CRSP Return Forecast

1/30/2021

```
library(data.table)
library(tidyverse)
library(magrittr)
library(xts)
library(lubridate)
library(xtable)
library(tidyquant)
library(timetk)
library(dplyr)
library(kableExtra)
library(knitr)
library(psych)
library(prettydoc)
library(readxl)
vwretd <- read_xlsx("~/Documents/Documents/Empirical Methods/valueweighted.xlsx")
vwretd$`Calendar Date` <- as.Date(vwretd$`Calendar Date`, format = '%d/%m/%Y')
yield <- read_xlsx("~/Documents/Documents/Empirical Methods/yield.xlsx")
yield$MCALDT <- as.Date(yield$MCALDT, format = '%Y-%m-%d')
t_10 <- read_xls("~/Documents/Documents/Empirical Methods/T10YFF.xls")
t_10$observation_date <- as.Date(t_10$observation_date, format = '%Y-%m-%d')
aaa <- read_xls("~/Documents/Documents/Empirical Methods/AAAFF.xls")
baa <- read_xls("~/Documents/Documents/Empirical Methods/DBAA.xls")
```

Problem 1: Return forecasting regressions

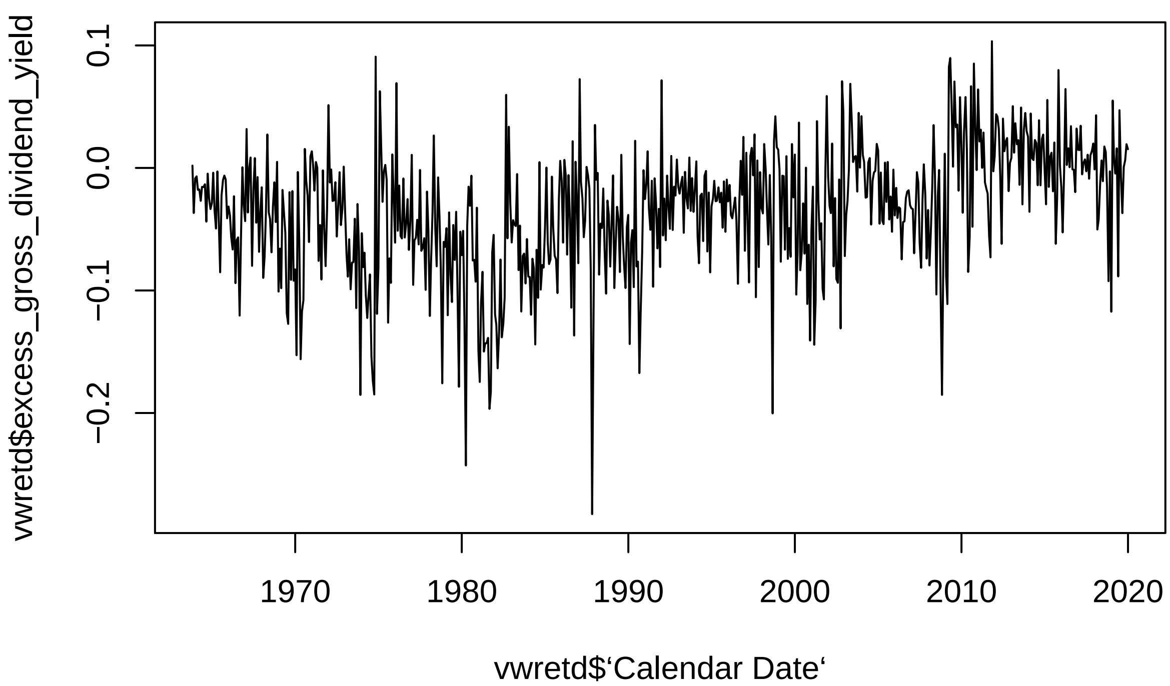
1. Using CRSP, get monthly market returns ex and cum dividends, as well as the monthly t-bill rate, from 1963 through 2019. Create the market dividend yield by summing the dividends over the last 12 months and divide by current price (you can do this using information extracted using the ex- and cum-dividend returns). Construct excess returns by subtracting the log of the 1-month gross t-bill rate from the 1-month gross cum-dividends returns. Note: to get to gross returns you may have to add 1 to the original data series.
   From the St. Louis Fed data page (FRED; https://fred.stlouisfed.org/), get monthly data on the term and default spreads for the same sample. For the former, use the "10- Year Treasury Constant Maturity Minus Federal Funds Rate," for the latter subtract "Moodyís Seasoned Aaa Corporate Bond Minus Federal Funds Rate" from "Moodyís Seasoned Baa Corporate Bond Minus Federal Funds Rate."

```
vwretd <- vwretd %>% mutate(dividend_yield = `Value-Weighted Return-incl. dividends` -
                            `Value-Weighted Return-excl. dividends`,
                            dividend = dividend_yield * `Level of the S&P 500 Index`,
                            annualized_dividend = NA, t_bill = NA)
for (i in 12:length(vwretd$`Calendar Date`)) {
  vwretd$annualized_dividend[i] <- sum(vwretd$dividend[i:(i-11)]) / vwretd$`Level of the S&P 500 Index`
}
for (i in vwretd$`Calendar Date`) {
```

