Trading Strategy based on Spread between Implied Volatility and Realized Volatility using Classification Tree

Shenrui Jin

Fordham University

December 20, 2016

Abstract

This study demonstrates that VIX could be used to generate indicator, since it is known as the ‘fear gauge’. The study tries to build the simple indicator based on the spread between the implied volatility and the realized volatility. It is a well-known phenomenon that the implied volatility is most of the time higher than the historical volatility. However, it is possible that the spread tightens or even becomes negative. This signals a stress and timing to short equity position. The study then analyzes the source of performances and demonstrates the weakness of this strategy. The result indicates that simple strategy doesn’t outperform because of correctness. Therefore, there are much more space to improve the strategy and keep the advantage by increasing the correctness. The study applies the Classification Tree in the spread between implied volatility and realized volatility to achieve better projection of recession. Overall, the analysis supports that spread between the implied volatility and realized volatility could be used to avoid biggest drawdown for portfolio.

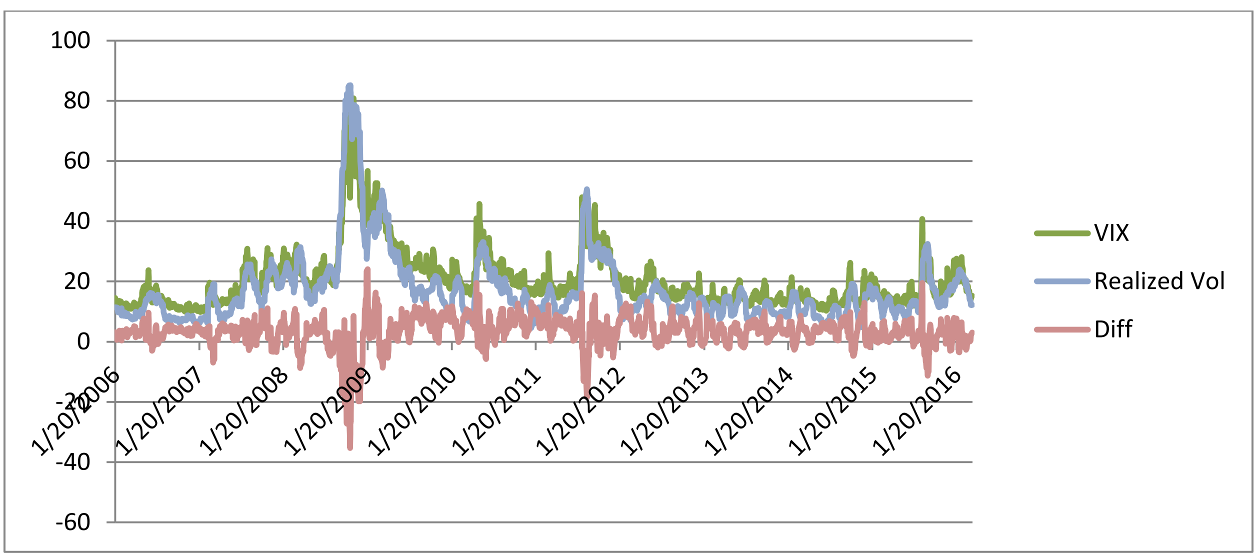
Keyword: VIX, implied volatility, realized volatility, trading signal, classification tree

1. Introduction

VIX is a trademarked ticker symbol for the CBOE Volatility Index, a popular measure of the implied volatility of S&P 500 index options; the VIX is calculated by the Chicago Board Options Exchange (CBOE). VIX is often referred to as the fear index or the fear gauge. the VIX represents one measure of the market's expectation of stock market volatility over the next 30-day period. The way to calculate this index is calculated using many out of money European Call or Put Options with 1-month tenor. Since VIX comes from the market price, so that it contains the views of investor toward the future market. When VIX increases, the level of fear in the market bumps up and people are buying protection. When VIX decreases, Market is more optimistic about the future. VIX is regarded as the implied volatility of the market, while realized volatility is calculated using the stock prices. One well known phenomenon is that implied volatility tends to be higher than realized volatility consistently. One possible reason is that option sellers are requiring more premium from the naked options. Buyers of options have limitation in the loss because they can lose no more than the premium they pay. While sellers take all the risk if the market go to the wrong directions, especially for call sellers. The loss caused from call option could be infinite. This means that there is an insurance premium involved, so that sellers of the option want to be compensated for their additional risk. The second reason is model misspecification. Implied volatility comes from model and market price. If the model is different, the implied volatility will be different as well. For example, Black Scholes Merton Model has an assumption that volatility is constant which is not true. So there is a bias between implied volatility and realized volatility.

