Cross Examination and Risk Analysis of the Consumer Staples and Consumer Discretionary Sectors

Shea Grant, Dominic Piselli, and Zachary Jones

QF 301: Advanced Time Series Analytics and Machine Learning

Prof. Feinstein

December 5th, 2023

Abstract

This paper presents a comprehensive analysis of the Consumer Staples (XLP) and Consumer Discretionary (XLY) sectors, utilizing advanced machine learning techniques to forecast future volatility and assess model performance. The machine learning methods implemented include linear regression models, random forest models with bagging and feature selection, and hierarchical and K-means clustering. Our results indicate that the linear regression models were highly effective, with minimal mean squared errors observed across all tests and highlights that simpler models with fewer significant predictors can be equally effective. The random forest models exhibited varying degrees of accuracy, with some combinations leading to increased mean squared errors, suggesting a trade-off between complexity and predictive power. Across the different models, the future volatility for XLP ranges from 0.1023895 to 0.1228586 and the future volatility for XLY ranges from 0.2453641 to 0.3394655. Future research may focus on utilizing these models to expand to more sectors or individual stocks and construct a diversified portfolio with varying volatilities.

Introduction

The landscape of the global economy presents a complex tapestry of sectors, each characterized by unique factors and growth potentials. Among these, the Consumer Staples and Consumer Discretionary sectors represent two critical components, distinguished not only by their inherent business characteristics but also by their contrasting risk profiles. This paper aims to conduct a thorough comparative analysis of these sectors, particularly focusing on utilizing different machine learning techniques to forecast the future volatility of these sectors.

The consumer staples sector, a cornerstone of the global economy, represents a wide array of products essential to daily life. This includes industries such as food and beverage, household goods, personal products, and tobacco. This sector is defined by its non-cyclical nature and price inelasticity, as the products comprising it are still purchased despite economic downturns (Chen, 1). Popular companies in this sector include Costco Wholesale, PepsiCo, The Coca-Cola Company, and The Proctor and Gamble Company. Consumer staples companies enjoy stable revenue streams and have embraced online sales channels during the rise of e-commerce and digital marketing. Due to this steadiness in consumption and defense against economic instability, the consumer staples sector is known to be quite nonvolatile in economic markets.

The Consumer Discretionary sector, in contrast, is characterized by goods and services that are non-essential and purchased by consumers when they have excess income. This sector encompasses industries like automobiles, textiles, consumer electronics, leisure equipment, and services such as hotels, restaurants, and other leisure facilities. It is considered cyclical, tending to outperform during economic expansions when consumer confidence and disposable income are high, but underperform during recessions and downturns (Team, 1) As a result, it maintains higher risk and volatility, especially during fluctuating economic climates.

