Six Stock Market Sectors Signaling Storm Clouds

G. K.

August 10, 2021

The Problem

Stock market analyst and MarketWatch contributor Mark Hulbert wrote an article published August 7, 2021 titled *Here's Another Sign the Bull Market is Near a Peak,* identifying **6 stock market sectors** to watch when trying to identify the top of a stock market bull trend before a downturn (into a bear market). His article can be found here:

https://www.marketwatch.com/story/heres-another-sign-the-bull-market-is-near-a-peak-and-this-one-bears-watching-11628233932?mod=home-page

Using historical data, Hulbert correlates the **3 months** leading up to a stock market top were configured as such: 3 stock market "sectors that typically perform the **worst** over the last 3 months of bull markets" (Energy, Financials and Utilities); and 3 stock market "sectors that typically perform the **best** over the last 3 months of bull markets" (Consumer Staples, Consumer Discretionary and Health Care) (Hubert, 2021).

(In no way is this analysis attempting to predict stock market movements, rather is merely seeking to support or interrupt the viability of the indicator that Hubert presents.)

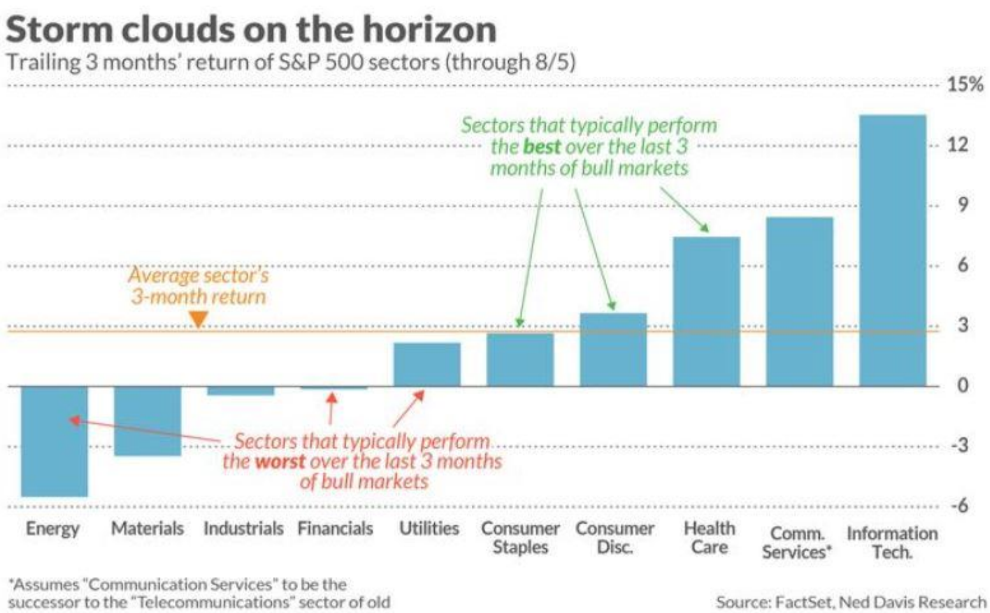


Figure 1. Stock market strength by sector in August 2021 (Huber, 2021).

The Data

First, we need to define and identify what a bull market is. It is a general upward trend in overall stock market prices beginning after a stock market crash (or the end of a bear market), and usually ending at the beginning of a stock market crash (or start of a bear market). Examples include June 1932 to February 1937 (following the stock market crash of 1929), and October 1990 to March 2000 (followed by the bursting of the dot-com bubble) (Frank, 2020). A full list can be found here:

https://www.cnbc.com/2020/08/18/heres-a-list-of-stock-bull-markets-through-time-and-how-this-new-one-stacks-up.html

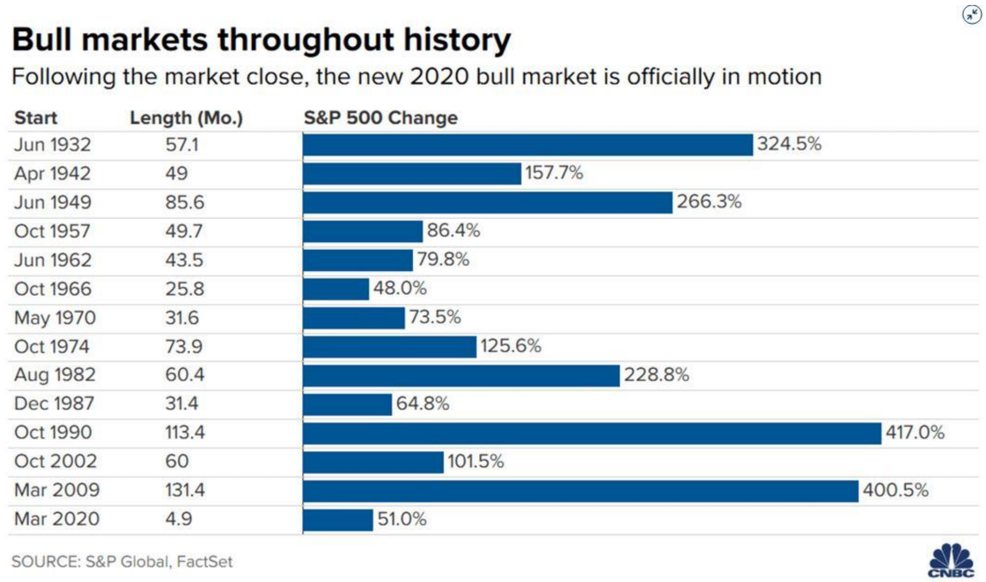


Figure 2. Bull market history from 1932 to 2020 (Frank, 2020).