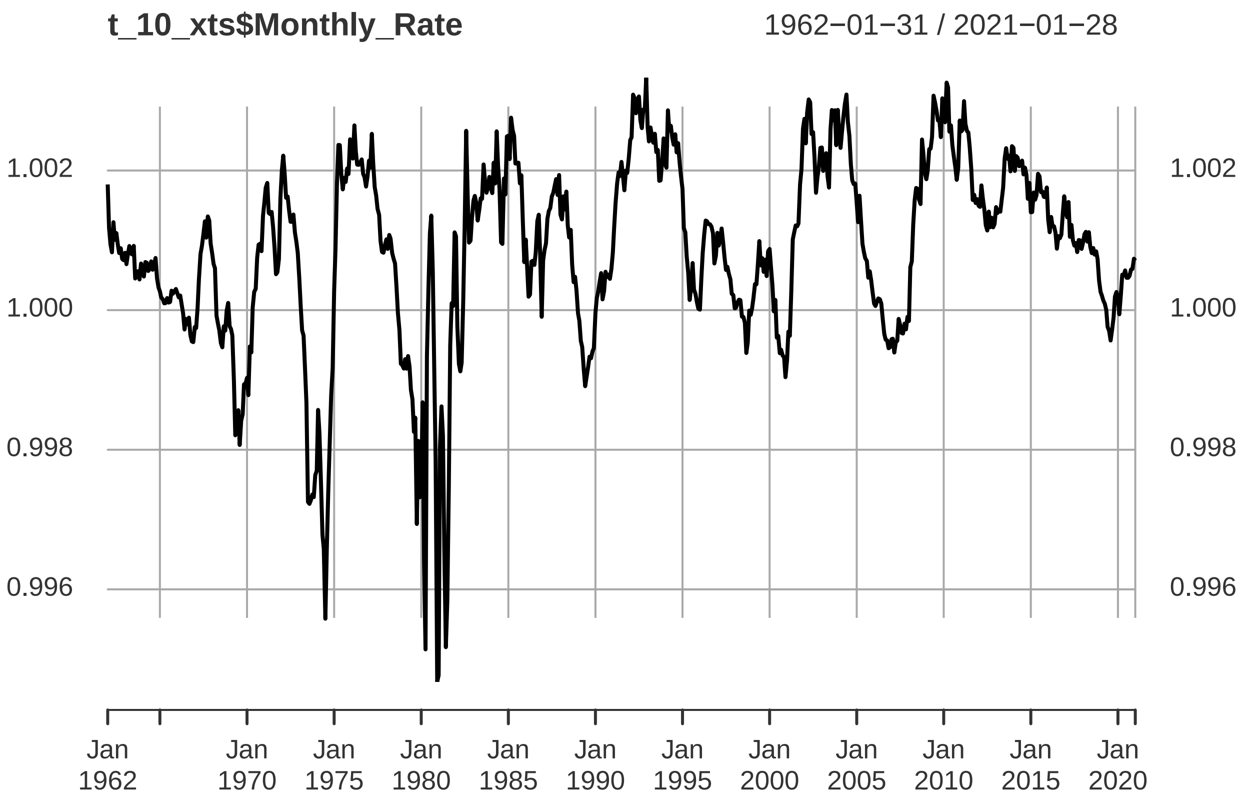
```r
  ifelse(i %in% yield$MCALDT, vwretd[`Calendar Date` = i]$t_bill <- yield[MCALDT = i]$TMYTM,
         vwretd[`Calendar Date` = i]$t_bill <- NA)
}


vwretd <- vwretd %>% mutate(t_bill = t_bill/100, gross_annualized_dividend = annualized_dividend + 1,
                            gross_t_bill = as.double(t_bill + 1), log_t_bill =
                            log(gross_t_bill), log_sp500 = log(1 + vwretd$`Value-Weighted Return-incl. d


baa <- left_join(baa,aaa, by="observation_date")
baa <- baa[complete.cases(baa),]
baa <- baa %>% mutate(default_spread = baa$DBAA - baa$AAAFF)

baa_xts <- xts(baa$default_spread, as.Date(baa$observation_date))
baa_xts <- apply.monthly((baa_xts/(100*252) + 1), prod)
colnames(baa_xts) = c("Monthly_Rate")

t_10 <- t_10[complete.cases(t_10),]
t_10 <- t_10 %>% mutate(daily_rate = t_10$T10YFF/(100*252))

t_10_xts <- xts(t_10$daily_rate, as.Date(t_10$observation_date))
t_10_xts <- apply.monthly((t_10_xts + 1), prod)
colnames(t_10_xts) = c("Monthly_Rate")
```

