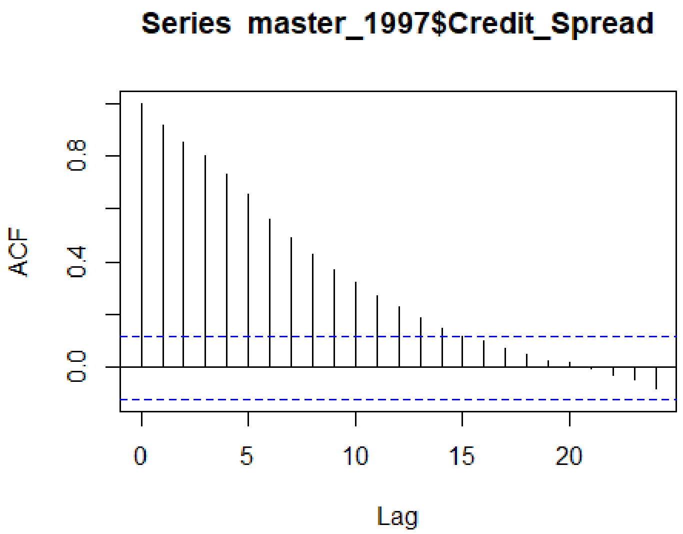
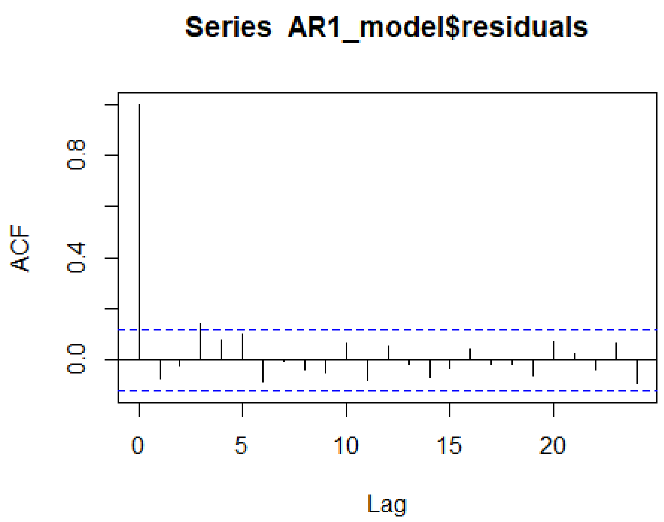
Predicting Credit Spreads

Credit spreads represent the difference in the yield of a risky security versus that of a risk-free security. In general, credit spreads are wider in times of economic uncertainty and narrower when there is less perceived risk. But credit spreads are also affected by supply and demand dynamics for bonds in addition to confidence in economic outlook. For the project, I would like to focus on the credit spread of US investment grade corporate bonds (e.g. BofAML US Corporate AAA) over US Treasuries (e.g. 10-Year Treasury Constant Maturity Rate). Then I would like to expand the study into corporate bonds of different rating categories. Using available economic, financial, and other data, I will develop a forecast of the direction and (perhaps) the magnitude of credit spread movements using machine-learning based tools and techniques. Also I will examine the various potential shortcomings of this approach, including non-transparency/"blackbox", model overfitting, and regime shift issues.

ACF of Credit spread to understand dynamics. Persistent series with high autocorrelations