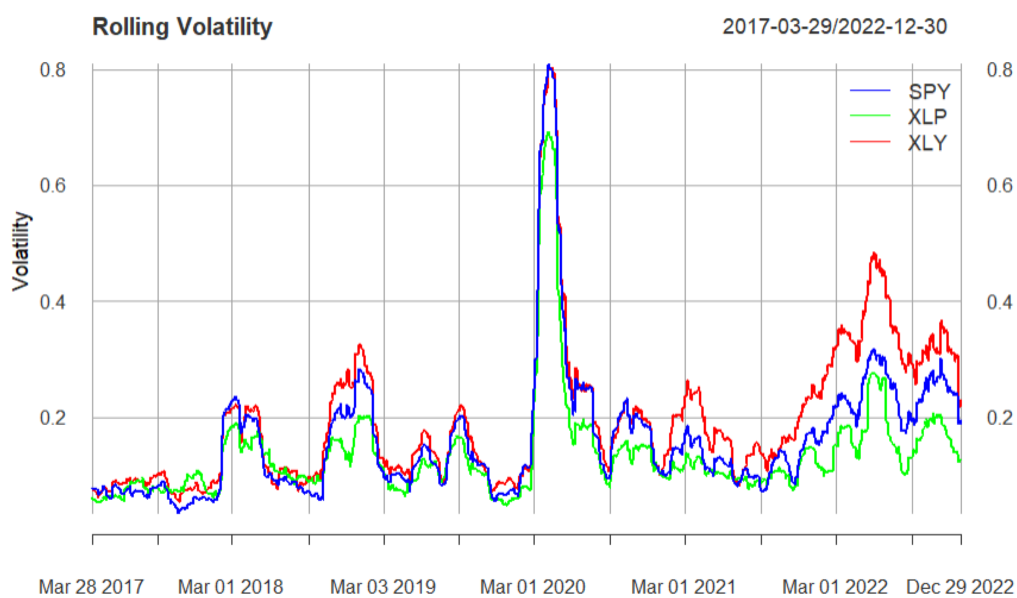
Methodology

Predicting future volatility of stocks is crucial for several reasons, especially in the realms of investment management, risk assessment, and strategic planning. By predicting future volatility, investors and fund managers can assess the risk associated with a particular stock or portfolio. This information is vital for making informed decisions about asset allocation, diversification, and risk tolerance alignment. For example, Investors can balance high-volatility stocks with lower-risk assets to achieve desired risk-return profiles. Understanding volatility trends assists in selecting stocks that align with the investor's investment strategy, whether it be conservative, moderate, or aggressive. The purpose of this study is not to create a diversified portfolio with different risk measures, but does attempt to aid in that process.

In order to gather the necessary data for analysis, we gathered the log returns of 30 random stocks from each sector from yahoo finance between January 1st, 2017 and January 1st, 2023. These dates provide us with an accurate representation of different types of markets and unforeseen market downturns such as the market collapse following the outbreak of COVID-19. A sample size of 30 stocks was chosen because a value equal to or greater than 30 allows us to make predictions and conclusions for the entire population. Upon calculating the log returns of all 60 stocks, a helper function was created to calculate the rolling volatility of each stock. First, the closing prices were converted into logarithmic returns. The logarithmic returns were then averaged over a 30 day period. Next, the squared deviations were calculated and averaged over the 30 day window. Lastly, the rolling volatilities were computed by taking the square root of the average squared deviations and were annualized through multiplying by the square root of 252, the amount of trading days in a year.

The first model we chose to run was K-means clustering within the two sector ETFs - consumer staples and consumer discretionary. The clustering was run on each ETF by itself to discover if there were any stocks in particular that came together within the individual sectors. At first the stocks were sorted into 4 different clusters but after looking at the results we thought it may be more beneficial to increase the number of groups the stocks were being sorted into because the majority of the stocks were grouped together in only 1 or 2 groups. After playing around with the number of clusters the model was running, we decided that 6 groups was a good balance to actually see how the 30 stocks in each ETF sector were coming together. As the number of groups increased higher than 6 most, of the clusters only contained a single stock without much information to analyze.

After analyzing the clustering of each of the stocks inside the two sector ETFs, we calculated the rolling volatility of the entire ETFs along with the rolling volatility of the S&P 500 using the same helper function that was created for our clustering. This data will be utilized in each of the regression techniques in the study, with 80% of the data used for training the models and the remaining 20% for testing purposes.



In an attempt to conduct a comprehensive approach to analyzing the financial data of the two sectors and predict future volatility, we decided to implement a series of linear regressions. These regressions allow us to make predictions on future volatility and assess which predictors are better than others. To cover all bases in terms of predictors, four total regressions were implemented. We first predict XLY using three XLY lags and three SPY lags. Next, we predict XLP using three lags for XLP and SPY. Using these two models, we predict the future volatility of both ETFs one year in advance. Following this, in order to discover how XLY and XLP work together, we implemented the next two regression models. These consist of predicting XLY and then XLP with three lags for all three ETFs, those being XLP, XLY and SPY. To see how predictions have changed for the new models, we predict the one year future volatility once more. Additionally, in order to assess the model accuracy and discover possible overfitting, the train and test mean squared error are computed for all four models.

Nonlinear models are also explored, specifically random forest regressions with bootstrapping and feature selection. The decision tree-based models were regressed with the same predictors as the multiple linear regressions. The feature selection models were extrapolated from analyzing the mean decrease in impurity plots and model accuracies were assessed through their train and test mean squared errors.

