1.

(a) The portfolios are constructed at the end of each June, are the intersections of 10 portfolios formed on size (market equity, ME) and 10 portfolios formed on investment (Inv). The size breakpoints for year t are the NYSE market equity deciles at the end of June of year t. Investment is the change in total assets from the fiscal year ending in year t-2 to the fiscal year ending in t-1, divided by t-2 total assets. The Inv breakpoints are NYSE deciles. The portfolios for July of year t to June of t+1 include all NYSE, AMEX, and NASDAQ stocks for which we have market equity data for June of t and total assets data for t-2 and t-1.

(b) We use portfolios of assets because it reduces the errors-in-variables problem. Cross-sectional regressions specify estimated betas as regressors. If the errors in the estimated betas are somehow correlated across assets then the estimation errors would tend to offset each other when the assets are put together into test portfolios. Thus, using portfolios as test assets allows for more efficient estimates of factor loadings which will make factor risk premia to be estimated more precisely.

(c) The value weighted portfolio refers to the composition based on weights of individual stocks proportional to their market capitalization. In equal weighted portfolio all the assets (or stocks) have the same weight.

2.

(a)

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.0.2
```

```
library(readxl)

AMZN <- fread("AMZN.csv", select=c(1, 5))
prices = AMZN$Close

AMZN = AMZN[-1,]
n <- length(prices);
AMZN$log_returns <- log(prices[-1]/prices[-n])*100

AMZN = AMZN[1:251]
data_capm <- fread("FF.csv", select=c(1, 2, 5))
```

```
## Warning in fread("FF.csv", select = c(1, 2, 5)): Stopped early on line 1147.
## Expected 5 fields but found 0. Consider fill=TRUE and comment.char=. First
## discarded non-empty line: <<Annual Factors: January-December >>
```

```
data_capm <- data_capm[892:1142]

Y1 = AMZN$log_returns

Mkt <- data_capm$'Mkt-RF'+ data_capm$RF
model1 = lm(Y1 ~ Mkt)
summary(model1)
```

```
##
## Call:
## lm(formula = Y1 ~ Mkt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.344  -5.275   0.007   5.532  36.704
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6651     0.7006   0.949    0.343
## Mkt           1.5002     0.1547   9.698   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.94 on 249 degrees of freedom
## Multiple R-squared:  0.2741, Adjusted R-squared:  0.2712
## F-statistic: 94.04 on 1 and 249 DF,  p-value: < 2.2e-16
```

```
confint(model1, level=0.95)
```

```
##                  2.5 %   97.5 %
## (Intercept) -0.7147282 2.044852
## Mkt          1.1955436 1.804931
```

i. R = 0.6651 + 1.5002(Mkt)

This model doesn't substract risk free rate from AMZN log returns and market excess returns which whould result in R = 0.7206 + 1.5025(Mkt). However, it doesn't change the conclusion.

ii. `2.5 %   97.5 %`

    (Intercept) -0.7147282 2.044852 Mkt 1.1955436 1.804931

iii. Alpha = 0.6651 and is statistically signifcantly different from 0. Its p-value is 0.343, it is larger than 0.05 or 0.1, therefore we do not reject the null hypothese that the alpha is statistically signifcantly different from 0. This indicates that Amazon stock does not seem to signifcantly outperform or underperform the overall market.

iv. B = 1.5002

H0: B = 1 H1: B > 1

t = (1.5002 - 1) / 0.1547 = 3.23

t > 1.645 (crtitical t)

.: We reject the null. It is more risky than the market and it will have higher return.

b)

i.

```r
P100 <- fread("100Portf.csv")
```

```
## Warning in fread("100Portf.csv"): Stopped early on line 716. Expected 101 fields
## but found 0. Consider fill=TRUE and comment.char=. First discarded non-empty
## line: <<Average Equal Weighted Returns -- Monthly>>
```

```r
P100 = P100[219:698]
P100 = as.matrix(P100)
P100 <- P100[,-1]

data_rf <- fread("FF.csv", select=c(1, 2, 5))
```

```
## Warning in fread("FF.csv", select = c(1, 2, 5)): Stopped early on line 1147.
## Expected 5 fields but found 0. Consider fill=TRUE and comment.char=. First
## discarded non-empty line: <<Annual Factors: January-December >>
```

```r
data_rf <- data_rf[663:1142]

y = as.matrix(data_rf$RF)
mkt = as.matrix(data_rf$'Mkt-RF')

Alphas = c()
Betas = c()

for(i in 1:ncol(P100)) {
    Y2 = P100[, i] - y
    model <- lm(Y2 ~ mkt)
    Alphas[i] <- model$coefficients[1]
    Betas[i] <- model$coefficients[2]
}

