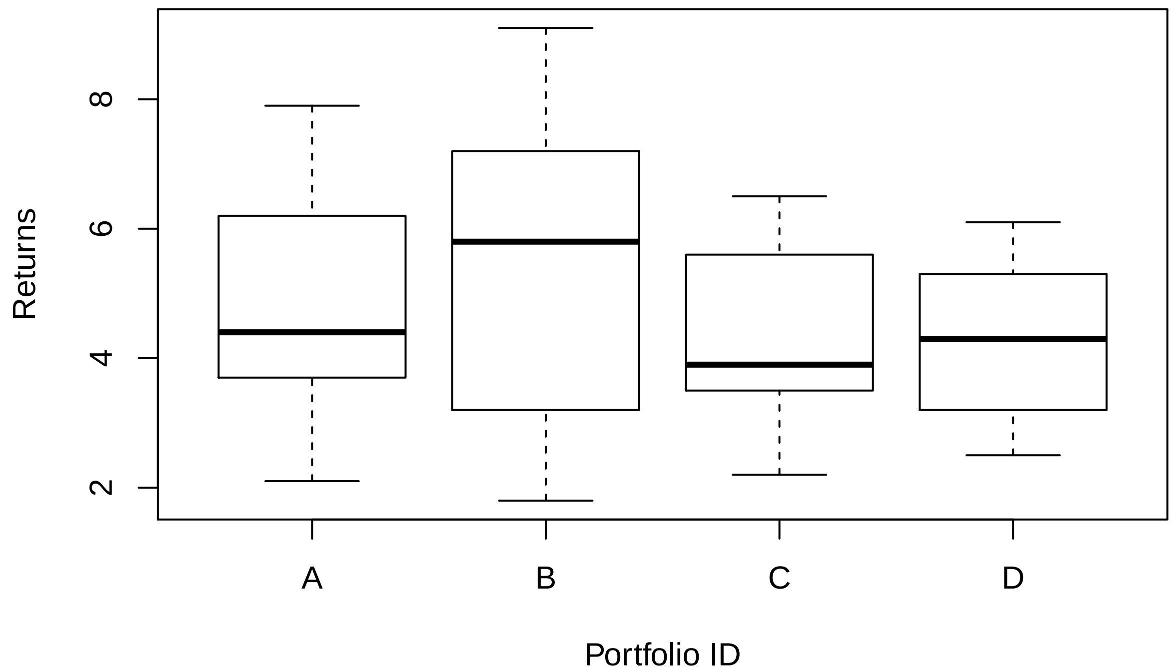
```
# Data entry of average annual returns for each portfolio
rA <- c(4.5, 3.8, 4.8, 3.7, 2.1, 3.1, 4.3, 7.9, 6.2, 6.6)
rB <- c(6.8, 5.2, 5.9, 5.7, 3.2, 1.8, 2.9, 8.2, 7.2, 9.1)
rC <- c(3.6, 4.7, 6.5, 3.5, 2.8, 2.2, 3.5, 5.6, 4.2, 5.9)
rD <- c(4.6, 3.2, 5.5, 2.9, 4.8, 4.0, 2.5, 5.3, 3.2, 6.1)
aggReturns <- c(rA, rB, rC, rD)

# Generate dataframe
```

```r
data <- data.frame(aggReturns)
data$portfolioID <- c(rep("A", 10), rep("B", 10),
                      rep("C", 10), rep("D", 10)) # Adding portfolio ID column
data$year <- rep(c(2005, 2006, 2007, 2008, 2009,
                   2010, 2011, 2012, 2013, 2014), 4)  # Adding year column
data$origin <- c(rep("domestic", 20), rep("foreign", 20))
data
```

```
##    aggReturns portfolioID year   origin
## 1         4.5           A 2005 domestic
## 2         3.8           A 2006 domestic
## 3         4.8           A 2007 domestic
## 4         3.7           A 2008 domestic
## 5         2.1           A 2009 domestic
## 6         3.1           A 2010 domestic
## 7         4.3           A 2011 domestic
## 8         7.9           A 2012 domestic
## 9         6.2           A 2013 domestic
## 10        6.6           A 2014 domestic
## 11        6.8           B 2005 domestic
## 12        5.2           B 2006 domestic
## 13        5.9           B 2007 domestic
## 14        5.7           B 2008 domestic
## 15        3.2           B 2009 domestic
## 16        1.8           B 2010 domestic
## 17        2.9           B 2011 domestic
## 18        8.2           B 2012 domestic
## 19        7.2           B 2013 domestic
## 20        9.1           B 2014 domestic
## 21        3.6           C 2005  foreign
## 22        4.7           C 2006  foreign
## 23        6.5           C 2007  foreign
## 24        3.5           C 2008  foreign
## 25        2.8           C 2009  foreign
## 26        2.2           C 2010  foreign
## 27        3.5           C 2011  foreign
## 28        5.6           C 2012  foreign
## 29        4.2           C 2013  foreign
## 30        5.9           C 2014  foreign
## 31        4.6           D 2005  foreign
## 32        3.2           D 2006  foreign
## 33        5.5           D 2007  foreign
## 34        2.9           D 2008  foreign
## 35        4.8           D 2009  foreign
## 36        4.0           D 2010  foreign
## 37        2.5           D 2011  foreign
## 38        5.3           D 2012  foreign
## 39        3.2           D 2013  foreign
## 40        6.1           D 2014  foreign
```

