VIX – Capstone Project Status Update
November 21, 2017

Name: David Hamilton

Email: deh284@nyu.edu

Research Question

The VIX index estimates expected volatility on the S&P 500 stock index. It does so by averaging the weighted prices of options on the S&P (puts and calls) over a wide range of strike prices. There are extremely liquid, exchange-based futures contracts based on the price of VIX as a means of trading and hedging said volatility.[1]

This project proposes to answer both a primary and secondary research question:

1. Primary: ***Is there potential for a VIX futures/SPX options arbitrage?***
   Initial research will attempt to determine the possible existence of a short-term disconnect between the price of VIX futures contracts and the composite value of the underlying S&P options on which they are calculated. The key issues which need to be resolved are:
   - Does such a short-term disconnect actually occur?
   - If the arbitrage exists, how often does it happen? How much, on average, does the price relationship get "out of line" and for how long does it last?
   - Are there specific times or market conditions when the disconnect is more prevalent?

2. Secondary: ***Is there a viable solution to capture the arbitrage?***
   Should the initial study discover meaningful inefficiency in the VIX/S&P relationship, the focus will shift to investigating and, if possible, designing a realistic, actionable strategy that can profit from the potential arbitrage. The single largest problem with any proposed implementation will be how to handle the SPX options leg. Trading a weighted strip of contracts from every strike price used to calculate the VIX is too cost prohibitive and

[1] VIX White Paper, Chicago Board Options Exchange (CBOE); 2015

therefore unrealistic. As a result, this secondary phase will most likely revolve around devising a proxy SPX solution. Key elements to consider:

- Is there a weighted subset of key strike prices (say 4 or 5 instead of 100+) that can be used as an approximation for the SPX leg?
- Can a loss function be applied to the optimization process? If so, multiple machine learning techniques could be deployed towards calculating a proxy solution.
- Can the intersection of transaction costs versus replication accuracy be quantified and examined for each possible solution? An example would be to apply Modern Portfolio Theory and evaluate projected relative performance against variance of the results.
- Is there an alternative instrument (or instruments), such as SPY options, that could be used more effectively as a proxy?

Data

The project dataset has been (and will continue to be) compiled from multiple sources. It currently contains end-of-day prices and supporting information from 2005 through mid-2016 on the following primary instruments:

- **Short-Term VIX Futures**: Pulled directly from Quandl via the provider's Python API. It was originally compiled by the Chicago Futures Exchange (CFE), the division of CBOE where VIX futures are listed. Individual contract prices have been parsed together to create a continuous time series that can be queried by a simple maturity designation (1-month, 2-month etc).
- **Short-Term SPX Options**: Downloaded from Wharton Research Data Services (WRDS). Like most blocks of historical options data, it is large in size (approx. 5 million observations) and required a great deal of cleaning and processing before being ready for use.

Intraday prices for futures and options, as well as other supporting instruments will be added shortly to complete the dataset. A two-step approach was necessary in order to ensure that key project functions and code could handle the unique challenges that accompany this type of

information: misaligned expiration dates between products, unexpected exchange holidays and multiple and detailed calculations required for each date observance (to name a few).

Methodology

Project implementation and analysis is also being conducted in two parts. The first phase has focused on building a one-dimensional time series using the lead VIX futures contract and single, at-the-money SPX options. Its purpose is to establish a simple, baseline metric that demonstrates their price relationship and examines the opportunity for potential arbitrage.