2. Plot your data.

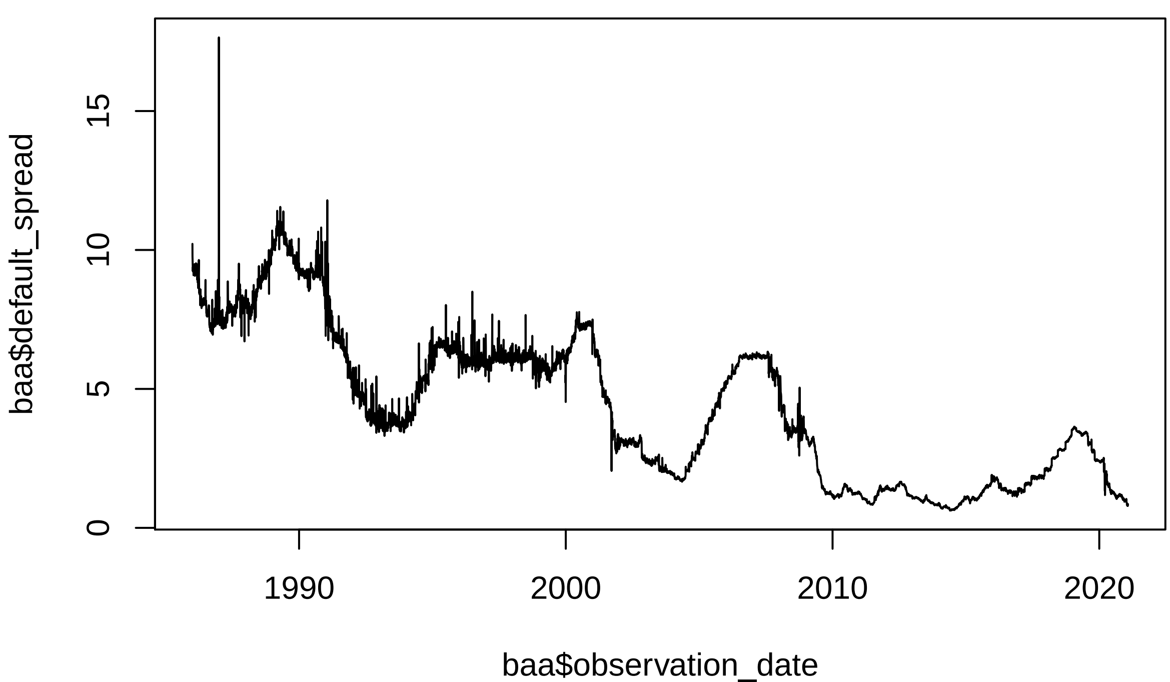
```r
vwretd <- vwretd[complete.cases(vwretd),]
plot(vwretd$`Calendar Date`, vwretd$excess_gross_dividend_yield, type ='l')
```



```
plot(t_10_xts$Monthly_Rate, type='l')
```




```
plot(baa$observation_date, baa$default_spread, type='l')
```



3. Using your three predictive variables (the lagged dividend yield, term spread, and default spread), forecast excess equity returns at the 1-month, 3-month, 12-month, 24-month, and 60-month horizons. Report your results from each of these regressions (regression coe¢ cients, standard errors, and R2s).