Periods of Interest

The data we will be using ranges from December 1998 to April 2020. This allows us to study the tail end of the dot-com bubble ending March 2000; the bull market from October 2002 to October 2007, followed by housing market crash of 2008; and the bull market from March 2009 to February 2020, which was ended by the Covid-19 crash. According to Mark Hubert, the final 3 months of a bull market are the most important and will be highly scrutinized.

This data is available on Kaggle (Onyshchak, 2020):

https://www.kaggle.com/jacksoncrow/stock-market-dataset

About the Data

An ETF (exchange traded fund) is a security that tracks an index (group of stocks), sector, commodity, or asset sharing commonality (Chen, 2021).

XLE – The first of our six sector ETF data sets comes from XLE, the ETF that represents the overall **Energy** market, mostly comprised of oil and natural gas related companies. The remaining five sector data sets are a as follows:

XLF – Financials

XLU – Utilities

XLP – Consumer Staples (Food, beverages, etc.)

XLY – Consumer Discretionary (Nonessentials such as retail, leisure, entertainment, automotive, etc.)

XLV – Health Care

One final (seventh) data set is **SPY** (S&P 500). Although not a sector, it is important because this index best represents the overall broad stock market covering 500 of the largest U.S. companies.

Data Set Columns

Each data set has seven columns, the Date, five prices (Open, High, Low, Close, Adjusted Close) and Volume. Although an investor may find the price fluctuations throughout the day useful, we do not; we only need a single price value on any given day. Therefore, the only two fields we will concern ourselves with is **Date** and **Close Price**.

Figure 3. Header of data set XLE (Onyshchak, 2020).

Cleaning the Data

First, let's read each of the CSV files.

```
xle <- read.csv("C:\\...\\week8\\assignment\\data\\XLE.csv")
xlf <- read.csv("C:\\ ...\\week8\\assignment\\data\\XLF.csv")
xlu <- read.csv("C:\\ ...\\week8\\assignment\\data\\XLU.csv")
xlp <- read.csv("C:\\ ...\\week8\\assignment\\data\\XLP.csv")
xly <- read.csv("C:\\ ...\\week8\\assignment\\data\\XLY.csv")
xlv <- read.csv("C:\\ ...\\week8\\assignment\\data\\XLv.csv")
spy <- read.csv("C:\\ ...\\week8\\assignment\\data\\SPY.csv")
```

Data	
spy	5353 obs. of 7 variables
xle	5353 obs. of 7 variables
xlf	5353 obs. of 7 variables
xlp	5353 obs. of 7 variables
xlu	5353 obs. of 7 variables
xlv	5353 obs. of 7 variables
xly	5353 obs. of 7 variables

Now let's create the dataframe structure to hold our finalized data.

```
# Empty df
mydf <- data.frame(matrix(ncol = 9, nrow = 5353))
```

```
# Naming columns
colnames(mydf) <- c('Date', 'XLEclose', 'XLFclose', 'XLUclose',
'XLPclose', 'XLYclose', 'XLVclose', 'SPYclose', 'Final3Months')
```

Next, let's copy over the date column from one of our data sets.

```
mydf$Date <- xle$Date
```

Remember, we only want the Closing Price from each of the ETF data sets. We will be taking that single (Close) column from each and appending it to our dataframe.

```
mydf$XLEclose <- xle$Close
mydf$XLFclose <- xlf$Close
mydf$XLUclose <- xlu$Close
mydf$XLPclose <- xlp$Close
mydf$XLYclose <- xly$Close
mydf$XLVclose <- xlv$Close
mydf$SPYclose <- spy$Close
```

Let's look to make sure we don't have any NULL values in these columns. So far so good.

```
sapply(mydf, function(x) sum(is.na(x)))
```

```
> sapply(mydf, function(x) sum(is.na(x)))
        Date     XLEclose     XLFclose     XLUclose     XLPclose     XLYclose
           0            0            0            0            0            0
    XLVclose     SPYclose Final3Months
           0            0         5353
```

The Final 3 Months – Binary Data

Remember from earlier, Mark Hubert postulates that the **3 months** prior to a bull market end are the dates to look at to indicate that the stock market is about to start a bear market down tread. Therefore, we will tag our data in the column **'Final3months'** with either **a 1 or a 0**. 1 flagging that the date was within 3 months (or 90 days) prior to a stock market crash, or 0 for dates outside the 3-month window.

First, we will fill the column with 0's.

```
mydf$Final3Months[is.na(mydf$Final3Months)] <- 0
```

Next, we must identify the 90 days before each stock market crash, along with their indexes in our dataframe. (Note: I used guess and check to find the index ranges.)