We need to take the following into consideration regarding each component leg:

- VIX futures are contracts on forward 30-day implied volatilities.[2] Their fair value is derived by pricing the forward 30-day variance that underlies their own settlement values.[3]
- The forward price of 30-day variance can also be found by taking methods used to calculate the VIX index (spot VIX) and applying them to a synthetic calendar spread of S&P 500 options bracketing the 30 days after the VIX futures expiration.[4]

This initial approximation can now be calculated as:

$$VIX_{1\to2}^2 = \frac{\sigma_2^2(T_2 - T_0) - \sigma_1^2(T_1 - T_0)}{(T_2 - T_1)}$$

where:

$VIX_{1\to2}^2$ = squared price of VIX future expiring on date T_1

σ_2^2, σ_1^2 = squared levels of implied volatility for at-the-money (ATM) SPX options expiring on T_2, T_1, respectively

[2] CBOE Futures Exchange (CFE), http://cfe.cboe.com/cfe-education/cboe-volatility-index-vx-futures/vix-primer/the-basics

[3, 4] CBOE Futures Exchange (CFE), http://cfe.cboe.com/cfe-education/cboe-volatility-index-vx-futures/vix-primer/vix-features

The second phase will attempt to build on the baseline by employing more precise methods to price VIX futures:

$$VIX_{1\to2} = \sqrt{\left[P_t - \hat{\sigma}^2_{VIX_{t\to2}}\right]}$$

where:[5]

$VIX_{1\to2}$ = price of VIX future expiring on date T_1

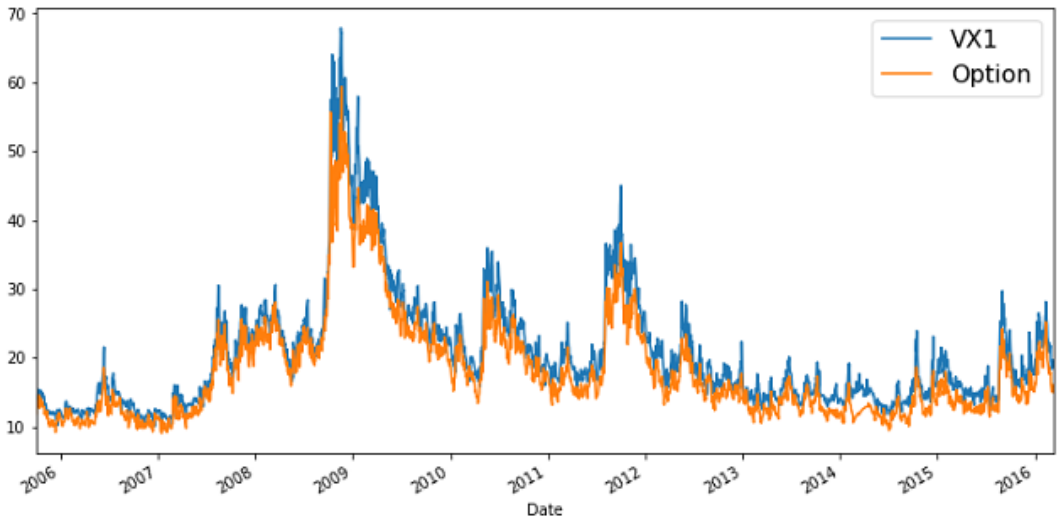
P_t = portfolio of SPX options with long, out-of-the-money positions expiring on date T_2 and short, out of-the-money positions expiring on date T_1

$\hat{\sigma}^2_{VIX_{t\to2}}$ = estimate of cumulative variance for $VIX_{1\to2}$ between T_0 and T_2

The use of long and short strips of SPX options with P_t in lieu of individual at-the-money contracts opens the possibility for optimized solutions to the static hedging of VIX futures and, hopefully, dynamic arbitrage capture. Application of various learning methods to the cumulative variance estimate could also be worth investigation, providing insight on the possible future behavior of $\hat{\sigma}^2_{VIX_{t\to2}}$, something that is extremely difficult to predict.