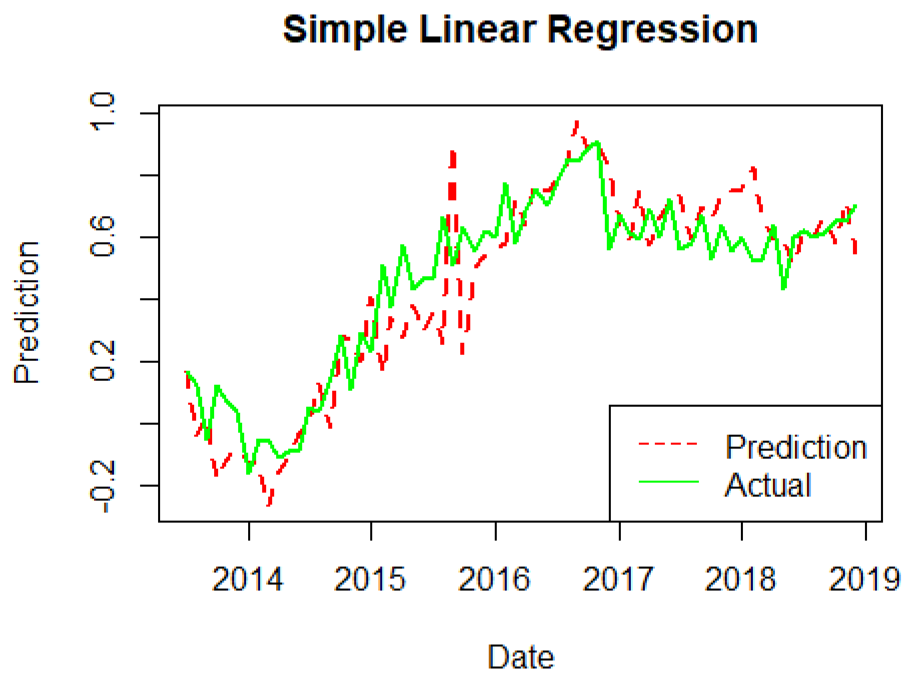
Taking lag 1 term purges the residuals sufficiently. So it follows AR(1) with high persistence



SLR

Next we divide into train(75%) and test data(25%) and run a simple linear regression

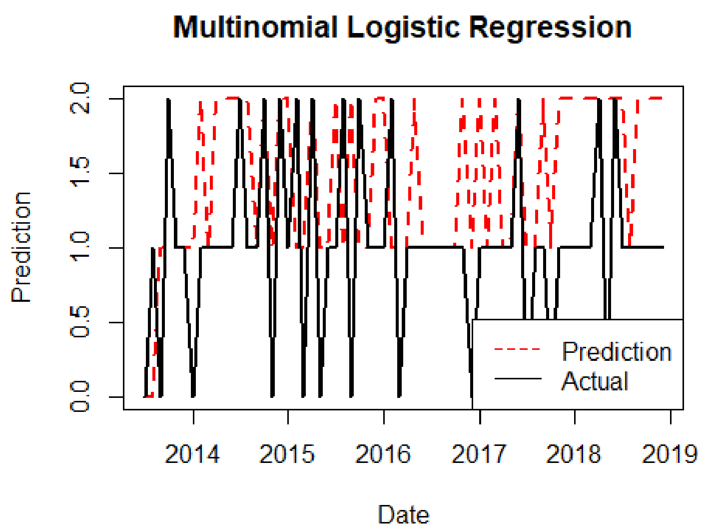
```
Call:
lm(formula = Credit_Spread ~ ., data = Reg_mat_train, na.action = na.omit)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.60638 -0.14404 -0.02293  0.09019  1.43605 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                2.58585    1.05399   2.453 0.016219 *  
T10Y2YM                    0.18212    0.20252   0.899 0.371087    
CPI_change_YOY             5.27272    2.71061   1.945 0.055095 .  
Unemployment              -0.05821    0.06060  -0.961 0.339529    
VIX                       -0.02189    0.01286  -1.703 0.092287 .  
VIX_chg                    0.49345    0.27659   1.784 0.078022 .  
Advance_retail_sales      -0.04582    0.03279  -1.397 0.166059    
St_louis_fed_stress_index  0.47054    0.27822   1.691 0.094494 .  
SP500_return_no_dividend   1.53678    1.26204   1.218 0.226749    
ICE_BAML_1yto3y           -0.27966    0.13046  -2.144 0.034950 *  
TEDRATE                   -0.23966    0.20967  -1.143 0.256267    
Corporate_slope           -0.69858    0.33439  -2.089 0.039727 *  
put_skew_ema               0.72538    5.25448   0.138 0.890532    
call_skew_ema             -0.69393    5.24136  -0.132 0.894988    
Credit_Spread_BBB         -0.56339    0.30450  -1.850 0.067795 .  
Credit_Spread_CCC          0.01450    0.03319   0.437 0.663345    
Credit_Spread_lag          1.23959    0.33170   3.737 0.000339 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2801 on 84 degrees of freedom
  (96 observations deleted due to missingness)
Multiple R-squared:  0.896,	Adjusted R-squared:  0.8762 
F-statistic: 45.23 on 16 and 84 DF,  p-value: < 2.2e-16
```