AVGs = c()
for(i in 1:ncol(P100)) {
    avg = sum(P100[, i] - y)/480
    AVGs[i] <- avg
}

model_final <- lm(AVGs ~ Betas)

rf = sum(y)/480
rf
```

```
## [1] 0.3048333
```

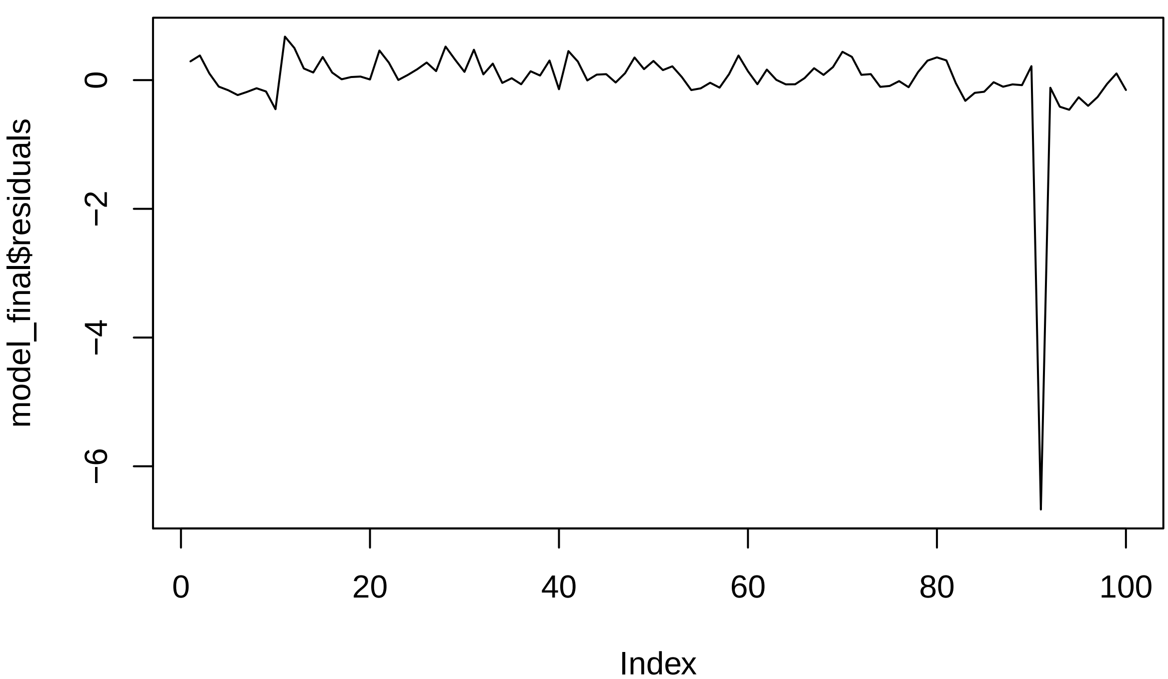
```r
avg_mkt = sum(mkt) /480
avg_mkt
```

```
## [1] 0.7549167
```

```r
summary(model_final)
```

```
##
## Call:
## lm(formula = AVGs ~ Betas)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6718 -0.0933  0.0756  0.2057  0.6757
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3931     0.5050   4.739 7.28e-06 ***
## Betas        -1.4733     0.4675  -3.151  0.00216 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7138 on 98 degrees of freedom
## Multiple R-squared:  0.092,  Adjusted R-squared:  0.08273
## F-statistic: 9.929 on 1 and 98 DF,  p-value: 0.002158
```

```r
plot(model_final$residuals, type="l")
```



```r
Box.test(model_final$residuals, lag=4, type=c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  model_final$residuals
## X-squared = 2.5542, df = 4, p-value = 0.635
```

ii. p-value for intercept is smaller than 0.05, therefore it is statistically significant to 0. The intercept/alpha is 2.3931 with a p-value of 0.00, which is lower than 0.05. This indicates that we will reject the null hypothesis and conclude it is statistically signifcantly to 0. Thus,the CAPM doesnt not hold. The beta is -1.4733, which is lower than 1 This indicates that it is more risky than the market.

iii. The mean of the residual looks steady at the first part, however there is a # sharp drop towards the end. Variance looks a bit unstable as well.

iv. In step 2, estimated intercept [2.3931] doesn't equal to average risk free rate [0.3048333] and estimated coefficient [-1.4733] doesn't equal to (average market return [0.754916] - risk free rate [0.3048333]). Therefore, CAPM model is not suitable.

3.

(a)

```r
P100 <- fread("100Portf.csv")
```

```
## Warning in fread("100Portf.csv"): Stopped early on line 716. Expected 101 fields
## but found 0. Consider fill=TRUE and comment.char=. First discarded non-empty
## line: <<Average Equal Weighted Returns -- Monthly>>
```

```r
P100 = P100[219:698]
P100 = as.matrix(P100)
P100 <- P100[,-1]