Results

When running the K-means clustering to both of the ETF sectors individually, it showed a lot of similar companies grouping together. Looking at the Consumer Staples (XLP) sector: two competitors, Dollar Tree and Target were in a cluster, food companies like Sysco Corp, Tyson Foods Inc, Lamb Weston Holdings Inc, and Bunge Limited were also all grouped together. The remaining companies were split fairly evenly between two other clusters having distribution companies like Walmart, Costco, General Mills, and PepsiCo in a cluster, with the exception of two outlying companies. Both Kraft Heinz Co and Keurig Dr Pepper Inc both were alone in their own cluster and did not find any other stocks to join. There were similarities with the clustering in the Consumer Discretionary (XLY) sector as well, Carnival Corp and Royal Caribbean Group were clustered together and Ulta and Lulu lemon were in their own group. Automotive companies like Ford Motor Co, General Motors Company, Aptiv PLC, and LKQ Corp were also all in the same cluster. The clustering between both of the ETF sectors were very positive and similar in the aspect that they were grouping together companies that were either alike in product or company design for staples and discretionary. Below shows where the remaining companies fell into each of the clusters,

K-means Clustering of XLP

```
  SYY.Close  TSN.Close   BG.Close  WBA.Close   LW.Close DLTR.Close  TGT.Close  KDP.Close  KHC.Close  WMT.Close   KO.Close  GIS.Close 
          1          1          1          1          1          2          2          3          4          5          5          5 
 COST.Close  KMB.Close  PEP.Close   CL.Close  HSY.Close  CHD.Close   PG.Close  MKC.Close  HRL.Close   UL.Close MDLZ.Close   DG.Close 
          5          5          5          5          5          5          5          5          5          5          5          6 
   KR.Close MNST.Close  CLX.Close  CPB.Close BF.B.Close   EL.Close 
          6          6          6          6          6          6 
```

K-means Clustering of XLY

```
TSLA.Close  MCD.Close   HD.Close SBUX.Close  TJX.Close ORLY.Close   TM.Close  AZO.Close  GPC.Close  NKE.Close AMZN.Close 
         1          2          2          2          2          2          2          2          2          3          3 
 CMG.Close  HLT.Close TSCO.Close EBAY.Close  HAS.Close  MAR.Close ROST.Close    F.Close   GM.Close  LEN.Close APTV.Close 
         3          3          3          3          3          4          4          4          4          4          4 
 NVR.Close  DRI.Close  LKQ.Close  WHR.Close LULU.Close ULTA.Close  RCL.Close  CCL.Close 
         4          4          4          4          5          5          6          6 
```

Upon completing the four linear regressions, it was discovered that all four were effective in predicting future volatility and were accurate models that reduced overfitting. Of the four models, the highest mean squared error we discovered was 0.0001, which occurred in the test data when predicting XLY with XLY and SPY. From this, we can conclude that no model was overfitting and each effectively completed its task in predicting with little error. That being said, some models would be better to use in later analysis than others. For example, both future volatility predictions were outstandingly similar. For the first two models using just two predictors, the 1-year ahead volatility predictions for XLP and XLY were 0.1023895 and 0.3320747, respectively. The next set of predictions for XLP and XLY following the models using all three ETFs as predictors were 0.1201972 and 0.3394654, respectively. These values make sense in the context of the study, as the Consumer Discretionary sector traditionally is more volatile than Consumer Staples. However, where the two types of models differed was the significance of their predictors. Seen below, as the number of predictors increases, the number of those that are statistically significant decrease and their respective p-values increase, proving less statistical significance.

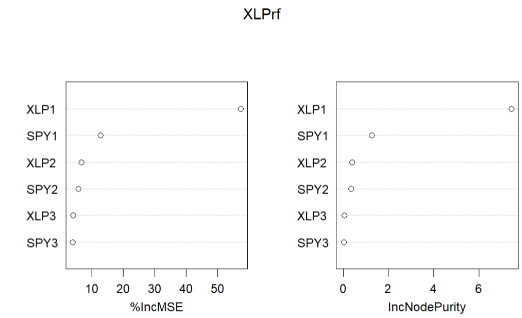
```
Call:
glm(formula = yXLP ~ ., data = XLPdf, subset = train)

Coefficients:
              Estimate Std. Error t value             Pr(>|t|)    
(Intercept)  0.0012154  0.0004095   2.968             0.003057 ** 
XLP1         1.0968613  0.0397362  27.604 < 0.0000000000000002 ***
XLP2         0.0084198  0.0545448   0.154             0.877349    
XLP3        -0.1328298  0.0384398  -3.456             0.000569 ***
SPY1         0.1514960  0.0341402   4.437           0.00000998 ***
SPY2         0.0067189  0.0469920   0.143             0.886331    
SPY3        -0.1425338  0.0339416  -4.199           0.00002883 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

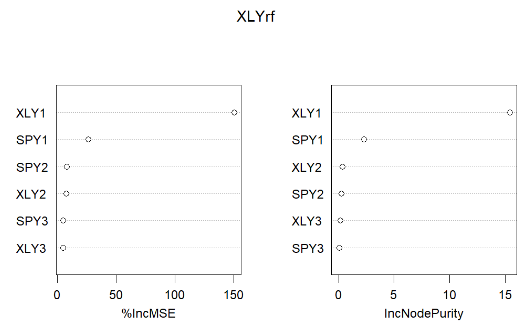
```
Call:
glm(formula = yXLP ~ ., data = XLPdf2, subset = train)