MSPE = 18.38

<u>Multinomial Logistic Regression</u>

Next we run a multinomial logistic regression where an up move > 0.5 std deviation is 2, down move < -0.5 std deviation is 0 and everything else is 1

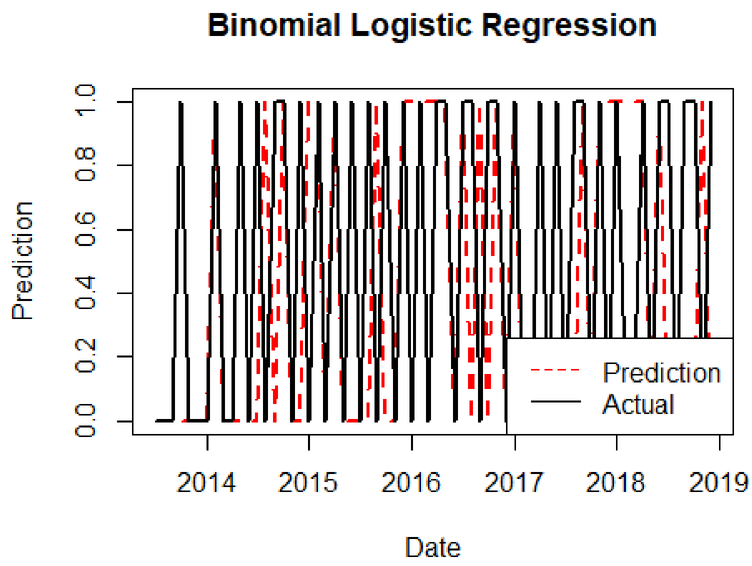


Prediction accuracy = 51.5%

This can be compared to 33% with random predictions

<u>Binomial Logistic Regression</u>

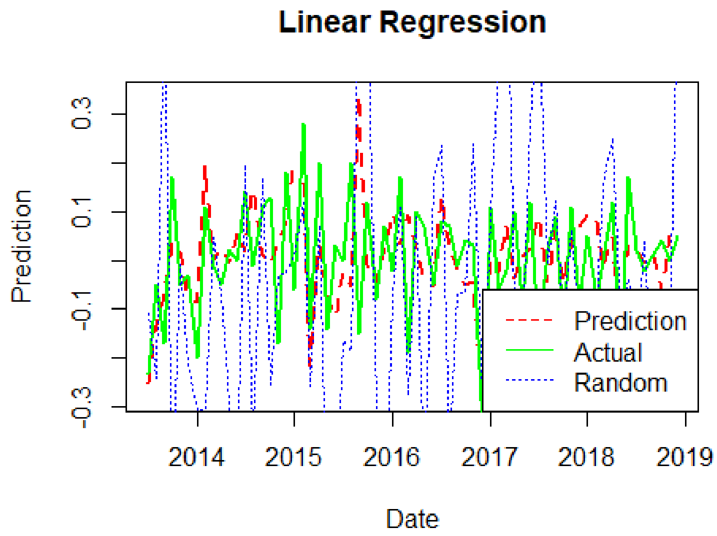
Upmove is 1 and downmove is 0



Prediction accuracy = 59% compared to 50% with random

SLR on difference series

```
Call:
lm(formula = Cred_spread_chg ~ ., data = Reg_mat_diff_train,
    na.action = na.omit)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.10424 -0.10239  0.00307  0.09677  1.66805 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.01415    0.01749  -0.809 0.419573    
TSY_slope_chg   -0.38381    0.12723  -3.017 0.002915 ** 
VIX_chg          0.29742    0.13086   2.273 0.024192 *  
Unemp_diff       0.02546    0.04833   0.527 0.598901    
TED_chg          0.18816    0.09183   2.049 0.041872 *  
Corp_slope_chg  -0.37883    0.08611  -4.400 1.83e-05 ***
BBB_chg         -0.07684    0.08264  -0.930 0.353727    
CCC_chg          0.01175    0.01441   0.816 0.415819    
Fed_stress_chg  -0.02118    0.13574  -0.156 0.876177    
ICE_BAML_3y_chg -0.25411    0.07342  -3.461 0.000668 ***
SP500_ret        1.22854    0.56729   2.166 0.031620 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2337 on 185 degrees of freedom
Multiple R-squared:  0.2343,	Adjusted R-squared:  0.1929 
F-statistic: 5.661 on 10 and 185 DF,  p-value: 2.228e-07
```



