Financial Markets Final Project

Spencer Gosden

April 2025

Topic: Exploring Implied vs. Realized Volatility as a Potential Trading Opportunity

1 Introduction

When discussing options contracts in class, I was extremely intrigued by the concept of implied volatility, and I wanted to explore it as a possible indicator of market movements. To do this, I looked at the VIX for the S&P 500. The VIX (short for Volatility Index) is a weighted average of the implied volatility for all options contracts with expiration dates between 23 and 37 days, roughly working out to average 30-day implied volatility. First, I wanted to investigate just how accurate the VIX was in anticipating volatility. From here, I sought to determine if it could be utilized in some way to predict large market shifts before they happen.

2 Data

For my data, I downloaded 25 years of historical data for both the S&P 500 and the Chicago Board of Exchange's VIX, from December 31st, 1999 to January 1st, 2025. I then did some simple feature engineering to create relevant features for my analysis and model. First, I converted returns of the S&P index to percent change, and created a new variable that tracked the rolling 30-day standard deviation of these returns. Additionally, I constructed a variable that tracked the implied volatility vs. the realized volatility by taking the difference of the S&P 500's 30-day rolling standard deviation and the VIX 30 days prior. Finally, I scaled all of the variables using a standard scaling to reduce bias in my model.

3 Technique & Analysis

First, I set out to investigate the relationship between the VIX implied volatility and S&P 500 realized volatility. Below is the graph of the scaled variables over time:

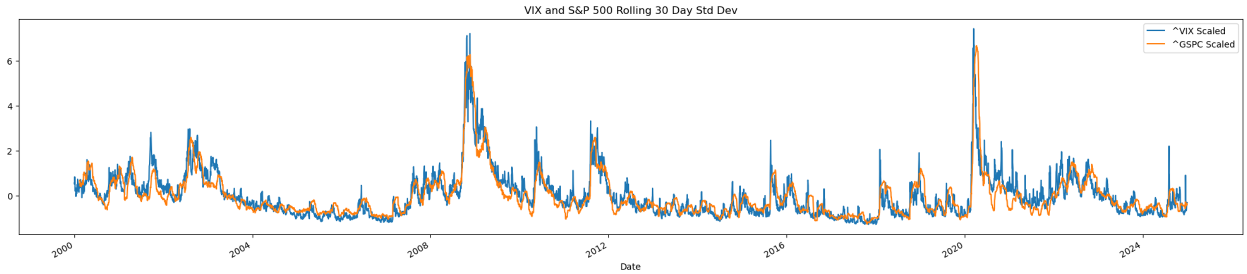

Figure 1: VIX Closing Price Scaled vs. S&P 500 30 Day Scaled Volatility

As we can see, the VIX moves very closely with the market's realized volatility. Logically, this isn't especially surprising, as options sellers begin demanding higher risk premiums during times of heightened volatility, even if volatility may decrease. Hence, the market's next 30-day volatility premium very closely resembles the most recent 30-day volatility, and typically overstates volatility observed 30-days later in the market. When we lag our VIX by 30-days and compare it with the market's realized volatility, this becomes quite apparent:

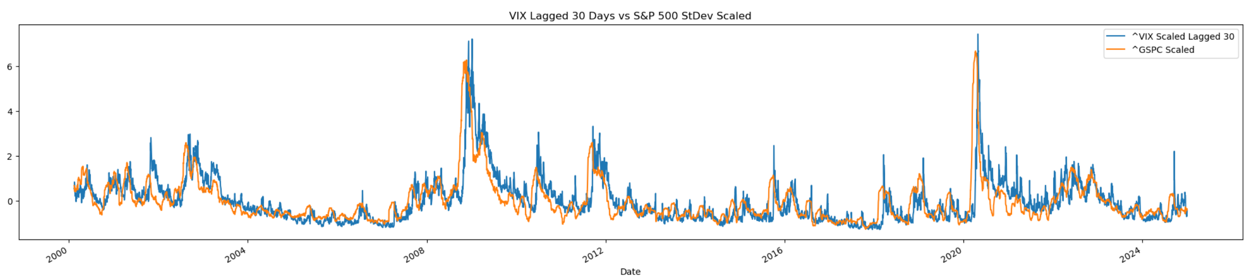

Figure 2: VIX Close Scaled with 30-day Lag vs. S&P 500 30 Day Scaled Volatility

While this graph also clearly demonstrates that trying to use VIX on its own as an indicator for market volatility would be foolish, it also shows an interesting interaction between implied and realized volatility. As we can see, the market often quickly decelerates from volatility peaks once the lagged VIX reaches it.

