VIX – An NYU CDS Capstone Project Proposal

Name: David Hamilton

Email: deh284@nyu.edu

I have over 15 years of institutional options trading experience on both the Buy and Sell Side in the US and Europe. Specific areas of focus in my career have included proprietary quantitative strategies (dispersion/correlation, relative value), market making and client facilitation.

This project will attempt to examine a highly visible, yet often poorly understood derivatives product using key elements of the CDS Masters curriculum including big data and machine/deep learning. Since it focuses on a relatively young, rapidly expanding area of the market, there is exciting potential for gaining valuable new insight(s) from research driven by data science.

What is the project?

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 has both a primary and secondary goal:

1. **Identify the potential of 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?

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

- 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. **Determine if there is 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 therefore unrealistic. *That means the project's second phase will 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 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.
   - Can a loss function be applied to the optimization process? If so, multiple machine learning techniques could be deployed towards calculating a proxy solution.
   - Is there an alternative instrument (or instruments), such as SPY options, that could be used more effectively as a proxy?

What does the data look like?

A detailed description of the type of data that is required to address the problem. For example, will the students primarily be working with data aggregated from twitter feeds or other social networks, is it medical data, biological data, or financial data? Along with detailed information on the type of data there should also be context regarding where the data can be found. Will the organization provide the majority of the data? Is the data accessible via other avenues/ sources? Is an NDA required? How much of the data is available? Do the students need to gather data, model data, both, or neither?

The data required will come in the form of Trade and Quote (TAQ) datasets for all VIX futures and S&P options contracts to be used in the study. Ideally, these should reflect securities prices at the tick-data level. If that degree of granularity is not practicable, then the smallest possible increment of time (1-second updates, 5-second updates) that fits within the scope of the project should be used. The requisite datasets already exist and can ideally be procured from various resources within NYU. Otherwise, there are a number of external entities and financial data vendors who, given the nature of the project, could potentially provide the information for little or no charge.

What is the proposed scope of the project?

Is this a project for 1, 2 or 4 people? Is this a smaller well-delimited problem or the next step in addressing a grand challenge in your field? How many hours of work are anticipated for completion of the project?

Ideally this would be a project for 1 or 2 people. Its stated goal attempts to address two questions facing the industry: does the market currently price/trade VIX futures (and similar products) in the best way possible and, if not, are there new opportunities and potential for profit by taking a different approach? The proper investigation of both issues should require a minimum of 150-200 hours of work (75-100 hours/person in the case of 2 people).

What are the rubrics of success?

The Center for Data Science will be grading all of the final projects. In order to do this, we will need information from the stakeholders around what progress/ success would look like. For example, is progress classified as the creation of an algorithm that improves upon the current process? Each problem must be accompanied by a detailed description of the rubrics of success.

In other words, how will this project be judged and graded? What are the quantitative and qualitative metrics that can be used as metrics for successful completion of the capstone project. Since each problem would ideally propose a different set of solutions, these rubrics must be specific to each individual problem.

Successful completion for the initial phase of the project should be fairly easy to define. If the necessary data is compiled and researched in the proper manner, it should be readily apparent if there is a quantifiable disconnect that occurs in the VIX/S&P relationship and, if so, how often does it happen and to what degree (on average) when it does?

The second part, assuming the discovery of a tangible inefficiency, will be harder to evaluate. Since the only true test of success for a trading strategy is to demonstrate profits in live market conditions, any measures used to calculate the efficacy of project-generated ideas will be, at best, an estimate. The key will be in identifying how accurately real-life conditions have been simulated. In the instance of possible arbitrage profits, additional qualitative evaluation can be performed on the methodology used (probability & statistical analysis, machine learning, signal processing etc.) to devise the potential trading solution(s).

VIX – Capstone Project Update Memo
November 9, 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 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.
- Can a loss function be applied to the optimization process? If so, multiple machine learning techniques could be deployed towards calculating a proxy solution.
- 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, currently nearing completion, 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:

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

where:

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

σ_2^2, σ_1^2 = squared levels of implied volatility for 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:

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

where:[5]

$VIX_{1\to 2}$ = 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

$\hat{\sigma}^2_{VIX_{t\to 2}}$ = estimate of cumulative variance for $VIX_{1\to 2}$ 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 $\hat{\sigma}^2_{VIX_{t\to 2}}$, something that is extremely difficult to predict.

Results

Phase 1 of the implementation is nearly completed and results from the baseline time series should be available within the next day or two. Initial data prep and coding took longer than expected, primarily due to challenges from the interaction between VIX and SPX products. The solutions were often difficult and time-consuming, even by options standards.

Discussion/Problems to Address going forward

As always, the biggest potential hurdle moving forward will be the addition of intraday historical prices to the dataset. There is reason to believe that the second round of cleaning/processing should be smoother, thanks to early lessons learned and the fact that the new material will likely becoming from a well-established provider/custodian. Special thanks are due to Marco Avellaneda at Courant, whose experience and expertise with VIX-related products, as well as his generosity in sharing both, has proven invaluable to date.

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