Next, we look at a side-by-side box plot of the 4 portfolio returns:

```
boxplot(data$aggReturns~data$portfolioID, xlab="Portfolio ID", ylab="Returns")
```



By inspection, we can see that the mean returns of portfolio A, C and D are similar and are different from that of portfolio B. We can also see that A, C and D have similar variance of returns wherease B has much larger variance of returns.

Hypothesis for regression

Next we test the hypothesis that the mean returns of all 4 portfolios are not statistically different from each other. In other words:

$$\begin{aligned} H_0 &: \mu_A = \mu_B = \mu_C = \mu_D \\ H_a &: \mu_A \neq \mu_B \neq \mu_C \neq \mu_D \\ &\text{where} \\ \mu_A, \mu_B, \mu_C, \mu_D &= \text{mean returns of A,B,C,D portfolios} \end{aligned} \tag{13}$$

Regression Analysis

We use dummy variable to run the ANOVA test.

```
# Portfolio Dummies for A
dummyA <- rep(0, 40)
dummyA[1:10] <- 1
```

```r
# Portfolio Dummies for B
dummyB <- rep(0, 40)
dummyB[11:20] <- 1

# Portfolio Dummies for C
dummyC <- rep(0, 40)
dummyC[21:30] <- 1

# Portfolio Dummies for D
dummyD <- rep(0, 40)
dummyD[30:40] <- 1

# Add dummy variable to the dataset
data <- cbind(data, dummyA, dummyB, dummyC, dummyD)
data
```

```
##    aggReturns portfolioID year   origin dummyA dummyB dummyC dummyD
## 1         4.5           A 2005 domestic      1      0      0      0
## 2         3.8           A 2006 domestic      1      0      0      0
## 3         4.8           A 2007 domestic      1      0      0      0
## 4         3.7           A 2008 domestic      1      0      0      0
## 5         2.1           A 2009 domestic      1      0      0      0
## 6         3.1           A 2010 domestic      1      0      0      0
## 7         4.3           A 2011 domestic      1      0      0      0
## 8         7.9           A 2012 domestic      1      0      0      0
## 9         6.2           A 2013 domestic      1      0      0      0
## 10        6.6           A 2014 domestic      1      0      0      0
## 11        6.8           B 2005 domestic      0      1      0      0
## 12        5.2           B 2006 domestic      0      1      0      0
## 13        5.9           B 2007 domestic      0      1      0      0
## 14        5.7           B 2008 domestic      0      1      0      0
## 15        3.2           B 2009 domestic      0      1      0      0
## 16        1.8           B 2010 domestic      0      1      0      0
## 17        2.9           B 2011 domestic      0      1      0      0
## 18        8.2           B 2012 domestic      0      1      0      0
## 19        7.2           B 2013 domestic      0      1      0      0
## 20        9.1           B 2014 domestic      0      1      0      0
## 21        3.6           C 2005  foreign      0      0      1      0
## 22        4.7           C 2006  foreign      0      0      1      0
## 23        6.5           C 2007  foreign      0      0      1      0
## 24        3.5           C 2008  foreign      0      0      1      0
## 25        2.8           C 2009  foreign      0      0      1      0
## 26        2.2           C 2010  foreign      0      0      1      0
## 27        