An Analysis of S&P 500 Index Prices Versus VIX Index Prices

The VIX Index has often been referenced by financial analysts during times of market downturn as a "fear gauge". Indeed, the index itself calculates a figure based on front and back month futures of the SPX futures contracts. Being weighted 2:1 in favor of put options, it stands to reason that a basket of such instruments would appreciate during times of market downturn and depreciate during times of market stability and growth. Yet still, I believe there are great misconceptions about the instrument itself that may be overlooked by new investors who may believe it to be a leveraged inverse ETF, as I used to do, but was proven incorrect through staggering losses from this misthinking.

It is my thesis that an evenly weighted (1:1) put/call ratio of the VIX would still see the index spike during market crashes and depreciation during periods of steady growth. Even a 2:1 weighted basket in favor of calls could possibly still hold the same relationship. This is because of the non-random distribution and trends of market patterns – "stairs up, elevator down" as is commonly said. Indeed, it is rarely the case even in the most bullish rallies where one may observe the S&P 500 Index appreciate by sizable double digit percentage gains, whereas they may often be seen on the reverse, when a market panic or crash occurs. As such, the idea of a gamma squeeze itself as it would relate to the index would typically only occur on the downside. It is therefore conceivable that a significant market decline as a result of fear or panic, as seen during the initial COVID panic and the 2008 Financial Crisis and smaller events occurring in between such as the VIX-pocalypse/Vol-mageddon of February 2018 and the Chinese market crash of August 2015, could result in such an extreme rise in implied volatility it could result in even a call-heavy basket of options appreciating. Conversely, a period of steady and expected decline as a result of changes to monetary policy have resulted in the same effect on the VIX as a period of steady growth – a continual depreciation of volatility ETFs and a period of low VIX prices, due to the "expected" price trajectory of the S&P 500 index itself.

This analysis will compare historical price relationships between the S&P 500 futures and VIX Index dating back to January 1st, 2000. It aims to study realized volatility in the period immediately after a significant rise in the VIX index, versus the behavior of the VIX during lengthy periods where market volatility remained limited. As the S&P futures trade in real time and the VIX index is calculated in real time, there should be no obvious relationship or evidence to suggest that one instruments behavior causes a change in the other, but instead, both instruments are correlated to each other in real time, but do not affect the other.

Files Used*:

spydataset2010.csv – historical records of SPDR S&P 500 ETF (date, price, volume)

vixdataset2010.csv – historical records of VIX Index Calculation (date, calculated price)

**Note that both datasets automatically omit weekends and American holidays where the instruments do not trade, so we do not need to filter them out during the cleaning process.*

Thesis

The volatility index (VIX) calculates anticipated short term (1-2 month) moves based on underlying option prices on the S&P 500 futures contract. While the VIX itself may be traded as its own futures contract, for the purpose of this analysis I will instead use a historical log of calculated VIX prices, as present day VIX-based ETFs have not yet been issued at the start of 2010, and those that do exist have been subject to contract rollover and decay, which is an unwanted friction for the purposes of this analysis. I will compare the VIX calculations against average S&P 500 prices represented by the SPDR S&P 500 ETF ($SPY) as well as the traded volume of the ETF. All data is retrieved from Yahoo! Finance.

Within this analysis I aim to establish the accuracy of statements such as "stairs up, elevator down" by comparing relative VIX prices during periods of market appreciation and decay. Additionally, I aim to compare S&P 500 volume against the VIX and the S&P 500 average price to determine if there is any meaningful correlation between the variables. It is my thesis that there will be significantly greater turnover of stock during periods of high volatility such as those of market crashes, due to elevated trading from market participants who may seek to use the opportunity to either sell and get out or buy in at a discount.

I aim to present my findings through visualizations built in PowerBI and Tableau, as well as present any quantitative findings made through data manipulation procedures in SQL. Upon completion, I aim to redo the data manipulation segment of this project using Python Pandas as an additional guide to completion.

Data Collection

Historical Data from 1/1/2010 – 2/8/2023 was extracted from Yahoo! Finance for the instruments 'SPY' and 'VIX' as csv files listed above. Each file itself contains a date, the opening and closing price, the high of day and low of day, and the volume (shares traded). Since the VIX is a calculation rather than a tradeable instrument, it lacks volume data. It is regrettable that the data does not contain the volume-weighted average price (VWAP), which would be a much more holistic indicator than the high/low/open/close in terms of representing the day's activity. As such, I decided to average the opening and closing price to create a new calculated column to represent the VWAP for analysis. The source data for the entirety of this analysis will come solely from the two extracted CSV files, which each include approximately 3300 days of data representing the total number of trading days during the aforementioned range.