December 23, 1999 – March 24, 2000, **[254 : 317]**
July 12, 2007 – October 11, 2007, **[2150 : 2214]**
November 20, 2019 – February 19, 2020, **[5263 : 5323]**

```
mydf$Final3Months[254:317] <- 1
mydf$Final3Months[2150:2214] <- 1
mydf$Final3Months[5263:5323] <- 1
```

A quick check to count how many 1's we have in our Final3Months column. 190 looks about right.

```
table(mydf$Final3Months)
```

```
   0    1
5163  190
```

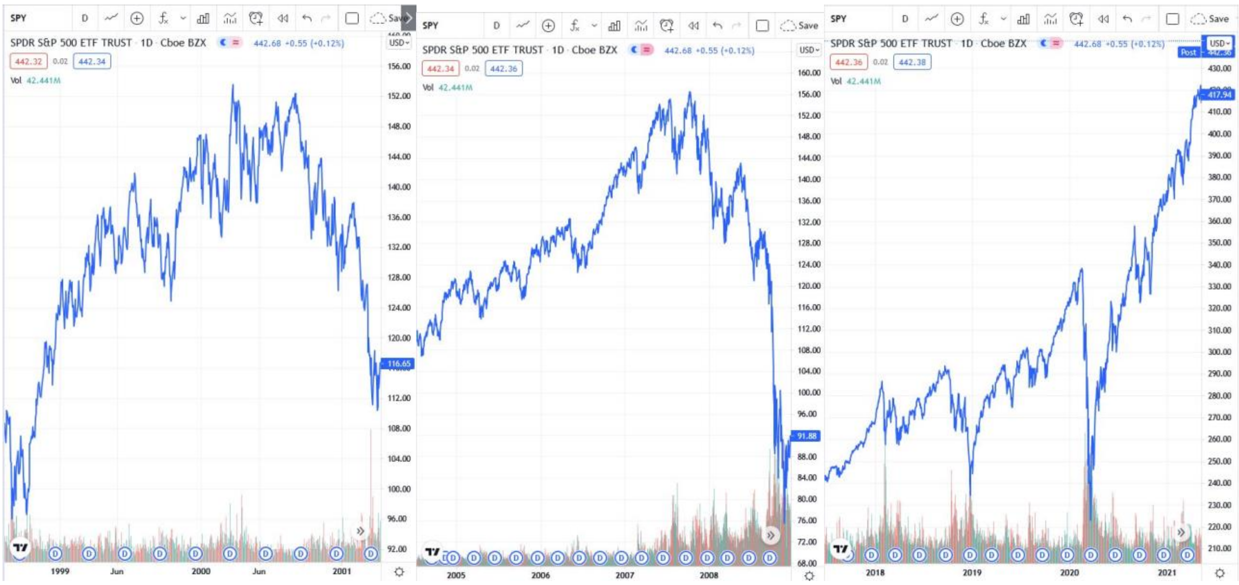


Figure 4. Charts of S&P500 during 3-months prior to market crashes in 2000, 2008, 2020.

Let's look at our final dataframe!

```
> str(mydf)
'data.frame':	5353 obs. of  9 variables:
 $ Date       : chr  "1998-12-22" "1998-12-23" "1998-12-24" "1998-12-28" ...
 $ XLEclose   : num  23.3 23.8 23.6 23.5 23.7 ...
 $ XLFclose   : num  18.9 19.2 19.3 19.1 19.3 ...
 $ XLUclose   : num  29.8 29.7 30.2 30.1 30.6 ...
 $ XLPclose   : num  26.5 27.1 27.1 26.9 27.5 ...
 $ XLYclose   : num  25.5 25.6 26 25.8 26.4 ...
 $ XLVclose   : num  25 25.6 25.8 25.4 25.9 ...
 $ SPYclose   : num  121 123 123 122 124 ...
 $ Final3Months: num  0 0 0 0 0 0 0 0 0 0 ...
> head(mydf, 5)
        Date XLEclose XLFclose XLUclose XLPclose XLYclose XLVclose SPYclose Final3Months
1 1998-12-22 23.26562 18.93785 29.82812 26.50000 25.46875 25.03125 120.6875            0
2 1998-12-23 23.75000 19.21710 29.70312 27.14062 25.57812 25.59375 123.2188            0
3 1998-12-24 23.62500 19.34403 30.25000 27.09375 26.04688 25.75000 122.6875            0
4 1998-12-28 23.50000 19.09017 30.09375 26.93750 25.81250 25.37500 122.3750            0
5 1998-12-29 23.73438 19.29326 30.59375 27.53125 26.37500 25.93750 124.3125            0
> tail(mydf, 5)
           Date XLEclose XLFclose XLUclose XLPclose XLYclose XLVclose SPYclose Final3Months
5349 2020-03-26    30.39    21.67    55.38    53.76   101.45    86.85   261.20            0
5350 2020-03-27    28.33    21.01    55.68    53.51    98.05    84.99   253.42            0
5351 2020-03-30    28.62    21.41    57.74    55.60   100.10    88.97   261.65            0
5352 2020-03-31    29.06    20.82    55.41    54.47    98.08    88.58   257.75            0
5353 2020-04-01    27.62    19.55    52.08    53.55    93.70    85.21   246.15            0
```