data_rf2 <- fread("FF.csv")
```

```
## Warning in fread("FF.csv"): Stopped early on line 1147. Expected 5 fields but
## found 0. Consider fill=TRUE and comment.char=. First discarded non-empty line:
## <<Annual Factors: January-December >>
```

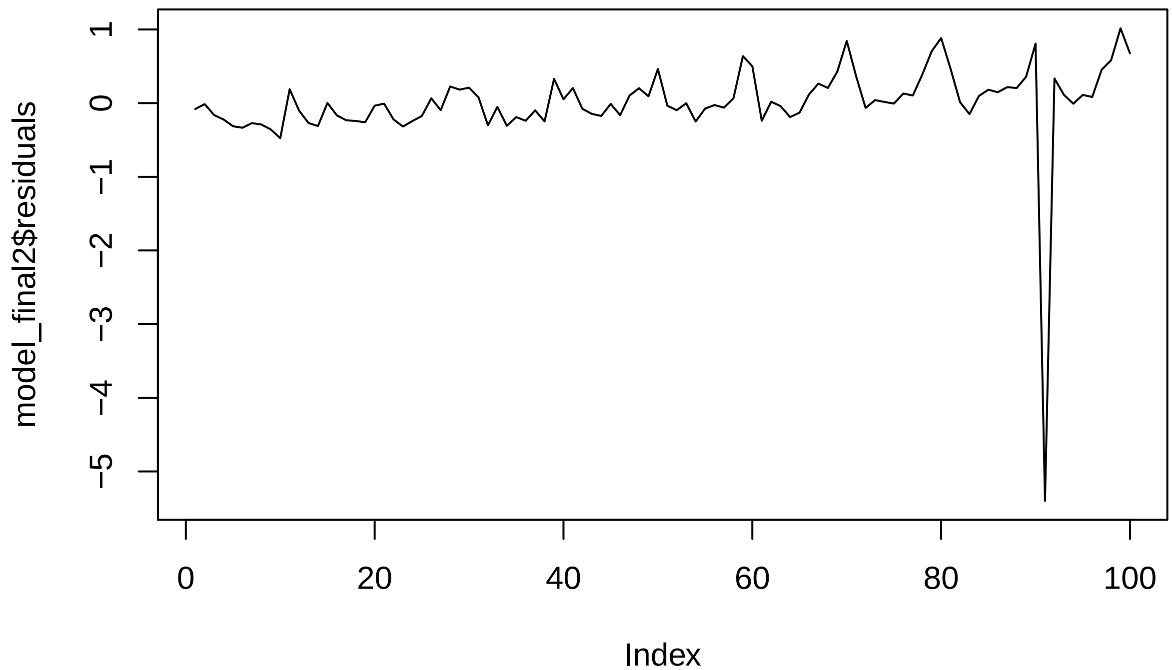
```r
data_rf2 <- data_rf2[663:1142]

y = as.matrix(data_rf2$RF)
mkt = as.matrix(data_rf2$'Mkt-RF')
SMB = as.matrix(data_rf2$SMB)
HML = as.matrix(data_rf2$HML)

Betas_mkt = c()
Betas_SMB = c()
Betas_HML = c()
```

```r
for(i in 1:ncol(P100)) {
    Y2 = P100[, i] - y
    model <- lm(Y2 ~ mkt + SMB + HML)
    Betas_mkt[i] <- model$coefficients[2]
    Betas_SMB[i] <- model$coefficients[3]
    Betas_HML[i] <- model$coefficients[4]
}

AVGs2 = c()
for(i in 1:ncol(P100)) {
    avg = sum(P100[, i] - y)/480
    AVGs2[i] <- avg
}

model_final2 <- lm(AVGs2 ~ Betas_mkt + Betas_SMB + Betas_HML)
```

```r
summary(model_final2)
```

```
##
## Call:
## lm(formula = AVGs2 ~ Betas_mkt + Betas_SMB + Betas_HML)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.3995 -0.1742 -0.0067  0.2032  1.0155
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.1142     0.7708   1.445 0.151574
## Betas_mkt    -0.6712     0.7230  -0.928 0.355553
## Betas_SMB     0.1949     0.1450   1.344 0.182080
## Betas_HML     1.2100     0.3100   3.903 0.000176 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.633 on 96 degrees of freedom
## Multiple R-squared:  0.3006, Adjusted R-squared:  0.2787
## F-statistic: 13.75 on 3 and 96 DF,  p-value: 1.558e-07
```

```r
plot(model_final2$residuals, type="l")
```




```r
anova(model_final2)
```

```
## Analysis of Variance Table
##
## Response: AVGs2
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Betas_mkt  1  7.687  7.6869  19.186 3.026e-05 ***
## Betas_SMB  1  2.738  2.7385   6.835 0.0103798 *
## Betas_HML  1  6.105  6.1048  15.237 0.0001762 ***
## Residuals 96 38.463  0.4007
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The mean and the variances of the residual look very volatile, especially there is a sharp drop on the index 90.

The intercept/alpha is 1.1142 with a p-value of 0.1515, which is higher than 0.05 or 0.1. This indicates that we will not reject the null hypothesis which is statistically signifanct to 0. CAPM doesnt not hold. The beta for makert excess is -0.6712, which is smaller than 1. This indicates that it is less risky than the market.

The intercept/alpha is 1.1142 with a p-value of 0.1515, which is higher than 0.05 or 0.1. This indicates that we will not reject the null hypothesis which is statistically significant to 0. CAPM doesn't not hold. The beta for market excess is -0.6712, which is smaller than 1. This indicates that it is less risky than the market.

The beta for SMB is 0.1949 with a p-value of 0.182080. Since the p-value is greater than 0.05 or 0.1, it is statistically signifcantly different from 0. Also, an increase by 1 of SMB will result an increase in the excess return by 0.1949

The beta for HML is 1.2100 with a p-value of 0.000176. Since the p-value is lower than 0.05 or 0.1, it is statistically signifcantly to 0. Also, an increase by 1 of HML will result in increase in the excess return by 1.2100.