The underlying data is monthly, so make sure to explain your choice of standard errors.

```r
reg_df <- merge.xts(baa_xts,t_10_xts,xts(vwretd$annualized_dividend/12, as.Date(vwretd$`Calendar Date`))
```

```
## Warning in merge.xts(baa_xts, t_10_xts, xts(vwretd$annualized_dividend/12, :
## 'join' only applicable to two object merges
```

```r
colnames(reg_df) <- c("Monthly_Rate_default","Monthly_Rate_term","Monthly_rate_dividend")
reg_df$Monthly_Rate_default <- reg_df$Monthly_Rate_default - 1
reg_df$Monthly_Rate_term <- reg_df$Monthly_Rate_term - 1

reg_df <- merge.xts(reg_df, xts((vwretd$`Value-Weighted Return-incl. dividends` - vwretd$t_bill/12), as
colnames(reg_df) <- c("Monthly_Rate_default","Monthly_Rate_term","Monthly_rate_dividend", "Excess_Equity

reg_out <- lm((reg_df$Excess_Equity_Monthly) ~ (reg_df$Monthly_Rate_default) + (reg_df$Monthly_Rate_term
summary(reg_out)
```

```
##
## Call:
## lm(formula = (reg_df$Excess_Equity_Monthly) ~ (reg_df$Monthly_Rate_default) +
##     (reg_df$Monthly_Rate_term) + (reg_df$Monthly_rate_dividend))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.220918 -0.023438  0.004558  0.026149  0.125280
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.014654   0.007673   1.910   0.0569 .
## reg_df$Monthly_Rate_default  -0.626645   1.541659  -0.406   0.6846
## reg_df$Monthly_Rate_term     -1.760503   2.904558  -0.606   0.5448
## reg_df$Monthly_rate_dividend -1.612739   4.603327  -0.350   0.7263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04281 on 404 degrees of freedom
## Multiple R-squared:  0.002682,   Adjusted R-squared:  -0.004724
## F-statistic: 0.3621 on 3 and 404 DF,  p-value: 0.7804
```

```r
forecast_returns <- function(xtsdata, regout, sd){
for (i in 1:(length(xtsdata$Excess_Equity_Monthly))){
  xtsdata$forecasted_returns[i] <- regout$coefficients[1] + (regout$coefficients[2] * xtsdata$Monthly_Ra
}
  return(xtsdata)
}
reg_df_multiply <- 1 + reg_df

multiply_for_period <- function(xts_data, k){

  for (i in k:length(xts_data$Excess_Equity_Monthly)) {
    xts_data$Monthly_Rate_default[i] <- prod(reg_df_multiply$Monthly_Rate_default[i:(i-k+1)])
    xts_data$Monthly_Rate_term[i] <- prod(reg_df_multiply$Monthly_Rate_term[i:(i-k+1)])
    xts_data$Monthly_rate_dividend[i] <- prod(reg_df_multiply$Monthly_rate_dividend[i:(i-k+1)])
    xts_data$Excess_Equity_Monthly[i] <- prod(reg_df_multiply$Excess_Equity_Monthly[i:(i-k+1)])
}
  return(xts_data[k:length(xts_data$Excess_Equity_Monthly),] - 1)
```

```r
}
reg_df_3month <- multiply_for_period(1 + reg_df,3)
reg_out_3month <- lm(reg_df_3month$Excess_Equity_Monthly ~ reg_df_3month$Monthly_Rate_default +
                       reg_df_3month$Monthly_Rate_term + reg_df_3month$Monthly_rate_dividend)
summary(reg_out_3month)
```

```
##
## Call:
## lm(formula = reg_df_3month$Excess_Equity_Monthly ~ reg_df_3month$Monthly_Rate_default +
##     reg_df_3month$Monthly_Rate_term + reg_df_3month$Monthly_rate_dividend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31727 -0.03197  0.00557  0.04024  0.23024
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          0.03253    0.01344   2.420   0.0160 *
## reg_df_3month$Monthly_Rate_default  -1.89917    0.91770  -2.069   0.0391 *
## reg_df_3month$Monthly_Rate_term     -3.44454    1.74553  -1.973   0.0491 *
## reg_df_3month$Monthly_rate_dividend  3.93480    2.71191   1.451   0.1476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07405 on 402 degrees of freedom
## Multiple R-squared:  0.0117, Adjusted R-squared:  0.004325
## F-statistic: 1.586 on 3 and 402 DF,  p-value: 0.1921
```

```r
reg_df_12month <- multiply_for_period(1 + reg_df,12)
reg_out_12month <- lm(reg_df_12month$Excess_Equity_Monthly ~ reg_df_12month$Monthly_Rate_default +
                        reg_df_12month$Monthly_Rate_term + reg_df_12month$Monthly_rate_dividend)
summary(reg_out_12month)
```

```
##
## Call:
## lm(formula = reg_df_12month$Excess_Equity_Monthly ~ reg_df_12month$Monthly_Rate_default +
##     reg_df_12month$Monthly_Rate_term + reg_df_12month$Monthly_rate_dividend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53075 -0.07016  0.01203  0.09615  0.43021
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           0.10612    0.02884   3.680 0.000266 ***
## reg_df_12month$Monthly_Rate_default  -1.62045    0.49760  -3.257 0.001226 **
## reg_df_12month$Monthly_Rate_term     -2.41113    1.02137  -2.361 0.018729 *
## reg_df_12month$Monthly_rate_dividend  3.83970    1.47643   2.601 0.009655 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1529 on 393 degrees of freedom
## Multiple R-squared:  0.02689,    Adjusted R-squared:  0.01946
## F-statistic:  3.62 on 3 and 393 DF,  p-value: 0.0133
```

```r
reg_df_24month <- multiply_for_period(1 + reg_df,24)
reg_out_24month <- lm(reg_df_24month$Excess_Equity_Monthly ~ reg_df_24month$Monthly_Rate_default +
                        reg_df_24month$Monthly_Rate_term + reg_df_24month$Monthly_rate_dividend)
summary(reg_out_24month)
```

```
##
## Call:
## lm(formula = reg_df_24month$Excess_Equity_Monthly ~ reg_df_24month$Monthly_Rate_default +
##     reg_df_24month$Monthly_Rate_term + reg_df_24month$Monthly_rate_dividend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.66601 -0.10373 -0.00612  0.12672  0.67032
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           0.23947    0.04692   5.104 5.26e-07 ***
## reg_df_24month$Monthly_Rate_default  -1.56659    0.39514  -3.965 8.78e-05 ***
## reg_df_24month$Monthly_Rate_term     -2.24093    0.90843  -2.467  0.01407 *
## reg_df_24month$Monthly_rate_dividend  3.22255    1.18669   2.716  0.00692 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2348 on 381 degrees of freedom
## Multiple R-squared:  0.0418, Adjusted R-squared:  0.03426
## F-statistic:  5.54 on 3 and 381 DF,  p-value: 0.000989
```

```r
reg_df_60month <- multiply_for_period(1 + reg_df,60)
reg_out_60month <- lm(reg_df_60month$Excess_Equity_Monthly ~ reg_df_60month$Monthly_Rate_default +
                        reg_df_60month$Monthly_Rate_term + reg_df_60month$Monthly_rate_dividend)
summary(reg_out_60month)
```