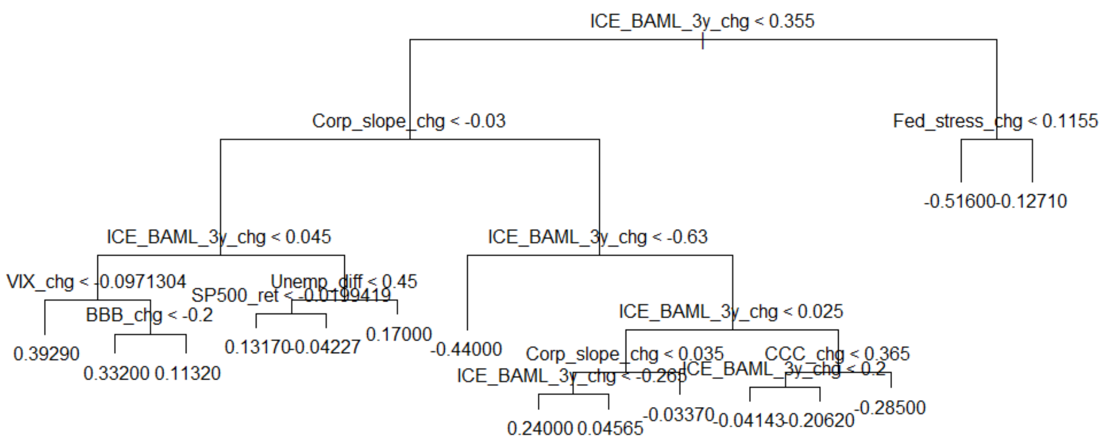
MSPE random = 6.536

MSPE prediction = 1.032

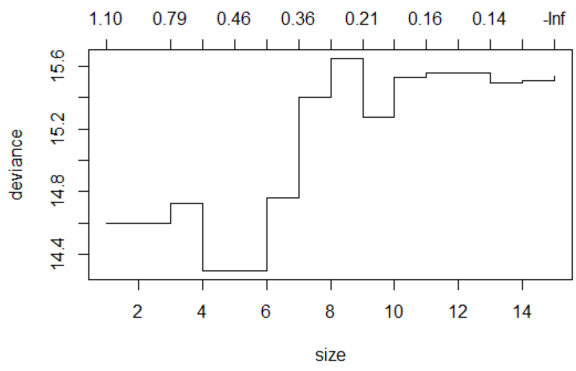
<u>LASSO</u>

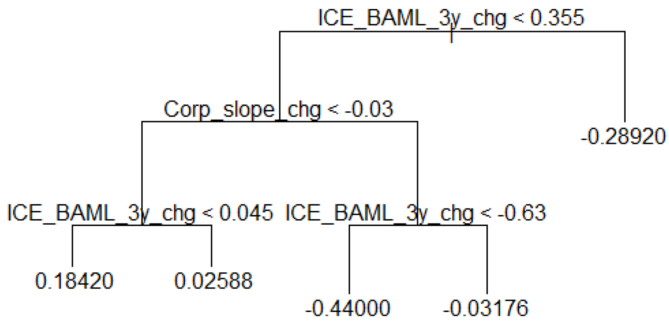
<u>CART</u>

Generate a tree based on difference series

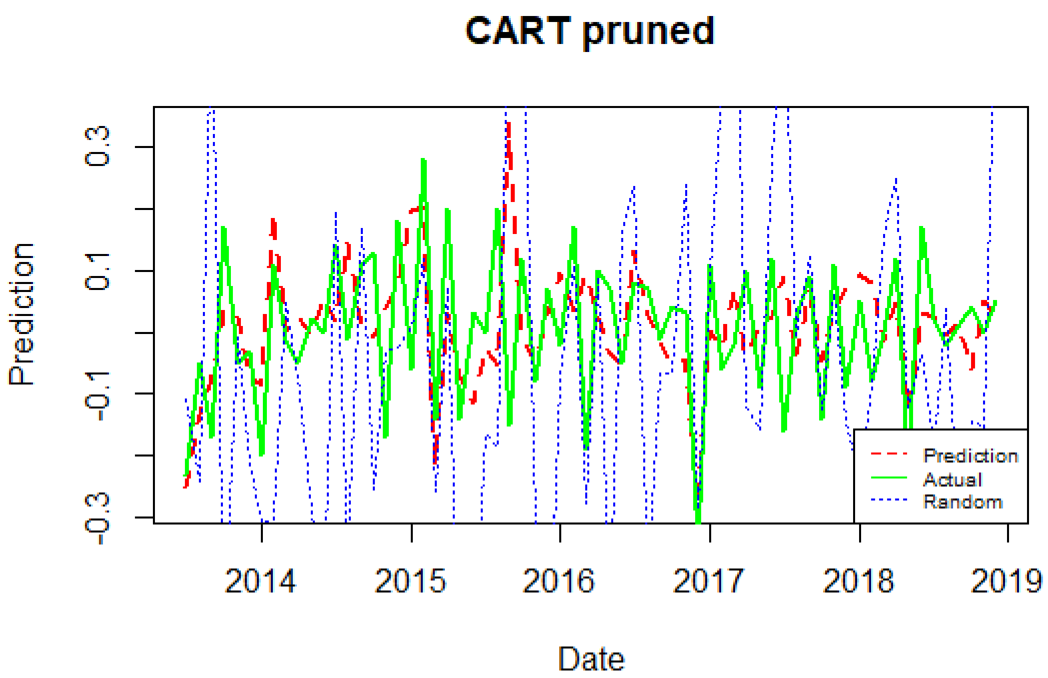


Using cross validation on tree to prune it


1.10 0.79 0.46 0.36 0.21 0.16 0.14 -Inf
deviance
15.6
15.2
14.8
14.4
2 4 6 8 10 12 14
size


ICE_BAML_3y_chg < 0.355
Corp_slope_chg < -0.03
-0.28920
ICE_BAML_3y_chg < 0.045
ICE_BAML_3y_chg < -0.63
0.18420
0.02588
-0.44000
-0.03176