(b)

i. From the graphs below we can conclude that betas are time varying and change during the particular interval.

```
P100 <- fread("100Portf.csv")
```

```
## Warning in fread("100Portf.csv"): Stopped early on line 716. Expected 101 fields
## but found 0. Consider fill=TRUE and comment.char=. First discarded non-empty
## line: <<Average Equal Weighted Returns -- Monthly>>
```

```
P100 = P100[219:698]
P100 = as.matrix(P100)
P100 <- P100[,-1]

data_rf2 <- fread("FF.csv")
```

```
## Warning in fread("FF.csv"): Stopped early on line 1147. Expected 5 fields but
## found 0. Consider fill=TRUE and comment.char=. First discarded non-empty line:
## <<Annual Factors: January-December >>
```

```
data_rf2 <- data_rf2[663:1142]

y = as.matrix(data_rf2$RF)
mkt = as.matrix(data_rf2$'Mkt-RF')
SMB = as.matrix(data_rf2$SMB)
HML = as.matrix(data_rf2$HML)

myfun <- function(data, i_begin, i_end) {
  Res = matrix(0, nrow=0, ncol=4)
  for(i in 1:ncol(data)) {
    Temp = c()
    Y2 = data[i_begin:i_end, i] - y[i_begin:i_end]
    model <- lm(Y2 ~ mkt[i_begin:i_end] + SMB[i_begin:i_end] + HML[i_begin:i_end])
    Temp <- model$coefficients
    Res <- rbind(Res, Temp)
  }
  colnames(Res) <- c("Incpt","Mkt_B","SMB_B","HML_B")
  return(Res)
}

Period1 <- myfun(P100, 1, 96)
Period2 <- myfun(P100, 97, 192)
Period3 <- myfun(P100, 193, 288)
Period4 <- myfun(P100, 289, 384)
Period5 <- myfun(P100, 385, 480)
```