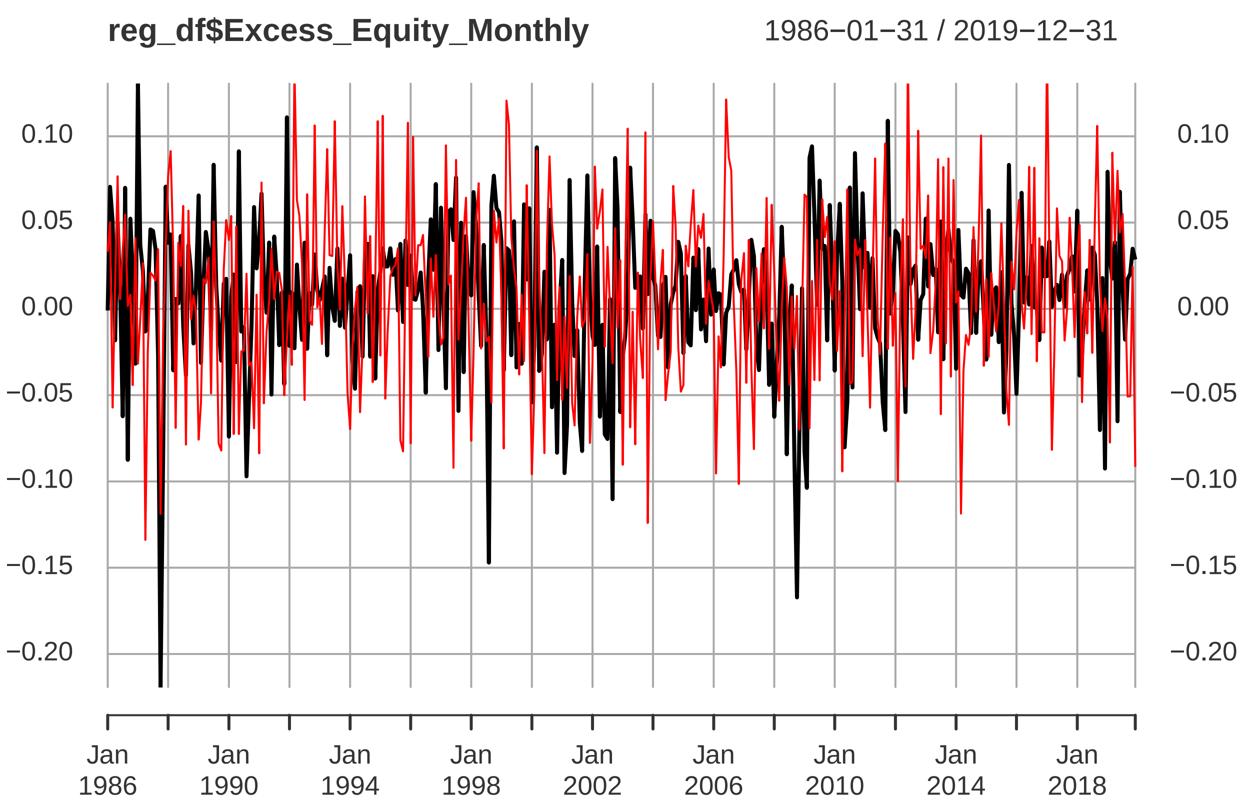
```
##
## Call:
## lm(formula = reg_df_60month$Excess_Equity_Monthly ~ reg_df_60month$Monthly_Rate_default +
##     reg_df_60month$Monthly_Rate_term + reg_df_60month$Monthly_rate_dividend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92415 -0.42612 -0.07446  0.31086  1.40287
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            0.5676     0.1318   4.308 2.16e-05 ***
## reg_df_60month$Monthly_Rate_default   -1.5770     0.3794  -4.157 4.07e-05 ***
## reg_df_60month$Monthly_Rate_term      -3.7141     1.3323  -2.788   0.0056 **
## reg_df_60month$Monthly_rate_dividend   5.2668     1.1982   4.396 1.47e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4965 on 345 degrees of freedom
## Multiple R-squared:  0.06038,    Adjusted R-squared:  0.05221
## F-statistic:  7.39 on 3 and 345 DF,  p-value: 8.219e-05
```

```r
reg_df <- forecast_returns(reg_df, reg_out, 0.05)

reg_df_3month <- forecast_returns(reg_df_3month, reg_out_3month, 0.08)

reg_df_12month <- forecast_returns(reg_df_12month, reg_out_12month, 0.1)

reg_df_24month <- forecast_returns(reg_df_24month, reg_out_24month, 0.2)

reg_df_60month <- forecast_returns(reg_df_60month, reg_out_60month, 0.5)
```