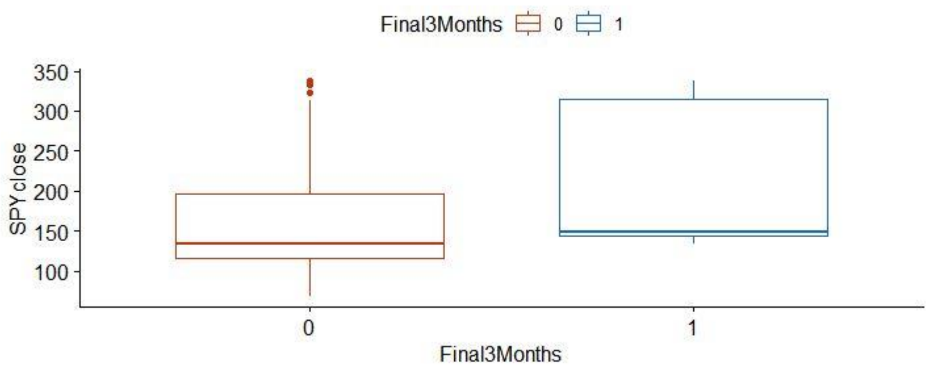
```
# To save a bit of memory, we no longer need our raw data from CSVs.
rm(xle, xlf, xlu, xlp, xly, xlv, spy)
```

```
# We need to update the data type for 'Final3Months' to factor
mydf['Final3Months'] <- lapply(mydf['Final3Months'] , factor)
```

Unimportant Visualization

Just for fun, maybe to help with intuition of the stock market, let's create a boxplot of the S&P 500 (SPY ETF) against our binary dependent variable, Final 3 Months, before a stock market crash. As expected, the range of overall stock market price during the final 3 months before a crash is higher than any other time. Interestingly, the median price between the two timeframes is relatively close.

```
library('ggpubr')
ggboxplot(mydf, x = 'Final3Months', y = 'SPYclose', color =
      'Final3Months', palette = c('#Bf360C', '#1565C0'))
```



Logistic Regression Model

Hypothesis

H0: There is **no** association among any of the 6 stock market sectors' (Energy, Financials, Utilities, Consumer Staples, Consumer Discretionary, Health Care) closing price and stock market peaks.

H1: The is an association among at least one of the 6 stock market sectors' (Energy, Financials, Utilities, Consumer Staples, Consumer Discretionary, Health Care) closing price and stock market peaks.

Splitting the data set – Train & Test

After splitting the data below, we end up with a ratio of 3611/137 in our training data set.

```
install.packages('caret')
library('caret')

set.seed(99)  # For repeatability
trainDataIndex <- createDataPartition(mydf$Final3Months, p=0.7, list =
      FALSE)
trainData <- mydf[trainDataIndex, ]  # 70%
testData <- mydf[-trainDataIndex, ]  # 30%
table(trainData$Final3Months)
```

```
   0    1
3615  133
```

Up Sampling Training Data

To deal with our highly imbalanced training data set, we will perform up sampling of our minority class from 133 to match the majority group of 3615.

```
'%ni%' <- Negate('%in%')  # define 'not in' function
set.seed(99)
up_train <- upSample(x = trainData[, colnames(trainData) %ni%
      'Final3Months'], y = trainData$Final3Months)
names(up_train)[9] <- "Final3Months"  # upSample() changed our column
nametable(up_train$Final3Months)
```

```
   0    1
3615 3615
```

Build Logistic Model

```
logitModel <- glm(Final3Months ~ XLEclose + XLFclose + XLUclose +
      XLPclose + XLYclose + XLVclose,
      family=binomial, data=up_train, maxit = 100)
summary(logitModel)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.9880  -0.2183   0.0282   0.6511   2.0423

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.124377   0.405082  -5.244 1.57e-07 ***
XLEclose    -0.126076   0.005543 -22.746  < 2e-16 ***
XLFclose    -0.028230   0.015674  -1.801   0.0717 .
XLUclose     0.884619   0.029335  30.156  < 2e-16 ***
XLPclose    -1.424396   0.045081 -31.596  < 2e-16 ***
XLYclose    -0.220330   0.019258 -11.441  < 2e-16 ***
XLVclose     0.717521   0.030406  23.598  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 10022.9  on 7229  degrees of freedom
Residual deviance:  4578.6  on 7223  degrees of freedom
AIC: 4592.6
Number of Fisher Scoring iterations: 7
```

Interpreting Coefficients

Looking straightaway at the p-values, we can see that with a 95% confidence level, the Financial sector (XLF) is not statistically significant at determining a stock market peak with a p-value of 0.07.

Interestingly, the remaining five sectors all have extremely small (much less than 0.05) p-values. Of the five significant factors, Energy (XLE), Consumer Staples (XLP), and Consumer Discretionary (XLY) had negative log odds of predicting a stock market peak.

Remember, our original motivation for this analysis is to compare our finding with Mark Hulbert, who postulates that during the final 3 months of a stock market bull run, a decreased strength of Energy (XLE), Financials (XLF), Utilities (XLU), and an increased strength of Consumer Staples (XLP), Consumer Discretionary (XLY), and Health Care (XLV) indicates a stock market peak.

Of the five significantly correlated sectors, Energy (XLE) has a negative coefficient estimate at -0.13, and Health Care (XLV) has a positive coefficient estimate at 0.72, which are in line with Mark Hulbert's claim. However, Utilities (XLU) had a positive coefficient estimate at 0.88, while Consumer Staples and Consumer Discretionary had negative coefficient estimates at -1.42 and -0.22, respectively.