Data Cleaning and Manipulation

I am fortunate that Yahoo! Finance keeps clean data records for the instruments I intend to analyze as the extracted .csv appears complete and accurate. There does not appear to be any additional need to validate, remove, or modify any of the source data at this time.

The first part of the manipulation procedure will result in four total columns: Date, SPY_Price, VIX_Price, and SPY_Volume. The SPY_Price and VIX_Price will be created by averaging the respective open and close price for each date. The dimensions of the resulting table will be approximately 3300x4, each record representing a trading day. The entirety of this process may be done in one SQL chunk as shown below:

```
SELECT [spydataset2010].[Date], ([SPY_Open] + [SPY_Close])/2 as SPY_PRICE,
([VIX_Open] + [VIX_Close])/2 as VIX_PRICE, [SPY_Volume]
INTO SPYVIX1 – Creates new table named SPYVIX1
FROM [spyvsvix].[dbo].[spydataset2010]
JOIN [spyvsvix].[dbo].[vixdataset2010]
ON [spydataset2010].[Date] = [vixdataset2010].[Date];
```

The second part of the manipulation procedure involves creating an additional calculated field named SPY_CHANGE_PCT to represent the percentage change in VIX_Price and SPY_Price from each previous observation (day). The first observation will therefore be zero, and each successive observation will be divided by a previous record. This can be done as follows:

```
SELECT *,
    CASE WHEN LAG(SPY_PRICE) OVER (ORDER BY date) = 0
         THEN 0
       ELSE ROUND(100*((SPY_PRICE / LAG(SPY_PRICE) OVER (ORDER BY
date))-1),3)
    END AS SPY_CHANGE_PCT,
        CASE WHEN LAG(VIX_PRICE) OVER (ORDER BY date) = 0
         THEN 0
       ELSE ROUND(100*((VIX_PRICE / LAG(VIX_PRICE) OVER (ORDER BY
date))-1),3)
    END AS VIX_CHANGE_PCT
FROM SPYVIX1;
```

The code above utilizes two case statements with the LAG function to access data from a different row as an input for a calculated column for a current row. The case statement is necessary to prevent a divide by zero error from the first row not having a row above to extract data from. As much as I'd like to say I'm an advanced SQL writer, the code above was produced largely using the help of ChatGPT as a coding assistant.

We may use the finished table above for part of the analysis. Additionally, I am interested in longer term trends; I will group rows by unique month/year, so there will be one observation for each month that has passed since 1/1/10. I will average the values of quantitative rows instead of averaging due to differences in business days per month, while also creating a new column (out of necessity) which tallies the total number of observations per grouped month as the column 'Business_Days'. Additionally, I will round the average SPY and VIX change column to 2 decimal places. I will save the resulting column to a new table, 'SPYVIX2'. This may be done using the following:

```
SELECT
  FORMAT(date, 'yyyy-MM') AS year_month,
  COUNT(*) AS Business_Days,
  AVG(SPY_Price) as AVG_SPY,
  AVG(VIX_Price) as AVG_VIX,
  ROUND(AVG(CAST(SPY_CHANGE_PCT AS decimal(10, 2))), 2) AS AVG_SPY_CHANGE_PCT,
  ROUND(AVG(CAST(VIX_CHANGE_PCT AS decimal(10, 2))), 2) AS AVG_VIX_CHANGE_PCT
INTO SPXVIXF2
FROM
  SPYVIXF1
GROUP BY
  FORMAT(date, 'yyyy-MM');
```

The two resulting columns were saved as tables under the names SPYVIXF1 and SPYVIXF2 respectively (cause I'm great at naming things, duh). These two cleaned and manipulated datasets will be directly used in the next step, the analysis phase. Note that the null values in the previous tables did not have an impact on the average change in the aggregated change value.