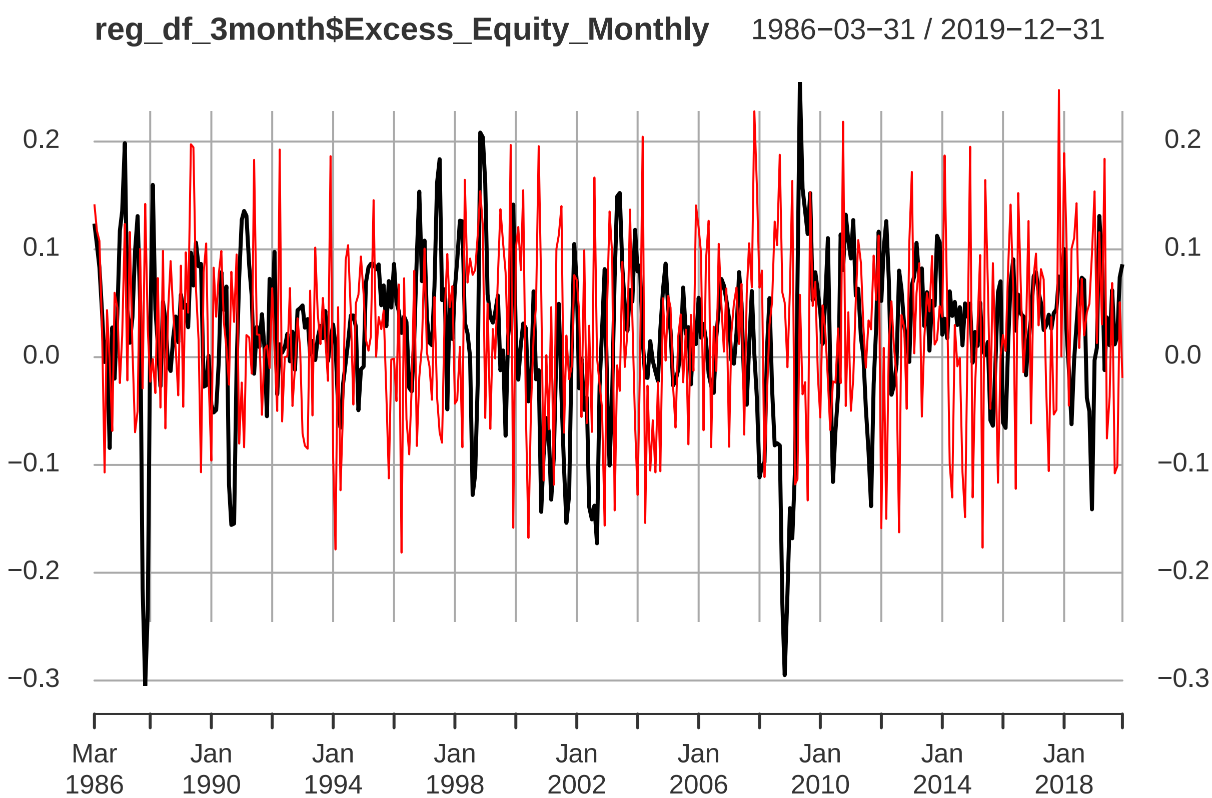
Our choice of standard errors is based on the fact that high p-values indicate a poor regression result and hence our coefficients might not be correct. So we take higher standard errors for the shock to ensure a margin of error in prediction.

4. Plot the estimated expected 12-month excess return that obtains from the forecasting regressions of the 12-month excess return regression. What type of periods are associated with high expected returns and what type of periods are associated with low expected returns. Does the patterns make sense to you? What are the economic stories you would tell to explain these patterns?