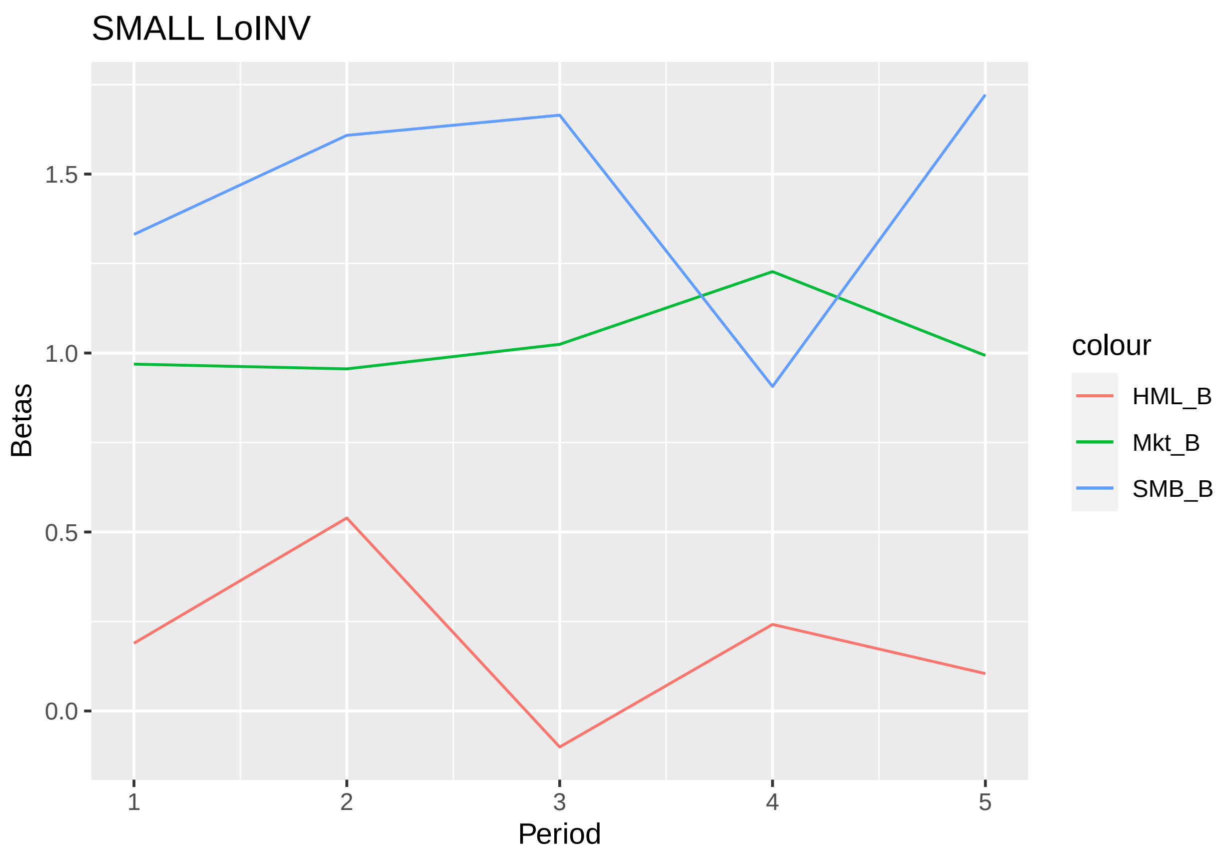
```r
P100_small <- matrix(0, nrow=0, ncol=4)
P100_small <- rbind(P100_small, Period1[1,], Period2[1,], Period3[1,], Period4[1,], Period5[1,])

df <- as.data.frame(P100_small)
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

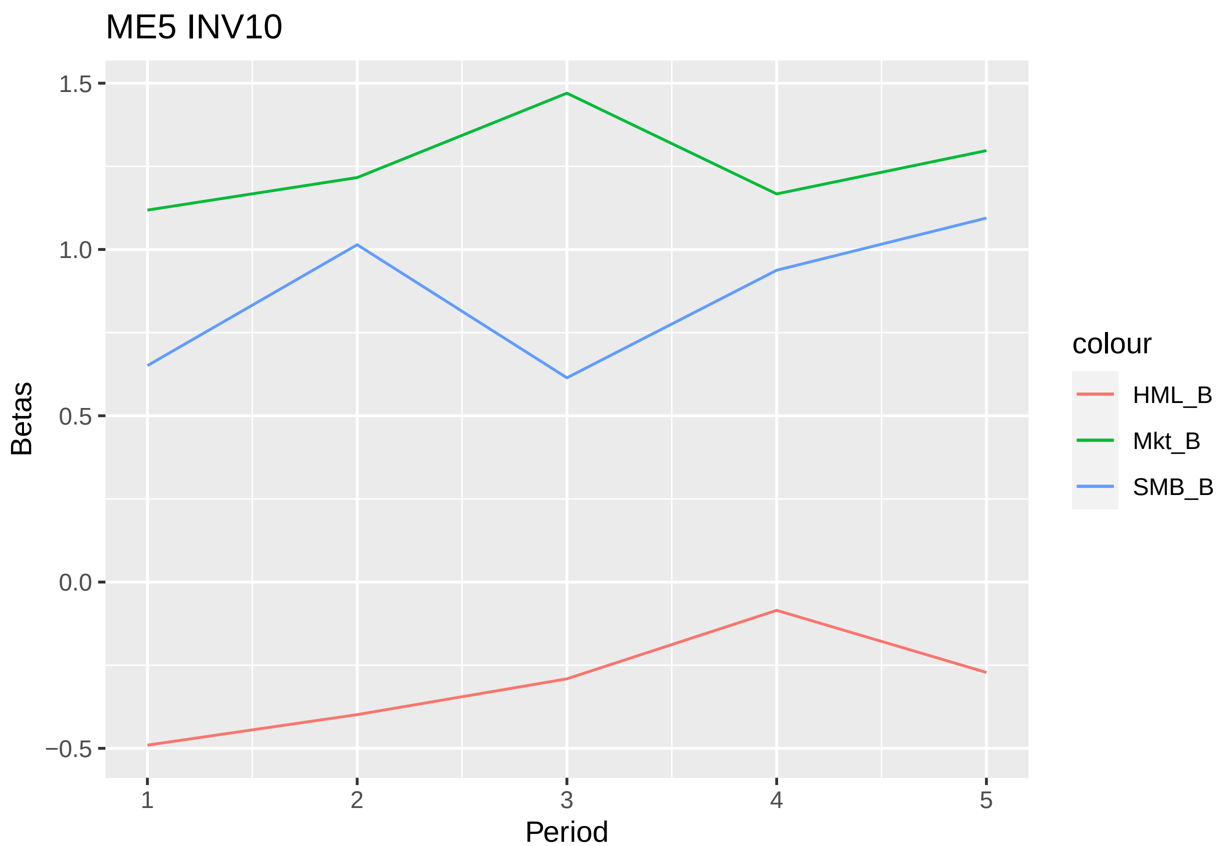
```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```r
ggplot(data = df, aes(x = as.numeric(row.names(df)))) +
  geom_line(aes(y = Mkt_B, color='Mkt_B')) +
  geom_line(aes(y = SMB_B, color='SMB_B')) +
  geom_line(aes(y = HML_B, color='HML_B')) + ggtitle("SMALL LoINV") + labs(y="Betas", x = "Period")
```

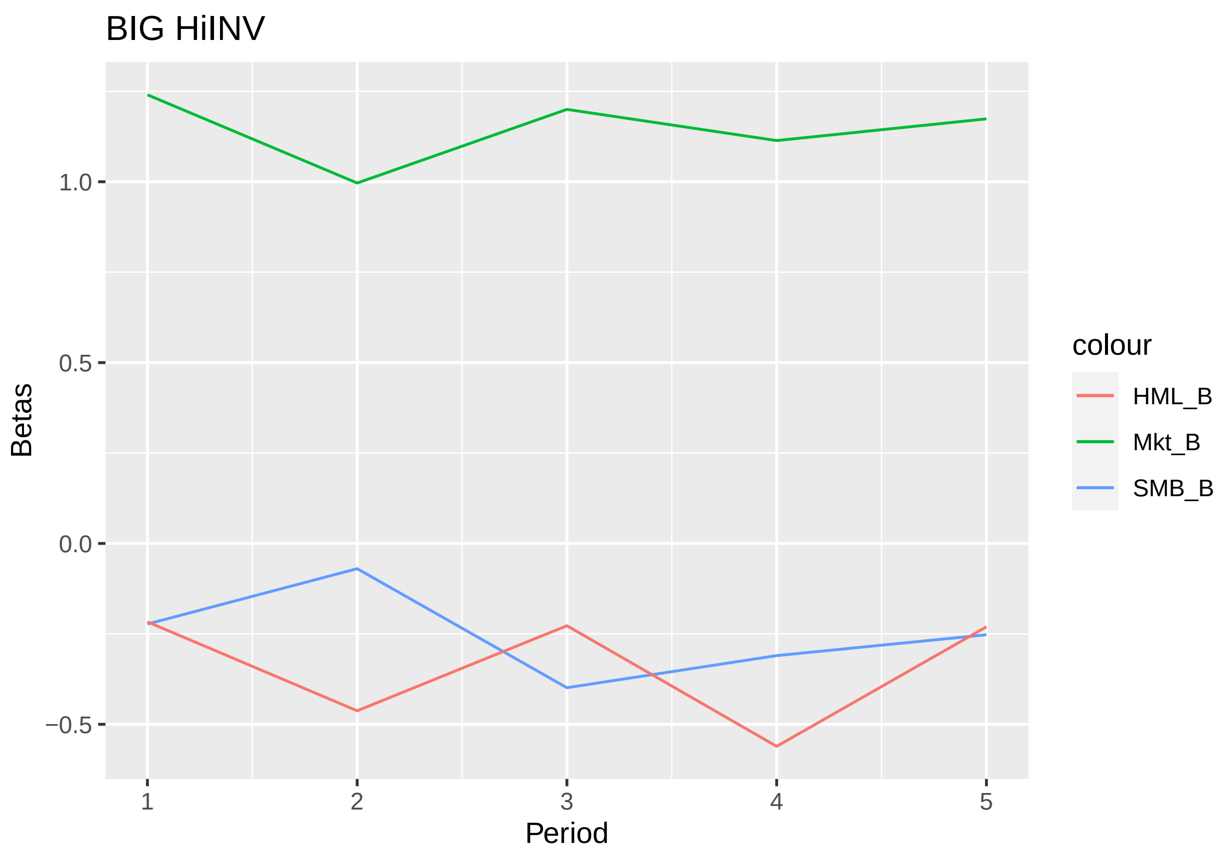