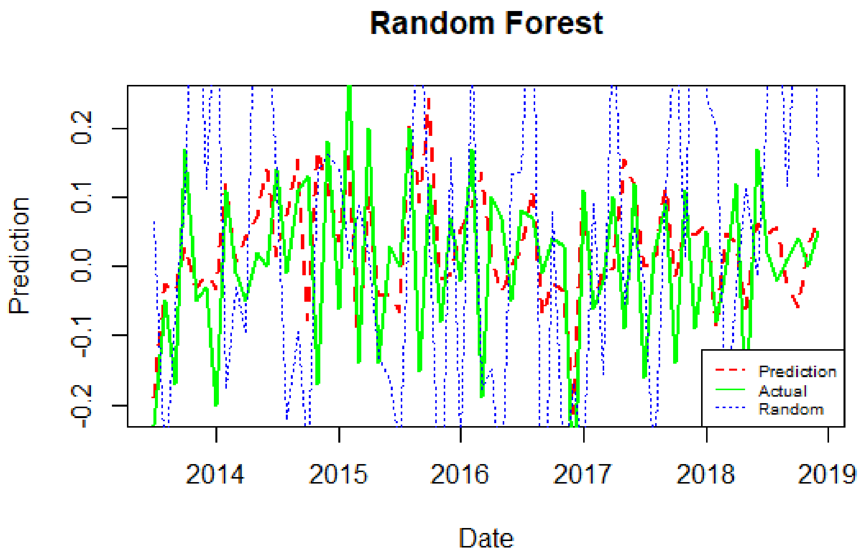
Pruned Tree


CART pruned
Prediction
0.3
0.1
-0.1
-0.3
2014 2015 2016 2017 2018 2019
Date
Prediction
Actual
Random

MSPE random = 5.355

MSPE prediction = 1.033

Random Forest

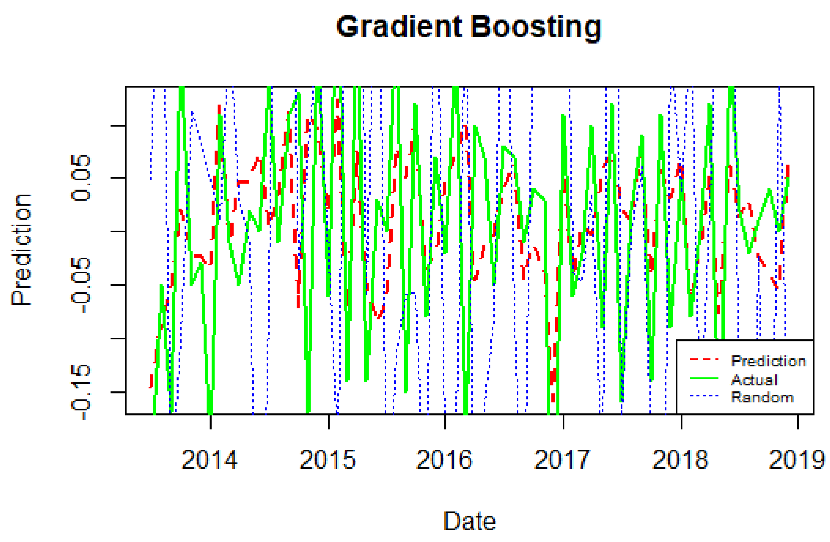


MSPE random = 6.086

MSPE prediction = 0.799

Gradient Boosting

	var	rel.inf
Corp_slope_chg	Corp_slope_chg	29.312012
ICE_BAML_3y_chg	ICE_BAML_3y_chg	25.723656
TSY_slope_chg	TSY_slope_chg	16.955657
BBB_chg	BBB_chg	7.767710
CCC_chg	CCC_chg	4.704455
VIX_chg	VIX_chg	3.845886
TED_chg	TED_chg	3.531239
SP500_ret	SP500_ret	3.221823
Unemp_diff	Unemp_diff	2.856536
Fed_stress_chg	Fed_stress_chg	2.081025



MSPE random = 6.007

MSPE prediction = 0.694

Gradient Boosting has improved MSPE, relative to other methods.