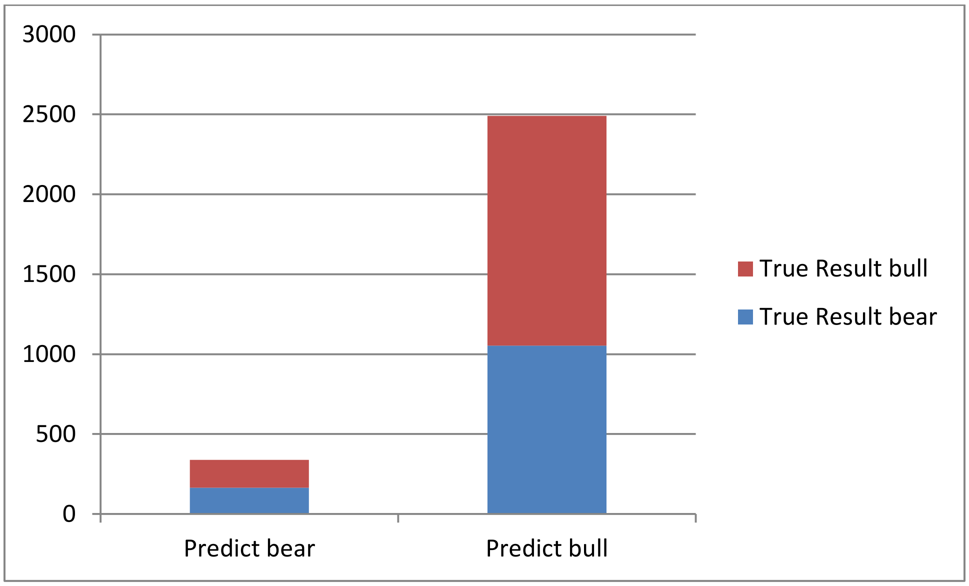
If we look deeper into the relationship between the VIX and realized volatility, we can find that it is possible that implied volatility is lower than realized volatility. This event usually happened when the market is very stress. So a generic idea is that we could use the VIX and realized volatility to build a signal for trading equity. Specially, positive equity returns tend to follow the positive spread between the implied volatility, while negative equity returns tend to follow the negative spread. It is developed from the seminal research from (Giot 2006). Giot (2006) suggests that there is a positive relationship between forward looking returns and extremely high level of the implied volatility and negative relationship between the return and extremely low level of the implied volatility. This means that an increase in the VIX level could coincide with a market moving downward. This is suggested by Whaley (2009) as well. This means that high volatility levels may suggest an over-sold market. This study combines the implied volatility and realized volatility to calculate technical trading rule which can outperform the VIX only strategy and buy-and-hold strategy.

Figure 1: Spread between VIX and Realized Volatility



This study also analyzes the source of the performance of VIX trading strategy. It turns out that the spread between the implied volatility and realized volatility doesn't give better correctness than we just think it is always bull market.

Figure 2: Correctness of Simple Spread Indicator



This study applies the Classification Tree into the spread indicator to explore better correctness and maintain the performance this strategy got before. Classification Trees is one of the algorithm trading methodologies. Algorithmic trading is a method where a computer is conducting a specific investment instead of a human. As described in the literature, these trading systems implement historical data with respect to well-defined rules, whereas traditional trading only implements a specific strategy. Classification tree will provide a better perspective. It has several advantages compared to other methodologies.