Results

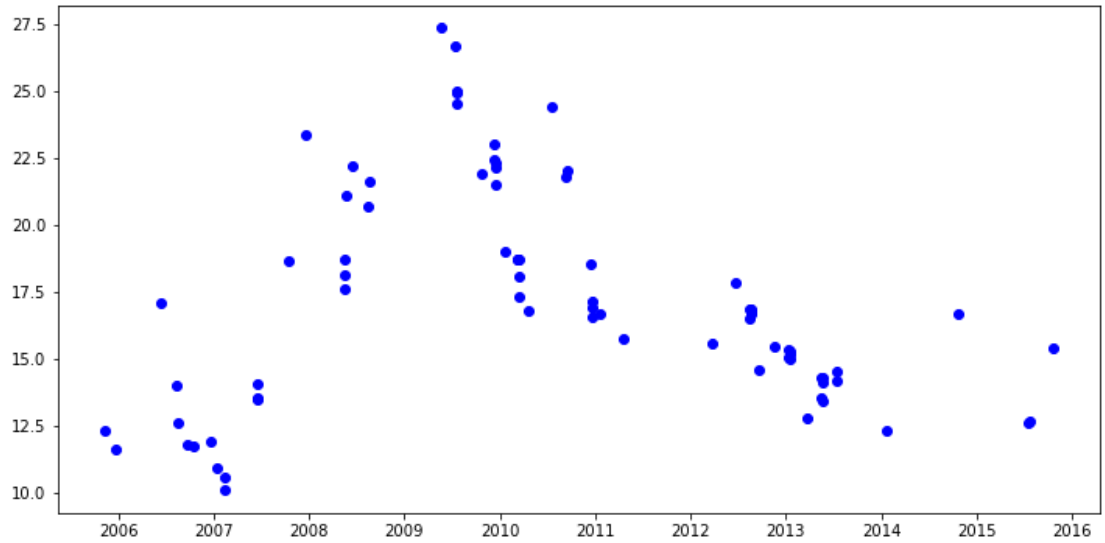
The baseline time series from Phase 1 indicated a reasonably steady relationship between the lead VIX futures contract and the at-the-money implied volatilities of the surrounding SPX options:



[5] "The VIX-VIX Futures Puzzle", Asensio, Ivan Oscar (https://www.uvic.ca/socialsciences/economics/assets/docs/seminars/Asensio.pdf)

The slight premium of the VIX closing price (median daily spread over options = 2.24) can be attributed to the absence of out-of-the-money (OTM) SPX put options, and the higher implied volatility at which they normally trade, from the calculation of the option calendar spread. This was an intentional omission to simplify the initial baseline time series. Phase 2 will employ a more detailed method which will include OTM options to better approximate the convexity element found in the 30-day forward starting variance that drives pricing for VIX futures.

Despite this stable VIX-Options pairing over a period of roughly eleven years, there are still 77 distinct instances (roughly 3% of daily observances) of a "disconnect" where end-of-day prices finish "inverted", with the value of ATM implied vols settling above that of the futures contract:



Closer examination of cases where the relationship traded "out of line" shows a majority occurred when the implied option leg was between 13 and 21.25, well within the "normal" historical price range for VIX and its near-term futures contracts:

	OPT	VX1-OPT_Spread
count	77.000000	77.000000
mean	17.170416	-0.725481
std	4.104026	0.617556
min	10.139066	-3.977454
25%	14.153389	-1.058194
50%	16.725489	-0.573158
75%	19.016475	-0.320114
max	27.360114	-0.002177

Discussion/Problems to Address going forward

The second phase, as previously mentioned, will now add intraday historical prices to the dataset. As this new information will be provided by a different vendor than the end-of-day figures, consistency needs to be verified between the two time series before building upon the original baseline study. The higher frequency data can then be studied to see if there are additional intraday "disconnects" in the pricing relationship between VIX futures and the SPX options leg.

Once all said instances have been identified, unsupervised learning techniques for dimensionality reduction (PCA, SVD) should be able to provide initial indications of the best individual options to approximate the SPX leg, as well as possibly predict future movements in the VIX/SPX spread. It will then be possible to apply more advanced methods for optimization (regression, classification) to determine the final recommended solution(s).

Next week's status update (Nov. 30) should include details regarding the addition of intraday data and early results from the modeling process.