Data Analysis

During the preliminary stage of the data analysis process, I derived interesting statistics from the daily dataset that supported the statement "stairs up, elevator down"; the idea that the market moves upwards slowly but downwards quickly. Below are the queries and corresponding notes I made with regards to these statistics:

```
--1. Average Business Days in a month, over this period?
SELECT AVG(Business_Days) as AVG_BD_Per_Month
from SPXVIXF2; -- The resulting query indicates that in the long term there is
approximately 20 business days per month, which I found to be surprising.

--2. Average daily S&P 500 Percent growth during this period?
SELECT round(AVG(SPY_CHANGE_PCT),4) as AVG_DAILY_CHANGE
from SPYVIXF1;--The result, 0.0434%, shows just how powerful compounding can be!
1*1.0004^3300 = 3.74! At such a minute rate, the S&P nearly quadrupled over a 13 year
period.

--3. Average daily S&P 500 Decline, from the subset of red days and the number of red
days
SELECT round(AVG(SPY_CHANGE_PCT),4) as AVG_DAILY_CHANGE, count(SPY_CHANGE_PCT) as
Red_Days
from SPYVIXF1
where SPY_CHANGE_PCT < 0;--We see that there have been 1440 red days, with the average
declining day being -0.64% (which makes sense, given that we removed green days)

--4. Average daily S&P 500 Increase, from the subset of green days and the number of
green days
SELECT round(AVG(SPY_CHANGE_PCT),4) as AVG_DAILY_CHANGE, count(SPY_CHANGE_PCT) as
Green_Days
from SPYVIXF1
where SPY_CHANGE_PCT > 0;--We see that there have been 1851 green days, with the average
green day being +0.58%. This provides evidence supporting the statement "stairs up,
elevator down"

--5. Average daily VIX Change during S&P 500 Decline, from the subset of red days
SELECT round(AVG(VIX_CHANGE_PCT),4) as AVG_DAILY_CHANGE
from SPYVIXF1
```

```
where SPY_CHANGE_PCT < 0;--The vix on average appreciates 4.33% on a red day

--6. Average daily VIX Change during S&P 500 increase, from the subset of green days
SELECT round(AVG(VIX_CHANGE_PCT),4) as AVG_DAILY_CHANGE
from SPYVIXF1
where SPY_CHANGE_PCT > 0;--The VIX on average depreciates 3.04% on a red day

--7. Total VIX green days
SELECT count(VIX_CHANGE_PCT) as Green_Days
from SPYVIXF1
where VIX_CHANGE_PCT > 0;--We see that there have been 1523 green days, more than the
1440 figure of SPY red days (VIX and SPY have a generally inverse relationship). It is
likely that the VIX rises on certain SPY green days

--8. Total VIX red days
SELECT count(VIX_CHANGE_PCT) as Green_Days
from SPYVIXF1
where VIX_CHANGE_PCT < 0;--Similarly, we see that there have been 1770 red days, less
than the 1851 figure of SPY green days.

--9. Total VIX red days when market is down
SELECT count(VIX_CHANGE_PCT) as Red_Days
from SPYVIXF1
where VIX_CHANGE_PCT < 0 AND SPY_CHANGE_PCT < 0;--Very interestingly, there are 296
instances where the vix depreciates on a red SPY day

--10. Total VIX green days when market is up
SELECT count(VIX_CHANGE_PCT) as Red_Days
from SPYVIXF1
where VIX_CHANGE_PCT > 0 AND SPY_CHANGE_PCT > 0;--Just as interestingly, there are 378
instances where the VIX appreciates on a green SPY day

--11. Total VIX red days when market is up
SELECT count(VIX_CHANGE_PCT) as Red_Days
from SPYVIXF1
where VIX_CHANGE_PCT < 0 AND SPY_CHANGE_PCT > 0;--The VIX depreciates in 1471/1770
instances (83.1%) when the market is up

--12. Total VIX green days when market is down
SELECT count(VIX_CHANGE_PCT) as Red_Days
from SPYVIXF1
where VIX_CHANGE_PCT > 0 AND SPY_CHANGE_PCT < 0;-- The VIX appreciates 1142/1440
instances (79.3%) when the market is down. It is more likely for the VIX to fall on a
green day than for the VIX to fall on a red day.

--13. Average VIX Change
SELECT AVG(VIX_CHANGE_PCT) as AVG_VIX_CHANGE
from SPYVIXF1; --Surprisingly the average VIX change during this period is 0.18%, which
would have resulted in astronomical gains over the period, which did not occur. This is
likely the result of rebalancing the calculation to adjust for contract rollover.

--14. # of SPY days >+2%
SELECT COUNT(SPY_CHANGE_PCT)
from SPYVIXF1
where SPY_CHANGE_PCT >2; -- 56/1851 (3.0%) green days saw gains >2%

--15. # of SPY days <-2%
SELECT COUNT(SPY_CHANGE_PCT)
```

```sql
from SPYVIXF1
where SPY_CHANGE_PCT <-2 -- 65/1440 (4.5%) green days saw gains >2%

--16. Average movement where SPY days >+2%
SELECT AVG(SPY_CHANGE_PCT)
from SPYVIXF1
where SPY_CHANGE_PCT >2; -- The average movement given that the SPY is at least +2% is 2.83%

--17. Average movement where SPY days <-2%
SELECT AVG(SPY_CHANGE_PCT)
from SPYVIXF1
where SPY_CHANGE_PCT <-2; -- The average movement given that the SPY is at least -2% is 3.12%

--18. Worst 10 days of the stock market
SELECT TOP 11 SPY_CHANGE_PCT, Date, VIX_Change_PCT
from SPYVIXF1
order by SPY_CHANGE_PCT asc; -- Ignoring the null value, the top 3 worst days for the S&P 500 occurred within a week apart, during the initial COVID crash. Surprisingly the corresponding VIX value for the worst day is quite small.

--19. Best 10 days of the stock market
SELECT TOP 10 SPY_CHANGE_PCT, Date, VIX_Change_PCT
from SPYVIXF1
order by SPY_CHANGE_PCT desc; -- The VIX rose on three of the top 10 best days for the S&P 500. The greatest single day increase is 5.8%, during the COVID crash. 8 of the top 10 best days were during March 2020.
```