1. Classification Tree is easy to understand and straightforward for market.

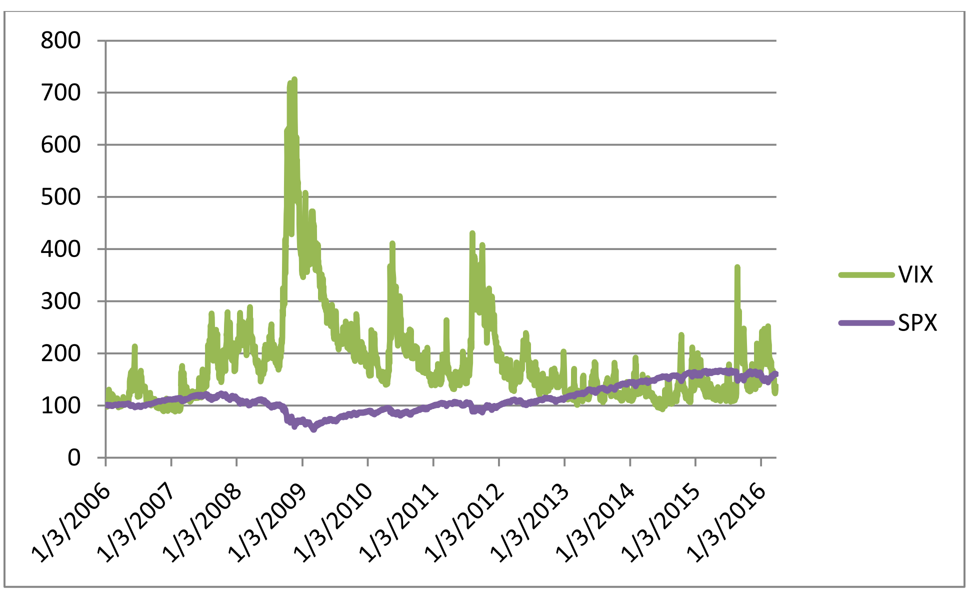
2. Equally appropriate for both ordered data and for categorical data or a mixture of them
3. Good at handling nonhomogeneous relationships with respect to conditional information.

This study will show how we could improve the correctness using the classification tree.

2. Data Sets

The Data that we use in this study comes from VIX, SPX. The time period under investigation spans from 1/1/2006 to 3/28/2016, which contains a total of 2574 observations per index.

Figure 3: SPX and VIX



3. Simple Strategy based on VIX

This study generates a trading signal which reflects the characteristic of 'fear gauge'.

First we calculate the historical volatility as follows:

$$\sigma_{hist_t} = \sqrt{\frac{252}{19} \sum_{i=1}^{20} \ln\left(\frac{Index_i}{Index_{i-1}}\right)^2}$$

The spread is calculated as:

$$Spread_t = Ivol_t - \sigma_{hist_t}$$

This trading signal has two states including normal market conditions and stressed market. In order to build the trading signal, we compute statistic:

$$Percentile_t = Percentile\ (Spread_{t-199}{:}\ Spread_t, 10\%)$$

This means that we will take 10% percentile for the last 200 spread observations.

It is natural to consider that we are in stressed market if $Spread_t < Percentile_t$, otherwise in normal market conditions. In order to back test the indicator, we will invest will invest SPX and exposure is 100% when we are in normal market condition, while we will short SPX and exposure is -100% when we are in stressed market. We can take short exposures to the SPX by shorting the first nearby future contract on SPX: ES1.

Figure 4: Back-Test on Simple Strategy

This strategy shows Sharpe 0.91 and Max Drawdown -27.62% which is quite impressive performance.

Table 1: Performance of simple strategy

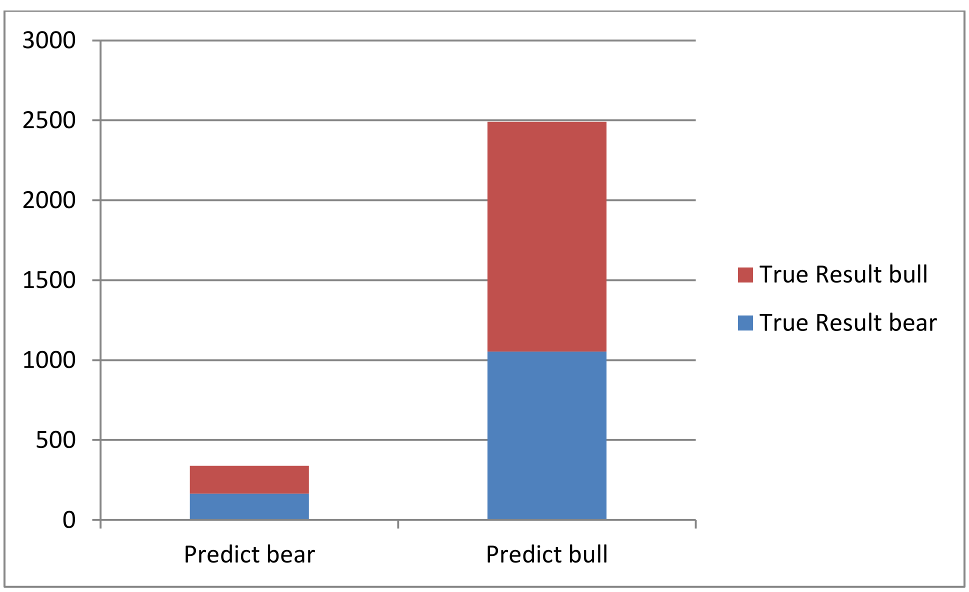
Statistic	Value
Volatility	21%
IRR	19%
Sharpe Ratio	0.91
Sortino	1.54
Max Drawdown	-27.62%