3.5           C 2011  foreign      0      0      1      0
## 28        5.6           C 2012  foreign      0      0      1      0
## 29        4.2           C 2013  foreign      0      0      1      0
## 30        5.9           C 2014  foreign      0      0      1      1
## 31        4.6           D 2005  foreign      0      0      0      1
## 32        3.2           D 2006  foreign      0      0      0      1
## 33        5.5           D 2007  foreign      0      0      0      1
## 34        2.9           D 2008  foreign      0      0      0      1
## 35        4.8           D 2009  foreign      0      0      0      1
## 36        4.0           D 2010  foreign      0      0      0      1
```

```
## 37        2.5           D 2011  foreign      0      0      0      1
## 38        5.3           D 2012  foreign      0      0      0      1
## 39        3.2           D 2013  foreign      0      0      0      1
## 40        6.1           D 2014  foreign      0      0      0      1
```

We run ANOVA Test on the equality of mean returns for portfolio A, B, C, and D:

```
reg1 <- lm(aggReturns ~ dummyA + dummyB + dummyC, data=data)
summary(reg1)
```

```
##
## Call:
## lm(formula = aggReturns ~ dummyA + dummyB + dummyC, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.800 -1.010 -0.125  1.305  3.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.2100     0.5515   7.633 4.94e-09 ***
## dummyA        0.4900     0.7800   0.628   0.5338
## dummyB        1.3900     0.7800   1.782   0.0832 .
## dummyC        0.0400     0.7800   0.051   0.9594
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.744 on 36 degrees of freedom
## Multiple R-squared:  0.1026, Adjusted R-squared:  0.02783
## F-statistic: 1.372 on 3 and 36 DF,  p-value: 0.2669
```

Remarks

By examining the F-statistics's p-value of .2669 > .05, we accept the null hypothesis that the difference in mean returns of the portfolios are insignificant, i.e. $\mu_A = \mu_B = \mu_C = \mu_D$ and the mean return is 4.21.

b. Does data indicate any one portfolio have a better performance than others. If so, which and by how much?

From the ANOVA in a., we observe that *dummyC* is least significant. So we drop it and rerun the regression with *dummyD*.

```
reg2 <- lm(aggReturns ~ dummyA + dummyB + dummyD, data=data)
summary(reg2)
```

```
##
## Call:
## lm(formula = aggReturns ~ dummyA + dummyB + dummyD, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -3.800 -1.164 -0.050  1.275  3.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.0667     0.5802   7.009  3.2e-08 ***
## dummyA        0.6333     0.7998   0.792   0.4336
## dummyB        1.5333     0.7998   1.917   0.0632 .
## dummyD        0.2970     0.7824   0.380   0.7065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.741 on 36 degrees of freedom
## Multiple R-squared:  0.1061, Adjusted R-squared:  0.03163
## F-statistic: 1.425 on 3 and 36 DF,  p-value: 0.2515
```

Remarks

We can observe similar results, i.e. *dummyD* is also insignificant. Thus, there are no one portfolio that has better performance than others.

c. Test whether the GFC recession in 2008 and 2009 had any effect on the mean returns of portfolios.