Conclusion – Rejecting the Null Hypothesis

Although the direction of odds correlating to an increased probability of a stock market crash do not line up perfectly with Mark Hulbert's theory, all is not lost. Our alternative hypothesis merely asks whether there is an association among any of the six economic market sectors to a stock market downturn.

With p-values less than 0.05 for Energy (XLE), Utilities (XLU), Consumer Staples (XLP), Consumer Discretionary (XLY), andHealth Care (XLV), we can reject the null hypothesis. There exists an association with a least one of the six market sectors with identifying the final 3 months just before a stock market peak.

Goodness of Fit - Deviance

Hypothesis

H0: Our logistic regression model fits.

H1: Our logistic regression model does not fit.

Chi-square test

```
anova(logitModel, test="Chisq")
```

```
Analysis of Deviance Table
Model: binomial, link: logit
Response: Final3Months
Terms added sequentially (first to last)
          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                       7229    10022.9
XLEclose  1    61.03       7228     9961.9 5.627e-15 ***
XLFclose  1  1939.44       7227     8022.4 < 2.2e-16 ***
XLUclose  1    12.83       7226     8009.6 0.0003416 ***
XLPclose  1  2089.61       7225     5920.0 < 2.2e-16 ***
```

```
XLYclose  1   504.46      7224     5415.5 < 2.2e-16 ***
XLVclose  1   836.99      7223     4578.6 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

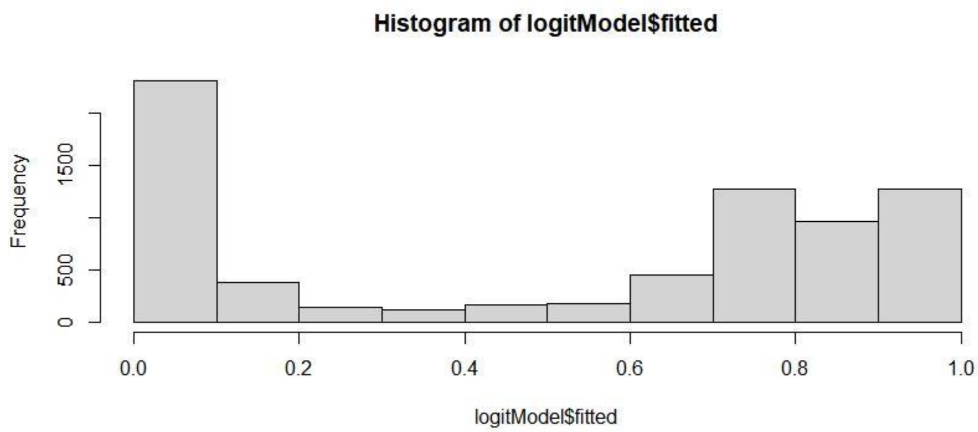
Conclusion

With each of the six market sectors (XLE, XLF, XLU, XLP, XLY, XLV) having a p-value of much less than 0.05, we can reject the null hypothesis that our model is an exact fit. (Note: This test is not incredibly useful, as accuracy of a model to a degree less than exact can still be useful) (McGill University, n.d.).

Histogram of fit

With a quick visualization of fit, we can see the further we go below or above 0.5 (in the range 0 to 1), the more frequent our fit becomes, supporting a decent fit of our model.

```
hist(logitModel$fitted)
```



Predictions

Test Data

Remember we saved 30% of our data as test data, which we will now use to evaluate the performance of our logistic regression model. The result is 82.6% (0.826) accurate in predicting the final 3 months before a stock market downturn, not bad.

```
pred <- predict(logitModel, newdata = testData, type = "response")
# pred greater than 0.5, it is 3 months before a stock peak
y_pred_num <- ifelse(pred > 0.5, 1, 0)
y_pred <- factor(y_pred_num, levels=c(0, 1))
y_act <- testData$Final3Months
# calculate accuracy
mean(y_pred == y_act)
```

```
[1] 0.8261682
```

Current Market Data

Let's see what our model thinks of the stock market today; are we in a market top like some are saying (Stankiewics, 2021)? The following closed stock market prices were recorded for August 10, 2021. With a result of 97.37% (0.9737) probability that we are within the last 3 months of a bull market, based on this model, I would be exiting the stock market.

XLE (49.79), XLF (38.39), XLU (67.49), XLP (71.68), XLY (181.35), XLV (133.03), SPY (442.40)

```
todaydf <- data.frame(Date = c(2021-08-10), XLEclose = c(49.79),
                      XLFclose = c(38.39), XLUclose = c(67.49),
                      XLPclose = c(71.68), XLYclose = c(181.35),
                      XLVclose =(133.03), SPYclose = c(442.40))
predict(logitModel, newdata = todaydf, type = "response")
```

