HW4

Group 9: Linqi Huang, Abhesh Kumar, Yu Onohara, Maitrayee Patil, Redmond Xia'

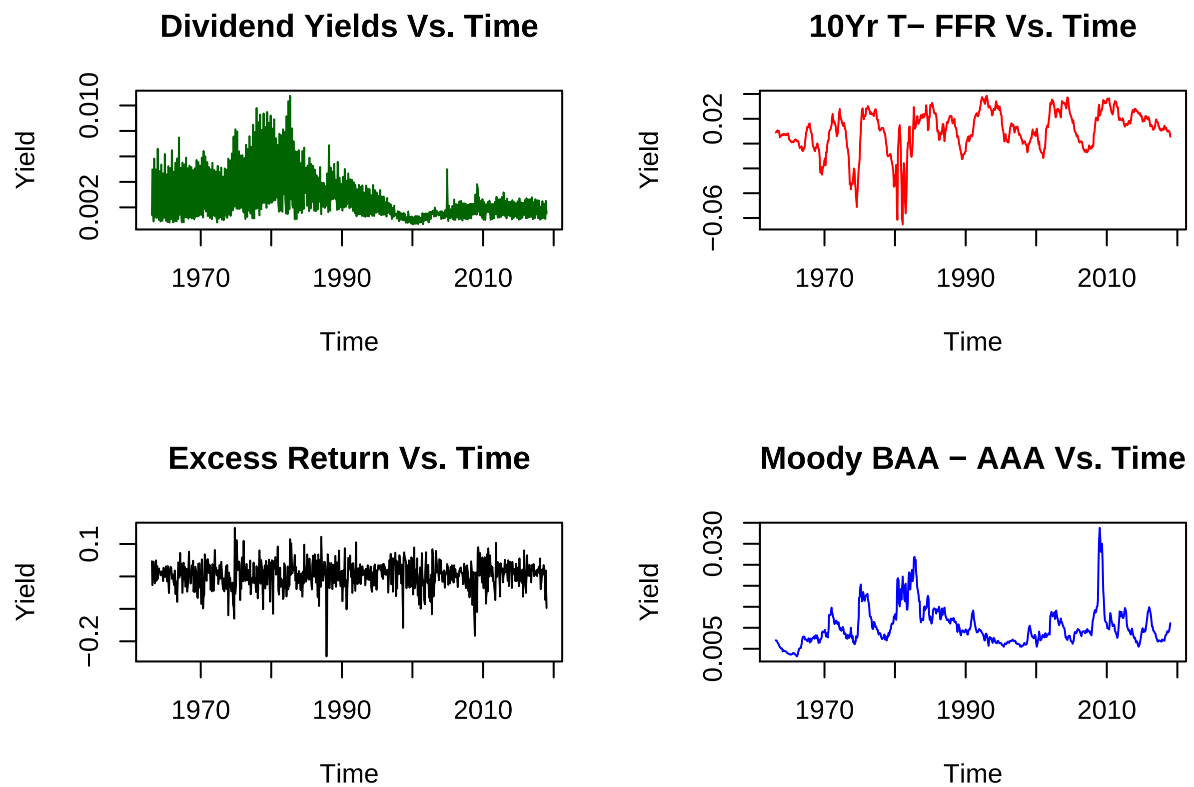
January 31, 2020

Problem 1

1

```r
# Cleaning and reading the data
suppressMessages(suppressWarnings(library(readr)))
sp500 <- read_csv("sp500.csv", col_types = cols(caldt = col_date(format = "%Y%m%d")))
sp500$divYield <- (sp500$vwretd - sp500$vwretx)
colnames(sp500)[1] <- "DATE"
sp500$DATE <- seq(as.Date("1963-02-01"),length = nrow(sp500), by ="months") - 1
rf <- read.csv("fedfund.csv")[-1]
rf$MCALDT <- as.Date(rf$MCALDT, format = "%m/%d/%Y")
colnames(rf)[1] <- "DATE"
rf$DATE <- seq(as.Date("1963-02-01"),length = nrow(rf), by ="months") - 1
rf$TMYTM <- rf$TMYTM / 12 / 100
AAAFFM <- read.csv("AAAFFM.csv")
AAAFFM$AAAFFM <- AAAFFM$AAAFFM / 100
AAAFFM$DATE <- seq(as.Date("1963-02-01"),length = nrow(AAAFFM), by ="months") - 1
BAAFFM <- read.csv("BAAFFM.csv")
BAAFFM$BAAFFM <- BAAFFM$BAAFFM / 100
BAAFFM$DATE <- seq(as.Date("1963-02-01"),length = nrow(BAAFFM), by ="months") - 1
AAFFM <- merge(AAAFFM,BAAFFM, by = "DATE")
AAFFM$BAA_AAA <- AAFFM$BAAFFM - AAFFM$AAAFFM
T10YFFM <- read.csv("T10YFFM.csv")
T10YFFM$DATE <- seq(as.Date("1963-02-01"),length = nrow(T10YFFM), by ="months") - 1
T10YFFM$T10YFFM <- T10YFFM$T10YFFM / 100
merged <- merge(merge(merge(sp500,rf, by = "DATE"),AAFFM[,c(-2,-3)],by = "DATE"),T10YFFM,by = "DATE")
merged$EXCESS <- log(merged$vwretd+1) - log(merged$TMYTM+1)
```

2



3

The standard error looks consistent for each regression. It also looks relatively large for small predictors. It's confidence bound is a little too large.