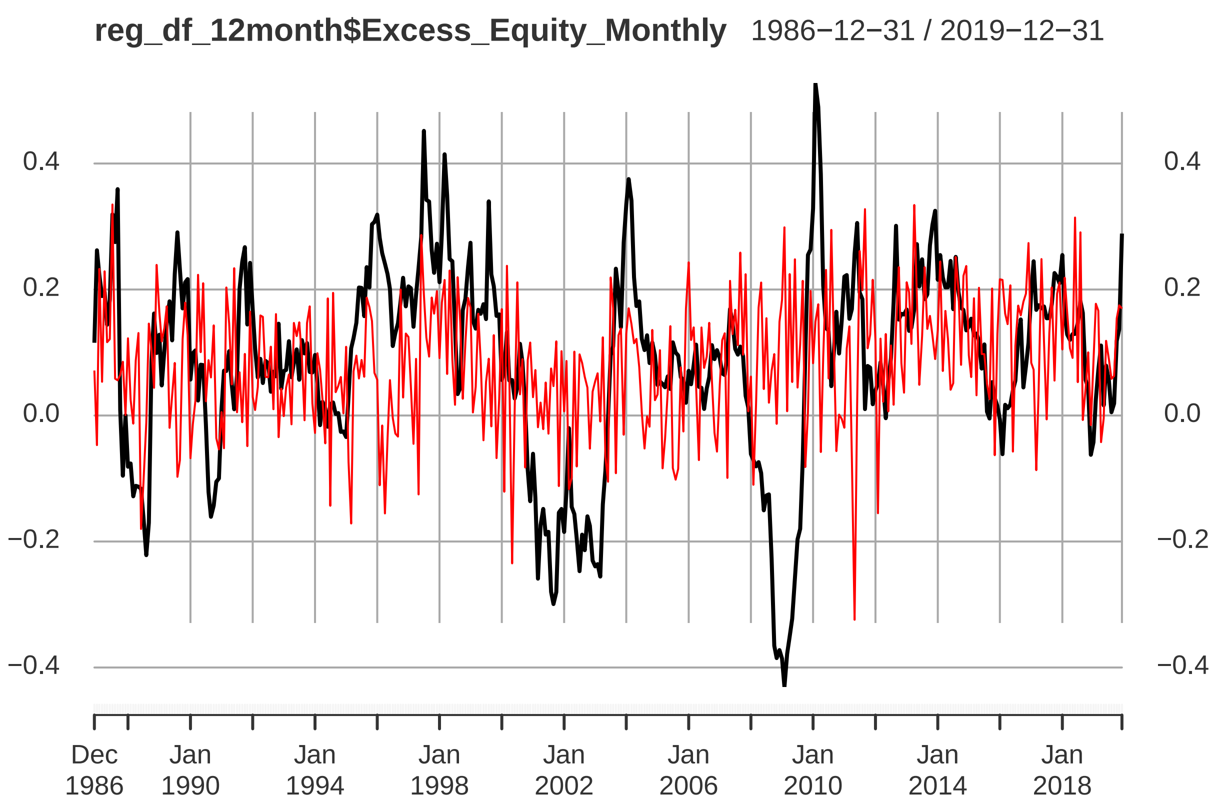
```r
dat1 <- plot(reg_df$Excess_Equity_Monthly, type = 'l')
lines(reg_df$forecasted_returns, col = 'red')
```



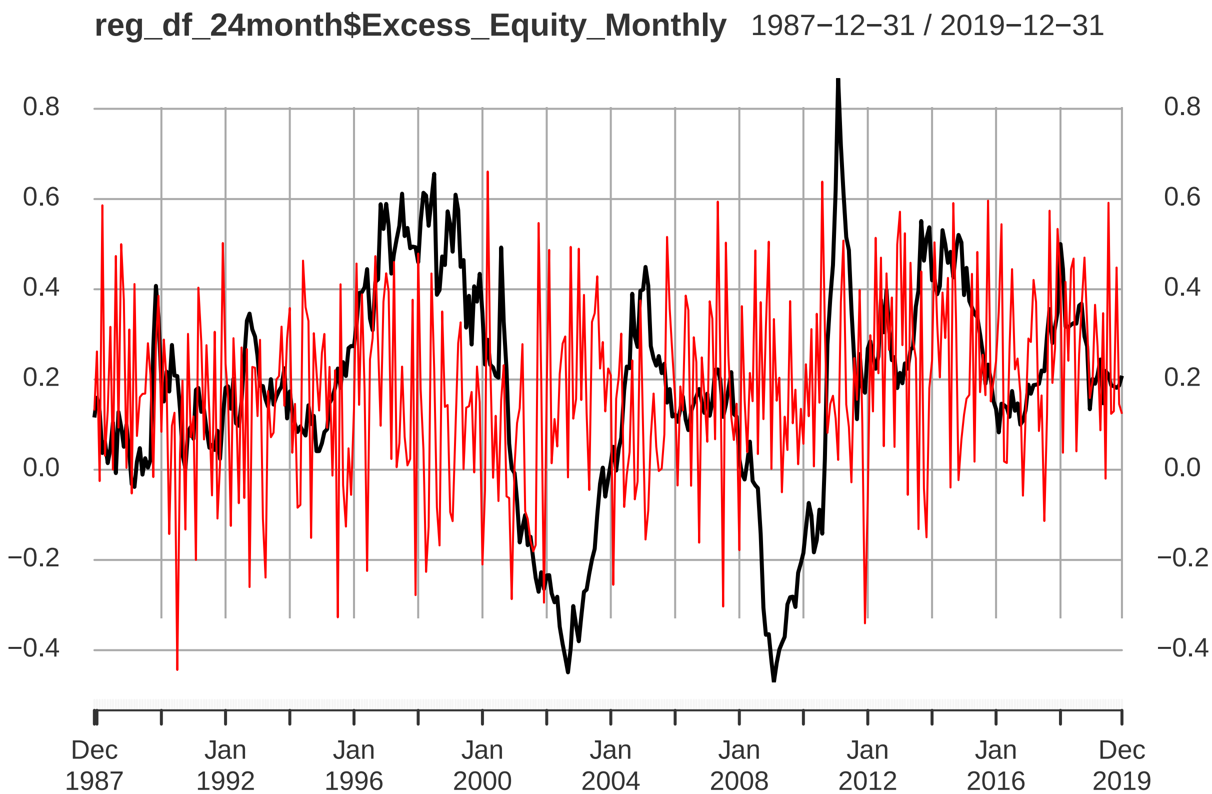

```r
dat2 <- plot(reg_df_3month$Excess_Equity_Monthly, type = 'l')
lines(reg_df_3month$forecasted_returns, col = 'red')
```