```
        1
0.9737061
```

Control Data – September 21, 2020

As a control, let's look at a date that was not in the last 3 months before a stock market peak. As of today, August 10, 2021, the bull market is still on, which started on March 24, 2020. Let's choose September 21, 2020 as our control date. This date is also outside of our data set, which ended on April 2, 2020. The result is a 7.85% (0.0785) probability of being in the last 3 months of a bull run, which is assuring that our model has some predictive power as a stock market crash did not occur 3 months after September 21, 2020.

XLE (31.97), XLF (23.93), XLU (57.66), XLP (62.83), XLY (141.51), XLV (103.18), SPY = (326.97)

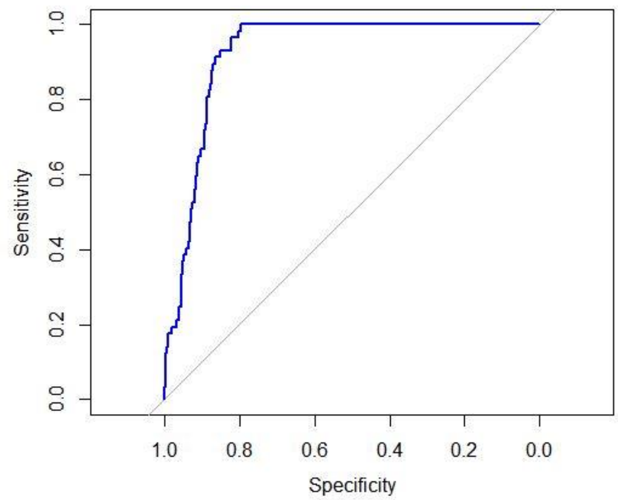
```
sep212020df <- data.frame(Date = c(2020-09-21), XLEclose = c(31.97),
                      XLFclose = c(23.93), XLUclose = c(57.66),
                      XLPclose = c(62.83), XLYclose = c(141.51),
                      XLVclose =(103.18), SPYclose = c(326.97))
predict(logitModel, newdata = sep212020df, type = "response")
```

```
         1
0.07847111
```

Area Under Curve – ROC (receiving operating characteristic)

We want to maximize the area under the curve for the most predictive power. Our ROC curve looks pretty good.

```
install.packages('pROC')
library('pROC')
roc(trainData$Final3Months, pred, plot = TRUE, col = "blue")
```



Assumptions

"Logistic regression does not assume the residuals are normally distributed nor that the variance is constant" (University of Cincinnati, 2017). However, independence among our independent variables is a concern, thus we must test for multicollinearity. With values much greater than 5 for XLE, XLU, XLP, XLY, XLV, great concern arises. We may have to throw out our model.

VIF

```
library(car)
vif(logitModel)
```

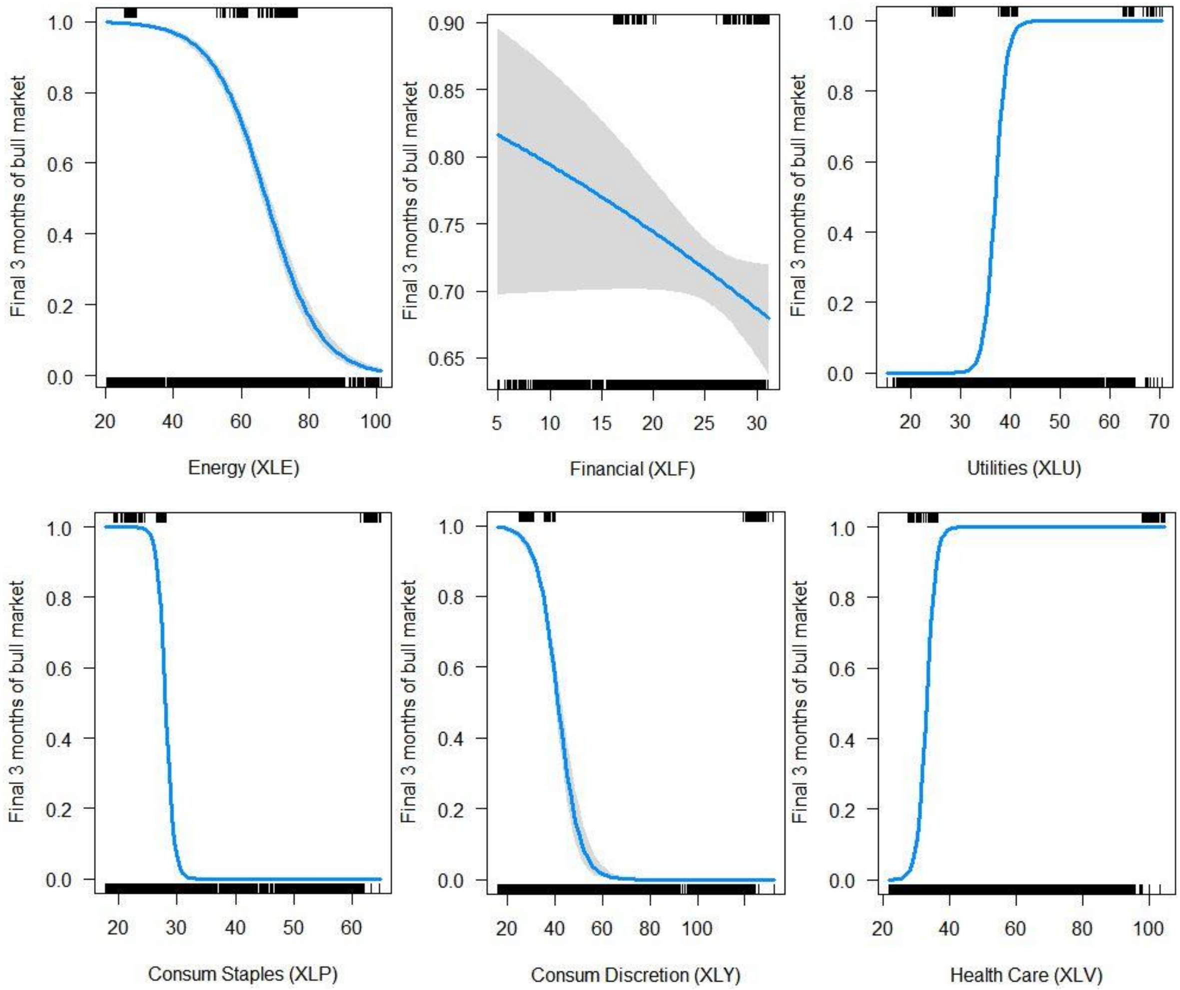
```
XLEclose   XLFclose   XLUclose   XLPclose   XLYclose   XLVclose
8.759147   4.731363  93.953278 269.886451 288.175345 401.868054
```

Visualization

To finish out our logistic regression model, let's just take a moment to visualize the binomial logistic regression line of our six market sectors. We can see why Financials (XLF) got such a high p-value.