Then we try to find out how the correctness for VIX.

For measuring the performance, in each time step, if the outcome of the bearish strategy be higher, we consider it as bearish market, while if the bullish create a better performance, we consider it as bullish market.

Figure 5: True Market States

Figure 6: Correctness of Prediction



With higher chance of having bull market, we can use the most likely outcome event as the default prediction i.e. always predicting the market is bullish, this way we can achieve very high correctness. Surprisingly we see that we achieve an overall better performance in the current model which is even worse than the default prediction. So normally by such prediction we should not end up in index higher index, however this is not the case.

Table 2: Average performance for each category

Prediction	Correct	Wrong
Bear	0.02642	-0.02204
Bull	0.00707	-0.00870

As we can see the performance in bearish is more dramatic then the bull situation. We have the following mean performances. It’s now clear that the reason for having a such a good index in the current model is that when we predict bear, we are correct about 50% of the times, with higher rate of return in its correct predication compare to wrong prediction. It’s the reason why we do overall better than majority voting. So we think that we could improve the performance by increasing the correctness.

4. Classification Tree for Prediction

The problem in hand is a forecasting problem, given the previous spread we need to forecast the market states. There is a rich classification libraries in R that we like to use for this problem. At the same time, we need to consider time dimensions as well. The data points are not i.i.d. . This creates a great limitation, so instead we look to expand the feature space in a way that time ordering problem be avoided.

Now we like to look at a different set of features and examine the possibility of learning about the market states through their structure. For each data point, we consider the following features.

For each point in time, we look at past 9 available observation of vol spread and find the regression line between these points. Then we use the slope, intercept and the variance of error of the regression line as our new 3 features including slope, intercept and sig.