```r
P100_ME5 <- matrix(0, nrow=0, ncol=4)
P100_ME5 <- rbind(P100_ME5, Period1[50,], Period2[50,], Period3[50,], Period4[50,], Period5[50,])

df1 <- as.data.frame(P100_ME5)
```

```
ggplot(data = df1, aes(x = as.numeric(row.names(df)))) +
  geom_line(aes(y = Mkt_B, color='Mkt_B')) +
  geom_line(aes(y = SMB_B, color='SMB_B')) +
  geom_line(aes(y = HML_B, color='HML_B')) + ggtitle("ME5 INV10") + labs(y="Betas", x = "Period")
```




```
P100_BIG <- matrix(0, nrow=0, ncol=4)
P100_BIG <- rbind(P100_BIG, Period1[100,], Period2[100,], Period3[100,], Period4[100,], Period5[100,])

df2 <- as.data.frame(P100_BIG)
ggplot(data = df2, aes(x = as.numeric(row.names(df)))) +
  geom_line(aes(y = Mkt_B, color='Mkt_B')) +
  geom_line(aes(y = SMB_B, color='SMB_B')) +
  geom_line(aes(y = HML_B, color='HML_B')) + ggtitle("BIG HiINV") + labs(y="Betas", x = "Period")
```



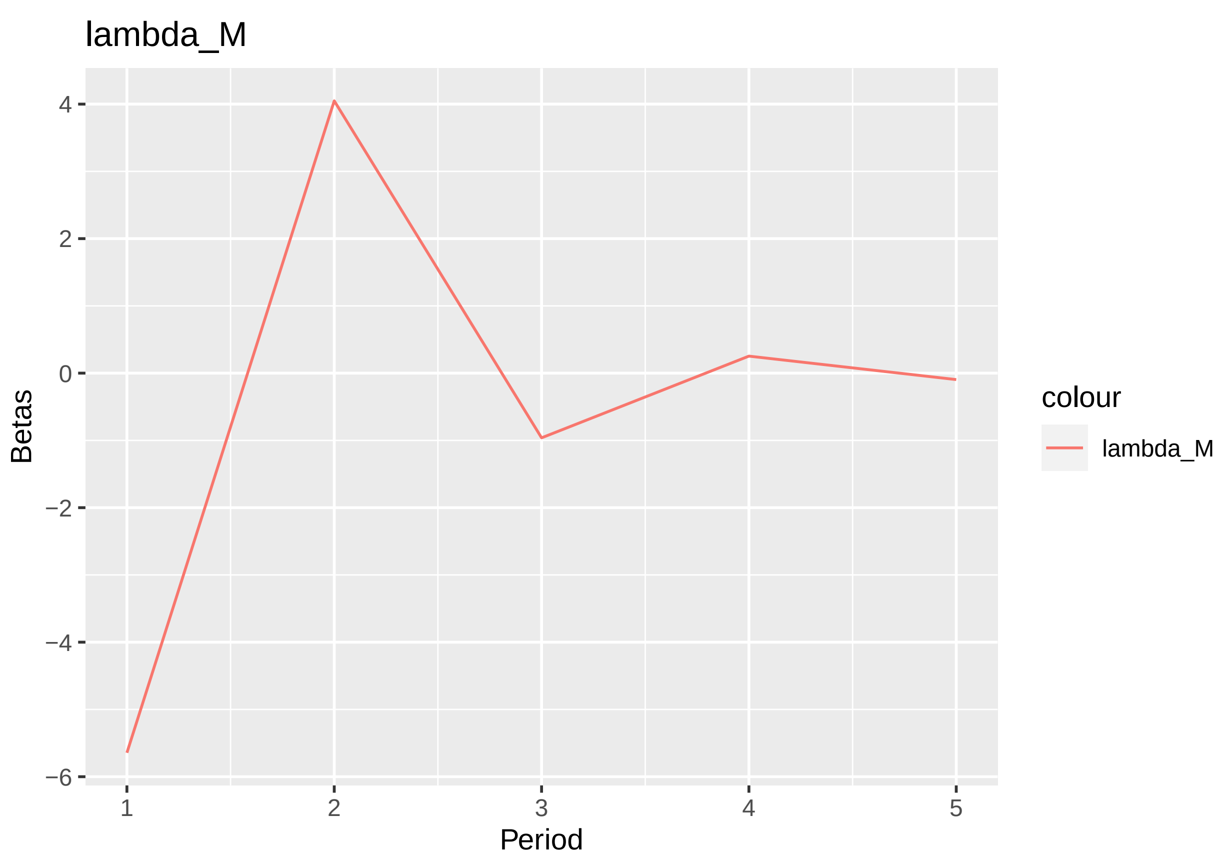
ii. From the graphs below we can conlude that the risk premias are time varying and volatile as well. Some of them seem to show a trend.