```
install.packages('visreg')
library(visreg)
visreg(logitModel, 'XLEclose', scale='response', rug=2,
       xlab='Energy (XLE)', ylab='Final 3 months of bull market')
visreg(logitModel, 'XLFclose', scale='response', rug=2,
       xlab='Financial (XLF)', ylab='Final 3 months of bull market')
visreg(logitModel, 'XLUclose', scale='response', rug=2,
       xlab='Utilities (XLU)', ylab='Final 3 months of bull market')
visreg(logitModel, 'XLPclose', scale='response', rug=2,
       xlab='Consum Staples (XLP)', ylab='Final 3 months of bull
market')
visreg(logitModel, 'XLYclose', scale='response', rug=2,
       xlab='Consum Discretion (XLY)', ylab='Final 3 months of bull
market')
```

```
visreg(logitModel, 'XLVclose', scale='response', rug=2,
       xlab='Health Care (XLV)', ylab='Final 3 months of bull market')
```



Kruskal-Wallis Test

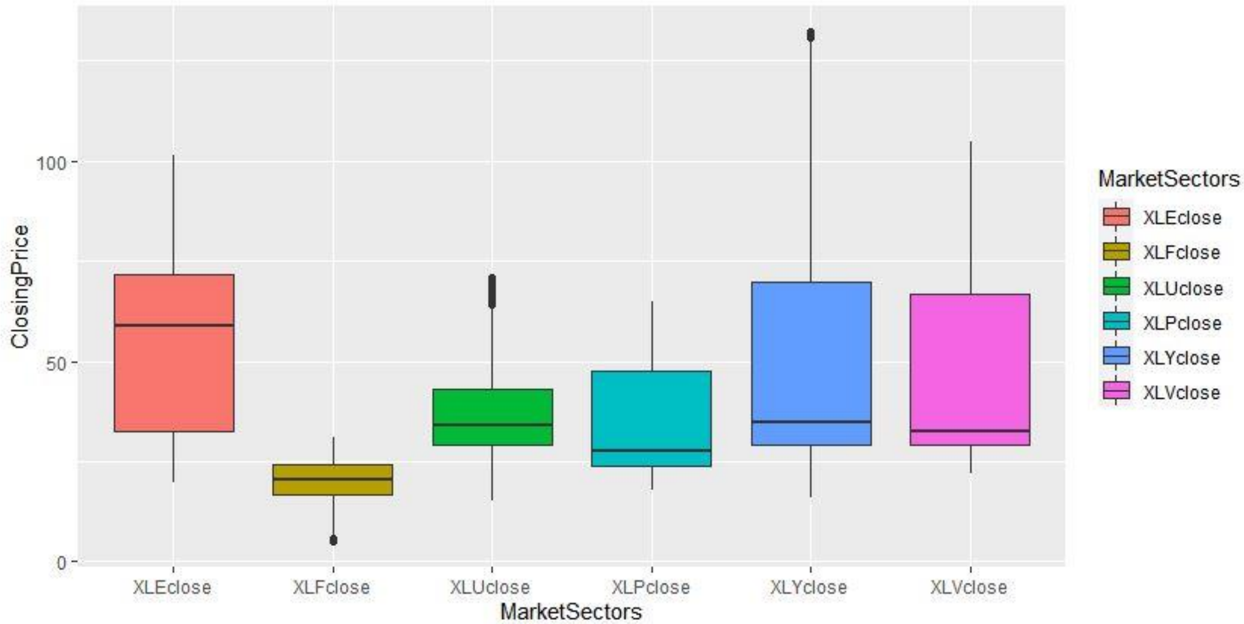
Since our VIF test did not work out as well as hoped, let’s see if we can salvage some use from our data using a nonparametric approach, the Kruskal-Wallis Test.

Boxplot by Sector

First, we need to reshape our original dataframe (‘mydf’) using melt() to fit all our market sectors into one column. Consumer Discretionary (XLY) has the broadest range, while Financials (XLF) has the narrowest.