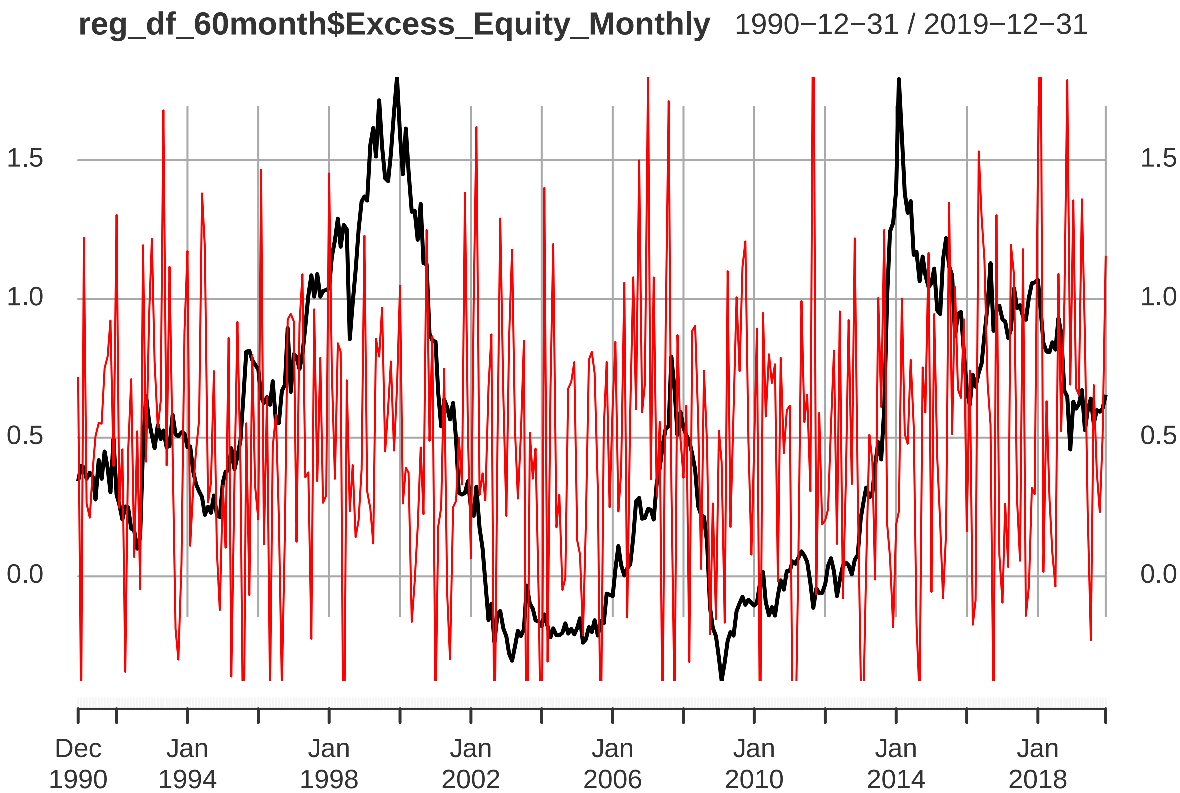

```
dat3 <- plot(reg_df_12month$Excess_Equity_Monthly, type = 'l')
lines(reg_df_12month$forecasted_returns, col = 'red')
```



```
dat4 <- plot(reg_df_24month$Excess_Equity_Monthly, type = 'l')
lines(reg_df_24month$forecasted_returns, col = 'red')
```




```
dat5 <- plot(reg_df_60month$Excess_Equity_Monthly, type = 'l')
lines(reg_df_60month$forecasted_returns, col = 'red')
```



We can observe that our regression results do not have robust coefficients. During the years 2000 (Dot-com bubble), 2008 (Sub-prime crisis) & 2014 we see huge spikes in excess returns. Expected returns are less during periods of less volatility (safe periods).

5. In the last class, we used our AR toolkit to infer expected returns at long horizons. Why do you think using, say, an AR(1)-type setting could be useful instead of simply regressing returns at di§erent horizons on lagged predictive variables as we have done in this problem?

   In our original model when we used the 3 predictive variables, we obtained very poor regression results. Using a Time-series model would be better because all factors affecting the variable would have already been take into consideration and only the shock will be the new effect we have to account for.