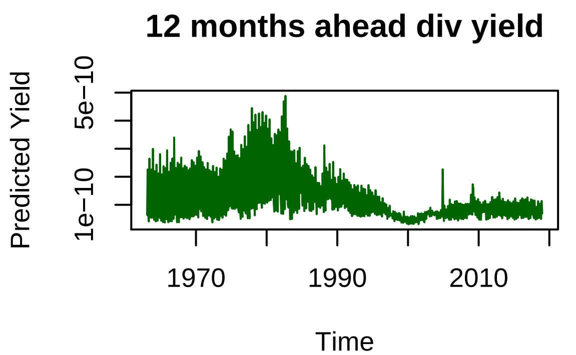
```
##
## Call:
## lm(formula = EXCESS ~ divYield, data = merged)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.250652 -0.022988  0.003715  0.027429  0.145371
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.003748   0.002777   1.350    0.178
## divYield    0.219932   0.882278   0.249    0.803
##
## Residual standard error: 0.04248 on 670 degrees of freedom
## Multiple R-squared:  9.274e-05,  Adjusted R-squared:  -0.0014
## F-statistic: 0.06214 on 1 and 670 DF,  p-value: 0.8032
```

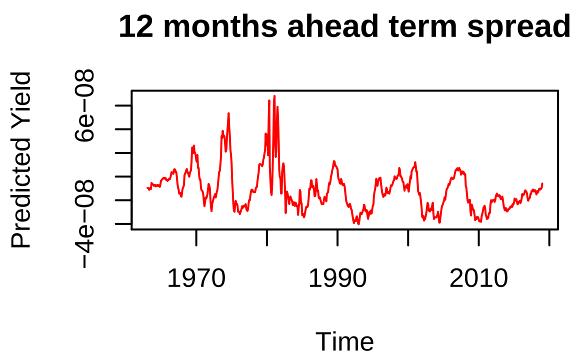
```
##
## Call:
```

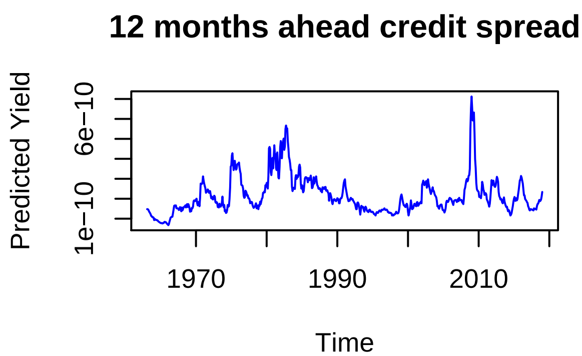
```
## lm(formula = EXCESS ~ T10YFFM, data = merged)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.254198 -0.023477  0.002497  0.027496  0.155040
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.001183   0.001927   0.614  0.53961
## T10YFFM     0.295270   0.097428   3.031  0.00253 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0422 on 670 degrees of freedom
## Multiple R-squared:  0.01352,    Adjusted R-squared:  0.01205
## F-statistic: 9.185 on 1 and 670 DF,  p-value: 0.002534
```

```
##
## Call:
## lm(formula = EXCESS ~ BAA_AAA, data = merged)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.250891 -0.023157  0.003785  0.027636  0.145160
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.002207   0.004082   0.541    0.589
## BAA_AAA     0.204745   0.364549   0.562    0.575
##
## Residual standard error: 0.04248 on 670 degrees of freedom
## Multiple R-squared:  0.0004706,  Adjusted R-squared:  -0.001021
## F-statistic: 0.3154 on 1 and 670 DF,  p-value: 0.5745
```

4.







Based on the plot, it seems like the term spread increases when dividend yields are really small. Dividends are small during bad times and really big during good times like during the hyper inflation of 80s. Credit spread has a weak correlation with dividend yield but it follows recession better. Thus these patterns do make sense. It shows that when there are extreme times, things are very volatile.

5

I believe using an AR(1) type setting could be useful instead because the regression doesn't match seasonality correctly. Instead the regression typically overfitts the data and so it follows the noise more closely. Thus we get an average error rate of 1%. The error rate are found below. This is very significant for monthly data who is yield is roughly just one percent.

```
errDiv <- mean( (forcast12m[[1]] - merged$divYield)^2)
sqrt(errDiv)
```

```
## [1] 0.003147867
```

```
errT10FMM <-  mean( (forcast12m[[2]] - merged$T10YFFM)^2)
sqrt(errT10FMM)
```

```
## [1] 0.01977712
```

```
errBAA <- mean( (forcast12m[[3]] - merged$BAA_AAA)^2)
sqrt(errBAA)
```

```
## [1] 0.0111974
```