```
library(reshape2)
newshape <- melt(mydf, id.vars='Date', measure.vars=c('XLEclose',
'XLFclose',
                    'XLUclose', 'XLPclose', 'XLYclose', 'XLVclose'))
names(newshape)[2] <- 'MarketSectors'
```

```
names(newshape)[3] <- 'ClosingPrice'
newshape %>% ggplot(aes(x=MarketSectors, y=ClosingPrice,
                        fill=MarketSectors)) + geom_boxplot()
```



Hypothesis

H0: There is **no** significant difference of medians among market sectors Energy (XLE), Financials (XLF), Utilities (XLU), Consumer Staples (XLP), Consumer Discretionary (XLY), Health Care (XLV).

H0: There is a significant difference of medians among market sectors Energy (XLE), Financials (XLF), Utilities (XLU), Consumer Staples (XLP), Consumer Discretionary (XLY), Health Care (XLV).

Kruskal-Wallis Test

```
kruskal.test(ClosingPrice ~ MarketSectors, data = newshape)
```

```
       Kruskal-Wallis rank sum test
data:  ClosingPrice by MarketSectors
Kruskal-Wallis chi-squared = 12698, df = 5, p-value < 2.2e-16
```

Conclusion

The p-value from our test is much less than 0.05. We can reject the null hypothesis. There is a statically significant difference of medians among our 6 stock market sectors.

Multiple pairwise-comparisons – Wilcoxon Signed Rank test

Since we know that there is a difference among our 6 different groups (from the Kruskal-Wallis Test), let's dive deeper to see which pairs in our group are different using the Wilcoxon Signed Rank test. The output below shows that all pairs have a p-value much less that 0.05, showing that all 6 economic market sector's pricing is significantly different from each other.

```
pairwise.wilcox.test(newshape$ClosingPrice, newshape$MarketSectors,
                     p.adjust.method = "BH")
```

```
     Pairwise comparisons using Wilcoxon rank sum test with continuity
correction
data:  newshape$ClosingPrice and newshape$MarketSectors
         XLEclose XLFclose XLUclose XLPclose XLYclose
XLFclose < 2e-16  -        -        -        -
XLUclose < 2e-16  < 2e-16  -        -        -
XLPclose < 2e-16  < 2e-16  < 2e-16  -        -
XLYclose < 2e-16  < 2e-16  < 2e-16  < 2e-16  -
XLVclose < 2e-16  < 2e-16  < 2e-16  < 2e-16  8.7e-08
```

References

Cannon, A. R., Cobb, G. W., Hartlaub, B., Legler, J. M., Lock, R. H., Moore, T. L., Rossman, A. J. & Witmer, J. A. (2013). STAT2 - Building Models for a World of Data. Retrieved from https://ismayc.github.io/teaching/sample_problems/multiple_logistic.html

Chen, J. (2021, March 3). Exchange Traded Fund (ETF). Investopedia. https://www.investopedia.com/terms/e/etf.asp

Frank, T. (2020, August 18). Here's a List of Stock Bull Markets Through Time and How this New One Stacks Up. CNBC. https://www.cnbc.com/2020/08/18/heres-a-list-of-stock-bull-markets-through-time-and-how-this-new-one-stacks-up.html

Hubert, M. (2021, August 6). Here's Another Sign the Bull Market is Near a Peak, and This One Bears Watching. MarketWatch. https://www.marketwatch.com/story/heres-another-sign-the-bull-market-is-near-a-peak-and-this-one-bears-watching-11628233932?mod=home-page

Kassambara. (2018, November 3a). Logistic Regression Essentials in R. STHDA. http://www.sthda.com/english/articles/36-classification-methods-essentials/151-logistic-regression-essentials-in-r/

Mangiafico, S. S. (2015). Multiple Logistic Regression. R Companion. https://rcompanion.org/rcompanion/e_07.html

McGill University. (n.d.). Goodness of Fit in Logistic Regression. Faculty of Medicine and Health Sciences. https://www.medicine.mcgill.ca/epidemiology/joseph/courses/epib-621/logfit.pdf

Onyshchak, O. (2020). Stock Market Dataset: Historical Daily Prices of Nasdaq-Traded Stocks and ETFs [Data file]. Kaggle. https://www.kaggle.com/jacksoncrow/stock-market-dataset

Prabhakaran, S. (2017, September 13). Logistic Regression – A Complete Tutorial with Examples in R. Machine Learning +. https://www.machinelearningplus.com/machine-learning/logistic-regression-tutorial-examples-r/

Stankiewics, K. (2021, August 4). Charts Suggest the Stock Market is Getting Close to a Near-Term Top, Cramer says. CNBC. https://www.msn.com/en-us/money/markets/charts-suggest-the-stock-market-is-getting-close-to-a-near-term-top-cramer-says/ar-AAMUnqi

University of Cincinnati. (2017, February 3). Logistic Regression. UC Business Analytics R Programming Guide. https://uc-r.github.io/logistic_regression

Zach. (2019, January 18). Kruskal-Wallis Test: Definition, Formula, and Example. Statology. https://www.statology.org/kruskal-wallis-test/