```
avgsfun <- function(data, beg_i, end_i) {
  res = c()
  for(i in 1:ncol(data)) {
      avg = sum(data[beg_i:end_i, i] - y[beg_i:end_i])/96
      res[i] <- avg
  }
  return (res)
}

Period1_avgs <- avgsfun(P100, 1, 96)
Period2_avgs <- avgsfun(P100, 97, 192)
Period3_avgs <- avgsfun(P100, 193, 288)
Period4_avgs <- avgsfun(P100, 289, 384)
Period5_avgs <- avgsfun(P100, 385, 480)


myfun_step2 <- function(avgs, data) {
  res = matrix(0, nrow=0, ncol=4)
  temp = c()
  mod <- lm(avgs ~ data[,2] + data[,3] + data[,4])
  temp <- mod$coefficients
  res <- rbind(res, temp)
```

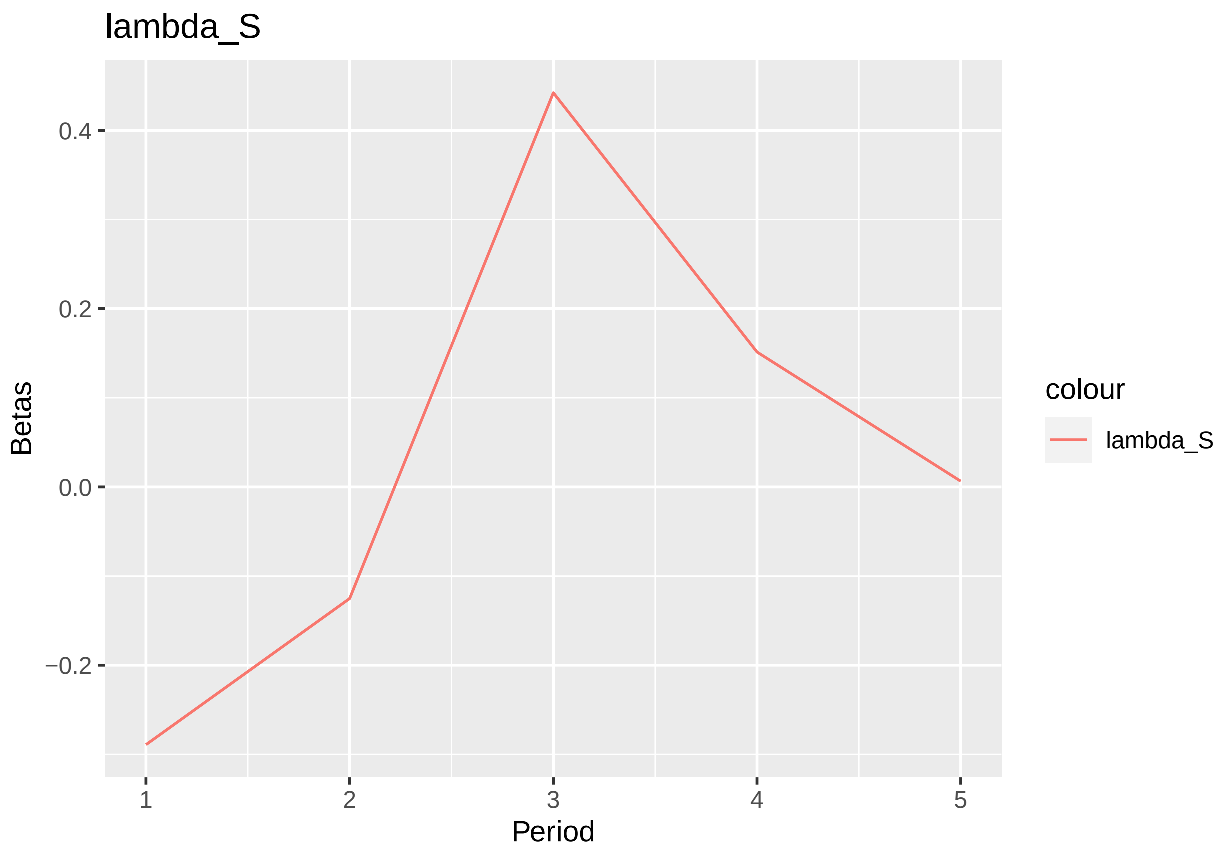
```r
  colnames(res) <- c("a","lambda_M","lambda_S","lambda_V")
  return(res)
}