Deviance Residuals:
      Min         1Q     Median         3Q        Max  
-0.068582  -0.002192  -0.000203   0.001893   0.080337  

Coefficients:
              Estimate Std. Error t value             Pr(>|t|)    
(Intercept)  0.0014966  0.0004623   3.237              0.00124 ** 
XLP1         1.1047327  0.0404200  27.331 < 0.0000000000000002 ***
XLP2        -0.0078938  0.0556907  -0.142              0.88731    
XLP3        -0.1289316  0.0386967  -3.332              0.00089 ***
XLY1         0.1070855  0.0482821   2.218              0.02676 *  
XLY2        -0.1276032  0.0663323  -1.924              0.05464 .  
XLY3         0.0114683  0.0475337   0.241              0.80939    
SPY1         0.0401545  0.0607149   0.661              0.50851    
SPY2         0.1401269  0.0858753   1.632              0.10301    
SPY3        -0.1515832  0.0597860  -2.535              0.01136 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In machine learning, the goal is to use simpler models without sacrificing model efficiency or accuracy. So since both types of models result in basically identical future volatility predictions and the second set of regressions contain less significant predictors, the best model to utilize would be that consisting of either XLP or XLY and SPY.

The first random forest model regressing XLP's rolling volatility onto SPY and XLP's one to three day volatility lags had a training mean squared error of 0.0000240 and 0.0000682 test mean squared error. When XLY's one to three day rolling volatility lags were also included in the model, the training mean squared error decreased to 0.0000228, while the test mean squared error increased to 0.0000687. The next model regressed XLY's rolling volatility onto SPY and XLY's one to three day volatility lags. The train mean squared error was 0.0000248 with a test mean squared error of 0.0001383, which was a slightly worse error than the XLP model's test mean squared errors. When XLP's one to three day rolling volatility lags were also added as predictors, the train mean squared error slightly decreased to 0.0000215, and test mean squared error also decreased to 0.0001311. Adding the three more predictors negligibly improved the mean squared errors and caused the models to grow in complexity. Looking at the mean decrease in impurity plots, the only important variables for both XLP and XLY were their respective 1 day volatility lags, as there was a significant gap between XLP and the next most important predictor, SPY 1 day volatility lag.

The feature selection random forest models performed worse than the previous random forest regressions. Regressing XLP's volatility onto XLP's one day lag resulted in an increased train mean squared error of 0.0000277 and a test mean squared error of 0.0000719. This model predicts the one year future volatility of XLP as 0.1204117. When XLY's volatility was regressed onto its one day volatility lag, the train mean squared error increased to 0.0072971 and the test mean squared error also increased to 0.0077186. The one year future volatility forecast based on XLY's feature selection model evaluated to 0.270459. Out of all the random forest models, the feature selection regressions were the simplest, as they only had one predictor. However, they performed worse in terms of mean squared error which indicates worse accuracy in their predictive capabilities.





