Distribution of Risk and Return in Variations of Volatility Arbitrage

Maksym Kosachevskyy

Professor Jia Li, Faculty Advisor

Professor Kent Kimbrough, Seminar Instructor

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University

Durham, North Carolina

2021

Acknowledgements

I would like to thank Professor Kent Kimbrough for his guidance and support throughout this project and for connecting me with the resources without which this thesis would not be possible. Next, I would like to thank Professor Jia Li for setting me on the right path and providing his expertise on the subject matter at every step of the way. I also want to thank Alec Ashforth for keeping a keen eye on my project and keeping me in check. Finally, I would like to thank John Little for helping me with R and wrangling through my code.

Abstract

The effectiveness of volatility arbitrage has been a source of debate for researchers. On one hand, some have found the strategy to be immensely profitable, indicating a potential structural mispricing in the options market. Other researchers have claimed these profits arise from hidden risk in the form of higher distribution moments like kurtosis and skewness or that the strategy is highly susceptible to jump risk. In this paper, I examine the risk and return of a set of options volatility arbitrage strategies over the last 6 years to determine the magnitude of a possible mispricing. I construct a portfolio of long straddles using the options in the decile with the greatest positive IV-HV difference and a portfolio of short straddles using the options in the decile with the greatest negative difference. I then calculate the Compound Annual Growth Rate and standard deviation of monthly and weekly strategies, find the optimal Sharpe ratio, and adjust for potential liquidity issues. I find that the combined monthly portfolio can be a strong performer if properly hedged but that only the long portfolio is necessary in the weekly strategy. Both weekly and monthly portfolios can highly effective investments if risk is managed correctly.

JEL classification: G11, G13, G14

I. Introduction

Options pricing theory in academic literature often relies on a broad assumption of informational efficiency resulting in an arbitrage-free market. Under this assumption, prices of options instantly reflect updated information, preventing any a strategy trading on this information to achieve excess return above the market. This belief is what has allowed investors to hedge directional risk, one-sided exposure to an asset, and volatility risk, an increase or decrease in the overall pattern of price movements, across asset classes (Jarrow, 2013). However, as tested by Cao & Han (2013), a "no-arbitrage approach [between options and the underlying] can only establish very wide bounds on equilibrium option prices" (p. 231). In addition,

One unique feature of options markets is a concept called implied volatility. Implied volatility is a metric that indicates the market sentiment of the stock's expected absolute price changes over the maturity of the option. It is not directly observed but derived through a pricing model such as Black-Scholes or the Binomial Model. One trading strategy known as volatility arbitrage rests on the assumption that implied volatility is not an accurate indicator of the actual volatility of the stock. Implied volatility should increase or decrease to match historical volatility levels, and this mispricing can be captured. On one hand, some researchers have found volatility arbitrage between implied and historical volatility to be immensely profitable (Goyal & Saretto, 2009), potentially challenging an efficient market hypothesis implied by pricing theory (Jarrow, 2013). Other researchers, on the other hand, attribute volatility arbitrage profits to risk to unaccounted for in the form of higher distribution moments like kurtosis and skewness (McGee & McGroarty, 2017). The distribution of returns in a volatility arbitrage strategy is oftentimes leptokurtic with fat tails – the majority of returns are small and positive, but there is a significantly higher than normal chance that a return will be many standard deviations away from the mean, especially on the left. This is due to unforeseeable risks such as jump tail risk and black swan events. In fact, according to Bollerslev and Todorov (2011) who decomposed the risk premium into its constituent parts, “on average more than half of the historically observed variance risk premium is directly attributable to disaster risk” (p. 2167) and “on average close to 5% of the equity premium (in absolute terms) may be attributed to the compensation for rare disaster events” (p. 2167). One possibility is that volatility arbitrage is profitable and predictable for some period of time but unprofitable in the long-run due to these major periodic losses.

From my experience trading options spreads, volatility does not appear to be priced correctly for numerous equity options, especially those with lower volumes and coverage by analysts. The observations I made while trading have led me to believe that for certain stocks there exist options with implied volatility that is not only higher than recent historical or future realized volatility – a bias that has been confirmed by previous research – but is even higher than when adjusting for foreseeable risks.