Hypothesis for Regression

Our main hypothesis is that all portfolios have similar mean returns.
Our recession hypothesis is that the recession of 08 and 09 has no effect on our 4 portfolio returns. In other words:

$$\begin{aligned} &\text{Main hypothesis} \\ &H_0 : \mu_A = \mu_B = \mu_C = \mu_D \\ &H_a : \mu_A \neq \mu_B \neq \mu_C \neq \mu_D \\ &\text{Recession Hypothesis} \\ &H_0 : \beta_{Recession} = 0 \\ &H_a : \beta_{Recession} \neq 0 \end{aligned} \tag{14}$$

ANOVA Test on Effect of Recession on Portfolio Returns

We first generate recession dummy variable as follow:

```
dummyRecession <- rep(c(0, 0, 0, 1, 1, 0, 0, 0, 0, 0), 4)
data$dummyRecession <- dummyRecession
```

Running the regression:

```
regRec1 <- lm(aggReturns ~ dummyA + dummyB + dummyC + dummyRecession, data=data)
summary(regRec1)
```

```
##
## Call:
## lm(formula = aggReturns ~ dummyA + dummyB + dummyC + dummyRecession,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0756 -1.0631 -0.0756  1.2080  3.2244
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.4856     0.5436   8.251    1e-09 ***
## dummyA           0.4900     0.7459   0.657   0.5155
## dummyB           1.3900     0.7459   1.864   0.0708 .
## dummyC           0.0400     0.7459   0.054   0.9575
## dummyRecession  -1.3781     0.6593  -2.090   0.0439 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.668 on 35 degrees of freedom
## Multiple R-squared:  0.2022, Adjusted R-squared:  0.111
## F-statistic: 2.218 on 4 and 35 DF,  p-value: 0.08709
```

Remarks

We can observe that the p-value of the F-statistics is statisticall insignificant so we accept the null hypothesis that the portfoflios have similar mean returns.

In addition, we can observe that the *dummyRecession* variable is statistically significant at $.0439 < .05$. Thus we reject the null recession hypothesis that recession has no effect on portfolio returns. From the result of the regression summary, we can see the effect to be -1.3781.

d. For each portfolio, find the mean return during the recession and no-recession period

From above, we note that the mean returns of each portfolio is normally 4.4856 and $4.4856 - 1.3781 = 3.1075$ during the recession period.

e. Test whether the portfolio's origin (domestic/foreign) had any affect on the average return to portfolios.

Hypothesis for Regression

Our origin hypothesis is that the portfolio's origin has no effect on our 4 portfolio returns. In other words:

$$
\begin{aligned}
&\text{Origin Hypothesis} \\
H_0 : \beta_{origin} &= 0 \\
H_a : \beta_{origin} &\neq 0
\end{aligned}
\tag{15}
$$

ANOVA Test on Effect of Origin on Portfolio Returns

We first generate origin dummy variable as follow with the dummy value for domestic origin being 1 and foreign origin being 0.

```
dummyOrigin <- c(rep(1, 20), rep(0, 20))
data$dummyOrigin <- dummyOrigin
```

Running the regression:

```
regOrigin <- lm(aggReturns ~ dummyOrigin, data=data)
summary(regOrigin)
```

```
##
## Call:
## lm(formula = aggReturns ~ dummyOrigin, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.350 -1.335 -0.130  1.295  3.950
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.2300     0.3866  10.942 2.64e-13 ***
## dummyOrigin   0.9200     0.5467   1.683    0.101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.729 on 38 degrees of freedom
## Multiple R-squared:  0.06936,    Adjusted R-squared:  0.04487
## F-statistic: 2.832 on 1 and 38 DF,  p-value: 0.1006
```

Remarks

We can observe that the p-value of the F-statistics is $.1006 > .05$ and thus, statistically insiginificant at 95%. The coefficient β_{origin} is also statistically insignificant as its p-value is $.101 > .05$. Thus we accept the null recession hypothesis that portfolio's underlying origin has no effect on portfolio returns.