Model	Train MSE	Test MSE	One year Forecast
XLP linear	0.0000589	0.0000393	0.1023895
XLP with XLY lags linear	0.0000585	0.0000398	0.1201972
XLP random forest	0.0000240	0.0000682	0.1228586
XLP with XLY lags random forest	0.0000228	0.0000687	0.1153698
XLP feature selection random forest	0.0000277	0.0000719	0.1204117
XLY linear	0.0000921	0.0001035	0.3320747
XLY with XLP lags linear	0.0000901	0.0000991	0.3394655

XLY random forest	0.0000248	0.0001383	0.2534778
XLY with XLP lags random forest	0.0000215	0.0001311	0.2453641
XLY feature selection random forest	0.0072971	0.0077186	0.270459

Discussion

Prior to conducting the models, our preliminary predictions for their performance and predictive capabilities indicated that as we introduced more predictors, the mean squared error would significantly decrease. However, that was not the case, as the models with more predictors performed almost identically. Therefore, the models that would be best to implement in the real world would be the volatility lags of SPY paired with only XLP or XLY as predictors, not both. We also expected that models predicting XLY would result in higher future volatility predictions and higher mean squared error values since the ETF is traditionally more volatile than XLP. The data was true to this prediction. If an individual chooses to predict XLP or XLY, we recommend that they do so with a multiple linear regression, as they had the lowest test mean squared errors when compared to the random forest models. Interestingly, the future volatility predictions for the random forest models are consistently lower than that of the linear regressions. This is due to the fact that random forests are better at capturing complex, non-linear relationships and are less likely to overfit to the training data due to their ensemble learning method. It builds multiple decision trees and merges their predictions, which can lead to more conservative volatility predictions. It additionally is less sensitive to outliers, which were in fact present in the data sets. This could also explain why the training mean squared errors were lower than the linear regressions. It is therefore possible that the future volatility predictions for the XLY random forest models are more accurate than the XLY linear regression models.

In the beginning stages of running the machine learning techniques there were some models that seemed to be overfitting or not very helpful in our analysis that ended up being left out of our project. Aftering running a simple decision tree on our data the train mean squared error was significantly higher than our test mean squared error telling us the model is overfitting the data not being very helpful. The naive bayes model was also similar, not being very helpful, because after running it the accuracy was very low and the confusion matrix was also very unbalanced.

The next steps in this project are to implement a way to gauge the accuracy of the one year volatility forecasts. Training the models on one year of data, predicting the one year volatility, and moving the timeframe one month ahead would be an effective way to find the standard deviation of the predictions, which would further reflect the accuracy of the one year forecasts. Another way to effectively expand upon this study would be creating a new portfolio based on varying levels of future volatility. In this expanded study, we could consider a more extensive comparative analysis between a larger number of ETFs, predict their respective future volatilities, and even extend the forecasting horizon beyond one year to analyze long-term trends. This would allow us to conduct more comprehensive models and create a diverse, extensive portfolio. Creating a portfolio would also be more applicable to real-world scenarios in which investors actively seek to manage risk while maximizing returns.

Resources

- Chen, J. (n.d.-a). *Consumer staples: Definition, role in GDP, and examples*. Investopedia. https://www.investopedia.com/terms/c/consumerstaples.asp#:~:text=The%20term%20consumer%20staples%20refers,well%20as%20alcohol%20and%20tobacco
- Team, T. I. (n.d.). *What is consumer discretionary? definition in economic indicators*. Investopedia. https://www.investopedia.com/terms/c/consumer-discretionary.asp

Appendix

Zachary-I wrote the code for clustering algorithms, rolling volatility function, predicting future volatility function, XLP, XLY, and SPY dataframes, linear regressions, random forests, feature selection, and train/test mean squared error evaluations. I contributed to part of the methodology, specifically the rolling volatility calculation method and random forests with feature selection included. I wrote the results of the random forest models and feature selection random forest regressions, and I added to the discussion how the multiple linear regression models performed best in terms of mean squared error and how we can further evaluate the accuracy of our one year forecasts. Lastly, I added some plots and tables to the paper.

Dominic- During the duration of this project I contributed by obtaining the adjusted close prices of our stocks from Yahoo Finance and created the data frames of our lags and log returns. After the rolling volatility was computed of our stocks inside the ETF sectors I changed around our clustering models in order to find a number of clusters that best fit the data along with drawing conclusions from our clustering models. I looked into the Naive Bayes model however was unable to create a useful model that helped give us useful information. Added a few visualization pieces to the final report after we gathered all of our results together.

Shea- For this project, most of my contribution came in the form of writing the final paper. This included writing the Abstract, introduction, much of the methodology, the linear regression section of the results, and the next steps / future studies section of the conclusion. Aside from the paper, I was involved in preliminary stock research and aided in the writing of r code.