Before starting this research, I wrote an algorithm to search and identify individual stocks with a volatility-bearish weekly options spread that is priced higher than would be expected given recent levels of historical volatility. I calculated the average 5-day price movement of a stock over 130 trading days, attempting to control for the execution date of an order. I then calculated the price of placing a market order for a long condor spread with calls expiring at the end of the trading week assuming poor liquidity for the spread: taking the ask when buying a long call and taking the bid when buying a short call. The inner legs of the condor would be the first available strike prices that were outside the estimated level of historical volatility (below for the smaller short call and above for the larger short call). I then calculated the maximum potential profit for the condor spread, filtering to only those spreads with positive profit potentials. Then, assuming the estimated five-day level of volatility in the upcoming five trading days, I identified the stocks with the largest "breathing room" to the breakeven point closest to their current stock price. Finally, I eliminated tickers with possible catalysts of volatility spikes within the expiry of the option, namely earnings reports, corporate announcements, and other announced events that historically have affected the price of stocks.

Despite more risk-averse trading practices – not competing on the spread of the option, using a condor as opposed to a butterfly or a short straddle, picking spreads with a "breathing room" that is approximately one-half standard deviation of the average volatility, and avoiding stocks with high or unpredictable idiosyncratic volatility – about 5-10 equities would still have a condor spread with a net premium higher than would have been predicted by the historical volatility. This strategy is one variant of the volatility arbitrage noted by Goyal & Saretto (2009). The similarity in our average performance despite variations in set-up, assumptions, and time period has intrigued me to understand these inefficiencies within options pricing, notably the risk factors contributing to this disparity, how the aforementioned strategy compares to other short volatility arbitrage strategies, the pricing of volatility between systematic and idiosyncratic risk, and how short volatility arbitrage performs in various market conditions. However, given the risks associated with a fat tail distribution, I do not believe there is a risk-free way to exploit this supposed inefficiency. Rather, I will distinguish variations of volatility arbitrage by their risk/reward profile. In this paper, I explore variations upon the volatility arbitrage strategy executed by Goyal & Saretto (2009) by changing variables such as maturity, liquidity, and hedge ratio. I then characterize distributions of their returns by consistency and magnitude of positive and negative returns through a set of statistical and financial pricing models, especially noting the strategies' resistance to unforeseeable black swan events. No other research that I have found has examined the impact of volatility arbitrage strategies since 2009. Furthermore, little research has been conducted looking at volatility arbitrage using weeklies. Finally, there has been almost no research to characterize different forms of volatility arbitrage beyond the variation in technique employed (straddles and delta-hedging).

First, I used data obtained from OptionMetrics over the time period January 1, 2015 to December 31, 2020. This range was chosen partly because I wanted to use recent data that had not been previously analyzed in other research papers and that covered the COVID stock market crash. A larger time scale of 10+ years would have been preferred, but I was limited by the computing power available to me and the large size of the data I was working with.

The baseline strategy is to calculate the difference between HV and IV for all options with monthly expirations at the start of the trading month, rank them into deciles, and then create a portfolio that trades long straddles with the largest positive differences and shorts straddles with the largest negative differences. This baseline strategy will receive a significant portion of my attention in this paper. I test a multipart hypothesis in that the following modifications will result in a significantly different stream of returns and create different risk/reward profiles from the baseline strategy.

- Stemming from Andersen et al (2016), I believe a weekly options strategy with ATM options compared to a monthly ATM options will be more effective in isolating volatility risk from jump risk. As a result, volatility arbitrage will produce more consistent returns, although its resistance to tail risk is subject to examination.
- Stemming from Cao and Han (2013), I believe lower liquidity (large bid/ask spread) is more effective in a short vol strategy but higher liquidity (small big/ask) is more effective

in a long vol strategy. Implementing specific liquidity requirements to the long and short side will result in higher returns with fewer drawdowns.

- Given my own experience, I believe a short vol strategy is more effective than a long vol strategy. A smaller proportion long hedge (not 50-50) would be more effective than a 50-50 hedge, resulting in higher returns but likely more significant drawdowns.

I will compare and contrast the average return, CAGR, standard deviation, distribution of returns, skewness and kurtosis, expected return using CAPM, and the Sharpe ratio when comparing performance of these strategies.