From here, I honed in on a clear objective: predicting large single day price ranges to formulate a trading strategy. I constructed a feature to track a scaled price range for each day, which then allowed me to derive a binary feature called 'Large Price Move', which was True for a specified level of price variation and False otherwise (true for roughly the top $\frac{1}{6}$ of market moves). As shown in the next graph, there certainly appears to be a correlation between these gaps in implied vs. realized volatility and next-day large price moves:

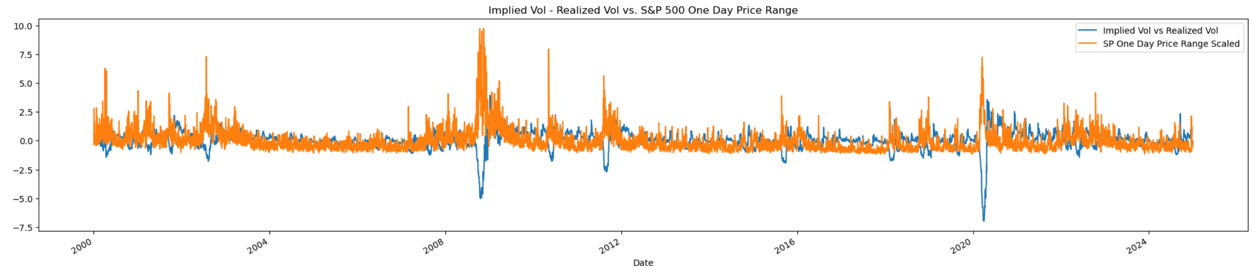

Figure 3: Difference in Implied and Realized Volatility vs. S&P 500 Scaled One Day Price Range

To analyze this trend and see if there is any potential for real-world applicability, I constructed a logistic regression model that uses the previous day's scaled closing price for the VIX, the difference between the prior day's implied and realized volatility, and the scaled S&P 500 trading volume to predict the binary 'Large Market Move' feature. I chose these features because they have moderate to strong correlations with the one day price range of the market, as seen in the correlation matrix below:

Figure 4: Correlation Matrix

Using an 80-20 test-train split (sequentially separated because we are dealing with time-series data), I yielded the following binary distributions:

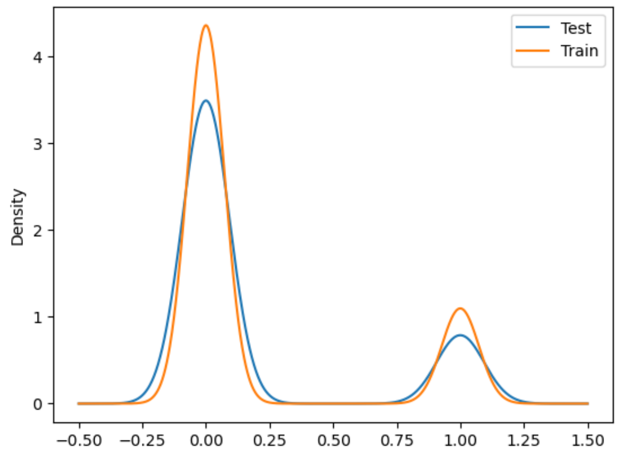

Figure 5: Test vs. Train Binary Distribution of 'Large' Market Moves

Because the goal is to make successful trades on large market moves, I weighted the model to punish Type 1 errors twice as much as Type 2 errors. This resulted in missing more large market moves, but significantly reducing the number of false positives. The test results were fairly promising, with 75.37% of predicted large market moves being true positives. Here's the full confusion matrix:

	Pred Neg	Pred Pos
True Neg	986	33
True Pos	129	101

Table 1: Test Confusion Matrix

Purchasing tight straddles on the S&P 500 for days when the model predicted large market moves certainly has the potential to be profitable. That said, without significant historical option chain data, there is no way to know for certain how well this strategy would perform. Additionally, I wanted to introduce our model to unseen data to see how well it would respond. Using stock data from 01-02-2025 to 04-10-2025, I generated the following confusion matrix:

	Pred Neg	Pred Pos
True Neg	19	0
True Pos	10	5

Table 2: Last 4 Months Confusion Matrix

While the model only accurately predicted $\frac{1}{3}$ of large market moves, there were zero false positives. This means that had we traded on the aforementioned straddle strategy, we never would have bought options on days when there weren't large market moves.

4 Conclusion

In this analysis, I demonstrated how the VIX more closely reflects current market volatility and risk premium, as opposed to an accurate prediction of market instability 30-days later. That said, looking at difference between observed and predicted volatility, as well as previous day VIX closing price, proved to be useful in accurately predicting days with large single day volatility. Finally, I proposed a strategy of using a long straddle option strategy on the market for days predicted to have high price ranges. Overall, I explored a unique financial instrument and its relation to the broader market, utilized data and machine learning to make market predictions, and proposed a potentially profitable trading strategy.