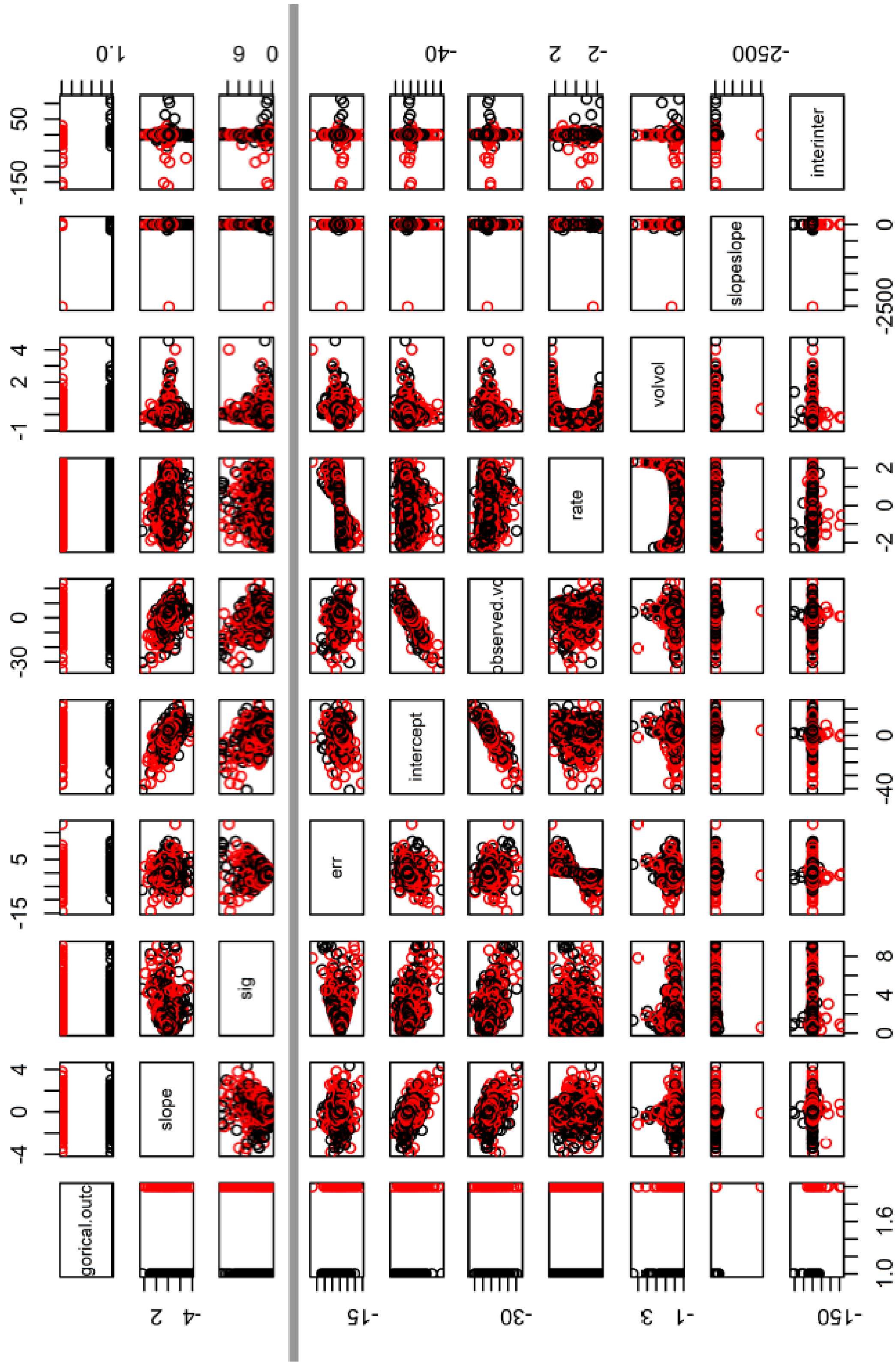
Moreover, we look at the difference of the last vol spread and the fitted value from regression line. Also we look at this difference divided by sig (Since we will use trees and the only classification provided by classification trees are step function we need to do these calculations and add them to data manually).

And finally we take the difference of the slope, intercept and sig from t and t-1 (difference divided by previous value).

In this way, we will have 9 different features including slope, sig, error, intercept, last spread, rate ,vol of vol, slope of slope and intercept of intercept. These features contain much more and higher dimension information about the VIX and realized volatility. In this way, we may find some logic using classification tree.

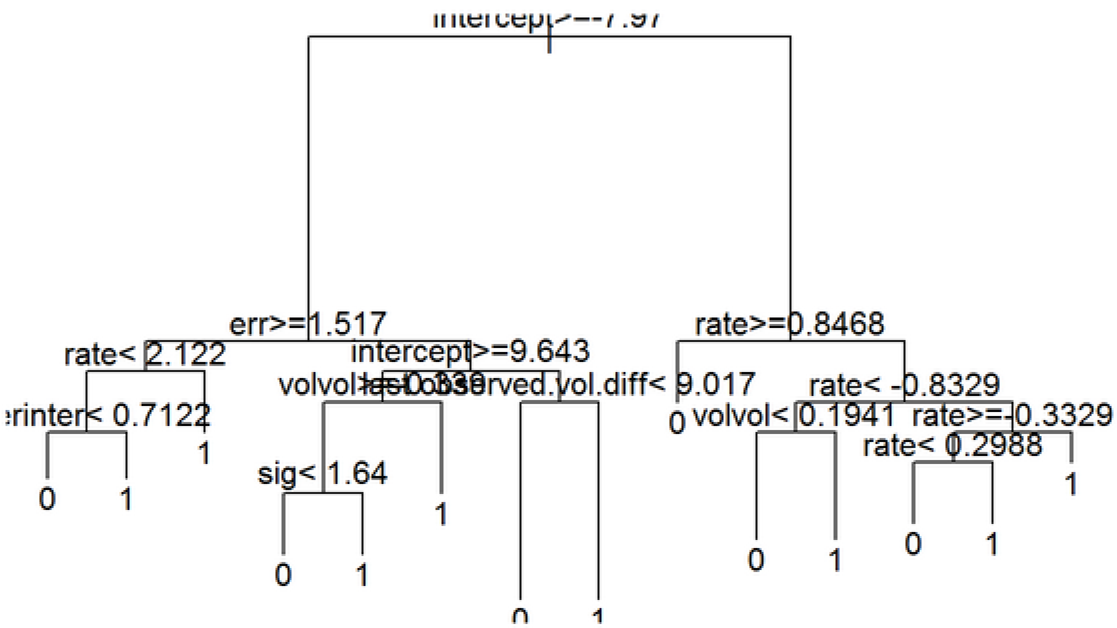
The goal is to use the tree to identify the possible trend in these variables that could help to identify the market states correctly.