The above query results confirm much of my initial hypothesis about an imperfect inverted correlation between the VIX and SPY due to the structure of the volatility index; while 2:1 weighted on puts, a drastic unexpected expansion or deflation of implied volatility is oftentimes more impactful than the direction of the underlying itself. The VIX has a greater tendency to depreciate on a down day than appreciate on a green day likely due to the payoff structure of the options it calculates; by depreciating in value during a majority of periods where volatility is low, it disproportionately appreciates during panics, capturing the tail end of returns during such periods.

We observe from the above results that it is more often for a given day at random to close green rather than red from the previous day (1851 vs 1440), but the magnitude of the average upward movement (0.58%) is less than the magnitude of the average downward movement (0.64%). Combining the two, the S&P 500 index appreciates an average of 0.0434% per trading day for the last 13 years, which seems insignificant, but resulted in the index almost quadrupling within this period. On a similar note, we observe the magnitude of the top 10 worst red days to be significantly greater than the top 10 green days (-5.9% vs +4.3%).

It is interesting to me the extent to which the expected relationship between the VIX and S&P 500 index is weaker than expected. The two move in opposite directions only approximately 81% of the time, despite the VIX being a heavily weighted put index. Additionally, the VIX is more likely to depreciate on a red day (21%) than appreciate on a green day (17%), which highlights the tail-end distribution of gamma; the VIX can explode in periods of extreme movements (even upwards), while it is more likely to depreciate when the market is stable, even if it slightly depreciates. In an extreme hypothetical scenario, consider the following: overnight, the S&P 500 index doubles. How would the VIX react? While the 2:1 weighted puts may approach 0, the calls are likely to increase by some extreme magnitude, such as 50x, which would altogether result in the vix increasing 1666%. While less extreme, a smaller (4-5%) move upward would also observe a large increase in the VIX if it was totally unexpected.

Conversely, suppose the S&P 500, after a period of explosive growth as hypothesized above, begins to see a period of market calmness, a slight daily depreciation of say -0.10% over an extended period. In such a scenario, the calls would naturally depreciate, but the puts would do so as well as the market, perhaps expecting a greater average daily movement, sees a significant repricing of the implied volatility of all options; the fall in implied volatility far exceeds any delta gains from a minute -0.10% move, resulting in both the calls and puts depreciating, ultimately resulting in a loss for the VIX even in an environment that sees the S&P 500 index depreciating. While such scenarios do not appear to occur often or frequently, the presence of one VIX-depreciating days in the top 10 worst days of the S&P 500 index and three VIX-appreciating days in the top 10 best days of the S&P 500 index supports the idea that changes in underlying implied volatility is oftentimes much greater an impact on underlying volatility prices than changes in delta, even in extreme circumstances such as day of major S&P 500 gains or losses.

Summary

The S&P 500 Index has a moderately strong but not perfect correlation when comparing the direction of the move alone against the VIX. Due to the nature of the VIX capturing changes in implied volatility, it is oftentimes the case that a sharp upwards move may result in positive VIX changes and a below expected move downward may result in a depreciation in the VIX. Furthermore, it is important to consider the current level of the VIX when analyzing how a change in the underlying S&P 500 index may affect volatility, as can be observed in volatility drops during days of significant index declines and an increase in volatility during days of market appreciation.

It is not advised to enter a position on a volatility instrument on the basis of speculating on the direction of the S&P 500 alone without regard for expected changes in volatility. Similarly however, it may not be ideal to purchase a volatility-linked product if one expects a sharp increase in expected volatility from an upwards movement due to the put-biased nature of the index. The investor may instead decide to purchase a leveraged ETF, futures on the underlying, or enter a delta positive options position.