Period1_Lambdas <- myfun_step2(Period1_avgs, Period1)
Period2_Lambdas <- myfun_step2(Period2_avgs, Period2)
Period3_Lambdas <- myfun_step2(Period3_avgs, Period3)
Period4_Lambdas <- myfun_step2(Period4_avgs, Period4)
Period5_Lambdas <- myfun_step2(Period5_avgs, Period5)

step2_matrix = matrix(0, nrow=0, ncol=4)
step2_matrix <- rbind(step2_matrix, Period1_Lambdas, Period2_Lambdas, Period3_Lambdas, Period4_Lambdas,
rownames(step2_matrix) <- c("1","2","3","4","5")
df_final <- as.data.frame(step2_matrix)
df_final
```

```
##            a   lambda_M     lambda_S   lambda_V
## 1  6.4850240 -5.6447175 -0.289214732  3.2497372
## 2 -3.6005368  4.0487726 -0.125189445  2.3996236
## 3  1.4493897 -0.9616746  0.442237564  0.3317643
## 4  0.3030035  0.2532065  0.151358507  0.1416540
## 5  1.3364768 -0.0956817  0.006419641 -0.3006786
```

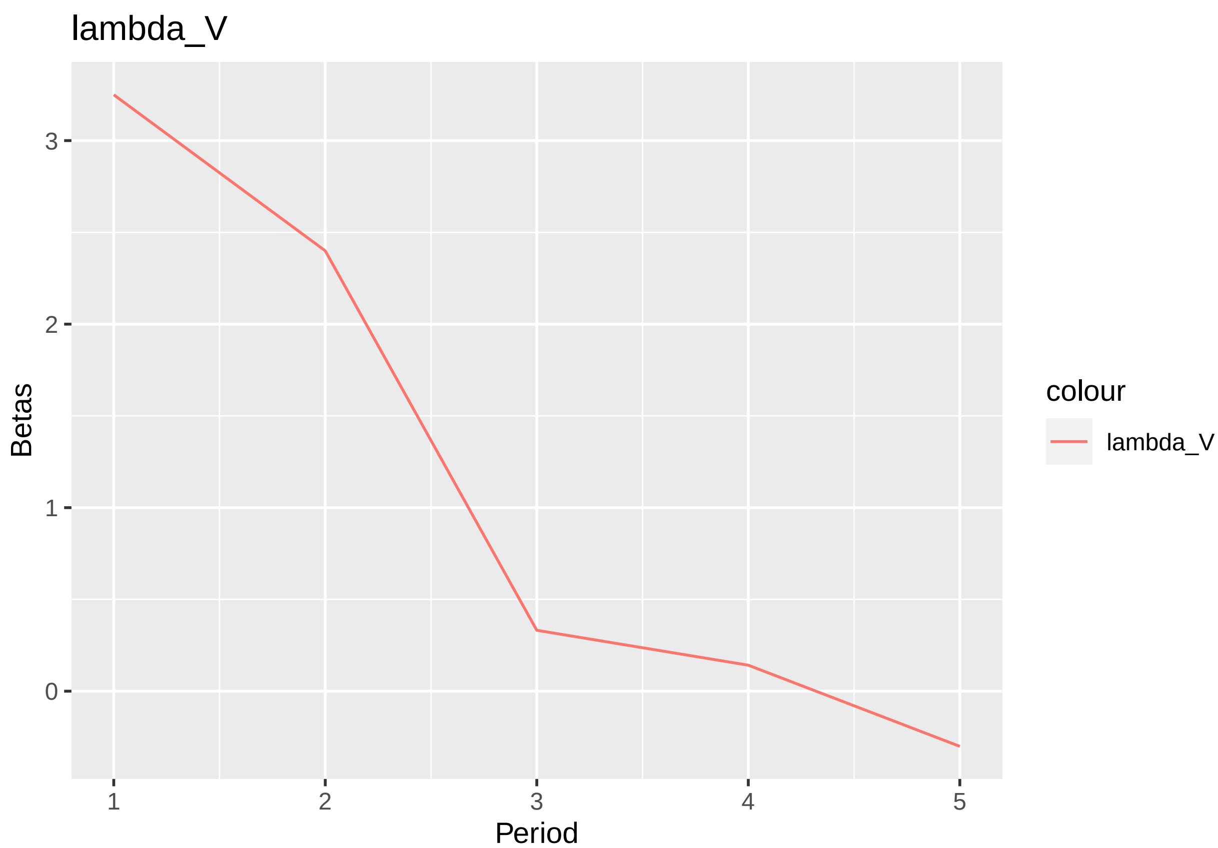
```r
ggplot(data = df_final, aes(x = as.numeric(row.names(df_final)))) +
  geom_line(aes(y = lambda_M, color='lambda_M')) + ggtitle("lambda_M") + labs(y="Betas", x = "Period")
```



```r
ggplot(data = df_final, aes(x = as.numeric(row.names(df_final)))) +
  geom_line(aes(y = lambda_S, color='lambda_S')) + ggtitle("lambda_S") + labs(y="Betas", x = "Period")
```




```r
ggplot(data = df_final, aes(x = as.numeric(row.names(df_final)))) +
  geom_line(aes(y = lambda_V, color='lambda_V')) + ggtitle("lambda_V") + labs(y="Betas", x = "Period")
```


lambda_V
Betas
3
2
1
0
1
2
3
4
5
Period
colour
lambda_V