Figure 7: Features Relationship



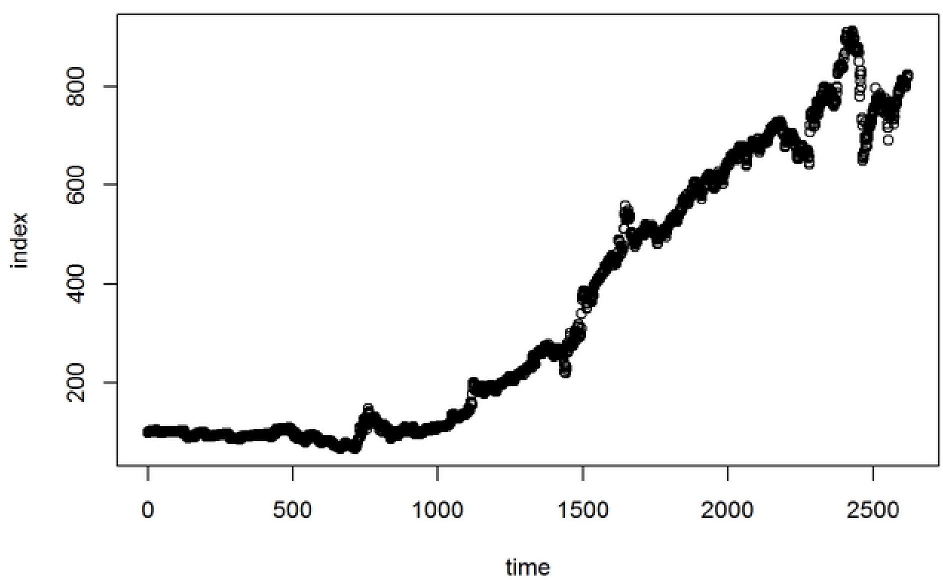
After calibrating the model using 2000 data points, we get the classification tree as follows:

Figure 8: Classification Tree for trading signal



If we use the same strategy as we used before, we could have much better performance.

Figure 9: Performance of Classification Tree strategy



Then we analyze the correctness again to see whether the model outperform the simple strategy. The rate of success on predicting bear is 0.5054545, while the rate of success on bull is 0.575693 and the bear prediction ratio is 0.1049618. we could see that we have better bear prediction correctness, which leads to higher performance.

5. Conclusion

This study set out to examine the relationship between the volatility spread and the market states. First trading rule is calculated by the simple indicator that Spread is lower than 10% percentile in last 200 days. This simple trading rule provides a decent performance. However, it doesn't have good prediction correctness if we look at the summery. Therefore, second trading rule is built using classification tree and other 9 higher dimensional features. The result reveals that, as opposed to the simple strategy, classification tree increases the correctness especially for bear market which could generate the large return. It turns out that performance of second strategy is twice as the first strategy.

Overall, the findings suggest that spread between implied volatility and realized volatility could signal the stressed market and advanced machine learning method could have better performance than the simple indicator.

Reference

James, K. and Camillo, L. (2007). Using VIX Data to Enhance Technical Trading Signals, Lakehead University.

Axel, S. (2008). Trading based on classification and regression trees.