II. Literature Review

Past research on volatility arbitrage using equity options has been too assumption-heavy, either noting that options are simply levered equities, that no arbitrage opportunities exist between equities and options, or that no arbitrage exists between implied and historic volatility, that it is close to an efficient market. Only in the last decade have researchers begun to identify options as an asset class with distinct features and have begun to test and challenge some of these theoretical assumptions.

Goyal and Saretto (2009) identify the volatility arbitrage strategy that is most similar to my own interest. They rank all options by the difference in historical volatility as calculated over the last twelve months (LTM) and implied volatility from the option's price, then trade a long-short vol approach through two portfolios of straddles, long in the decile with the highest positive difference and short in the decile with the largest negative difference. They find that the approach is profitable, noting a 22.7% average monthly return. The researchers hypothesize that investor overreaction to recent increases in volatility are the causes for an overpricing. They key finding of a statistically significant and highly profitable arbitrage between implied and historical volatility serves as an important basis for my own research.

However, there are a few significant assumptions that I would like to challenge and set-up behaviors I would like to change. For one, their paper portfolio of trades uses the mid-price of the bid and ask rather than the spread as given. Not competing on spread returned only 3.9% monthly. Because the size of the bid-ask spread is a measure of liquidity, having poor liquidity could create unfavorable buying and selling prices which would break the long-term profit of a portfolio. Furthermore, from my experience, the stock candidates for a short vol strategy have been far more numerous than in a long vol strategy. In fact, I believe the principle of mean reversion applies with greater significance to a short vol strategy since it profits from decreases in volatility over the duration of the contract while a long vol straddle will lose money from natural decreases in vol. I believe the greatest volatility changes (and overreactions) are in the upward direction, whereas decreases are more gradual. As a result, I would like to look at the short vol portfolio independently or with greater weighting. Most significantly, the researchers used options data in a time when options were not as common or utilized as they are in the present day. Examining trends in performance during more recent periods is necessary to gauge performance in the long-term, especially to examine the effects of this strategy after the publication of the paper.

McGee and McGroarty (2017) confirmed that there does exist an upward bias when estimating future realized volatility as a linear regression of implied volatility, often described as a risk premium. However, they claim that an OLS estimate assumes that the cost of volatility forecast errors is symmetrical and ignores the impact of higher moments of the return distribution such as skewness and kurtosis. Volatility arbitrage strategies through delta hedging do not account for these higher moments, which may be a source of the unpriced risk. The researchers created a framework for pricing in these higher moments and concluded that due to the heavy costs of tail events, a volatility arbitrage strategy will result in negative expected growth. “Despite its upward bias, the market pricing of implied volatility is efficient to the extent that trading the upward bias does not generate a long-term return premium over the period of the study [between Jan 1996 and May 2013] (McGee & McGroarty, 2017, p. 13). The strategy employed by Goyal & Saretto (2009) is delta-neutral and uses these moments as controls but does not adjust for these higher moments, which could be a factor in the price of volatility. However, McGee and McGroarty trade the same “static” portfolio of options for the duration of the study, studying the continuous upward bias of implied volatility for the portfolio for a set of predefined stocks. On the other hand, Goyal & Saretto (2009) find different candidates for volatility arbitrage at the start of every trading period, not being subject to systematic overpricing, rather looking for idiosyncratic mispricing. This may affect the type of risk factors influencing arbitrage candidates.

Cao and Han (2013) note other relevant findings. “On average, delta-hedged options have negative returns, especially when the underlying stocks have high idiosyncratic volatility” (p. 232). In fact, a portfolio that bought delta-hedged calls from the bottom quintile of idiosyncratic volatility and sold delta-hedged calls from the top quintile returned 1.4% per month in their study. Furthermore, “average delta-hedged option return is significantly more negative when the underlying stocks or the options are less liquid and when the option open interests are higher. These results are consistent with option dealers charging a higher option premium when the options are more difficult to hedge and option demands are higher” (p. 232). What this signifies is that non-directional exposure to options is negatively correlated with the idiosyncratic volatility of the underlying and is an opportunity for long-short arbitrage, regardless of being a put or call. Less liquidity lowers returns for a long strategy but, although it is not directly stated, may increase returns for a short strategy since a short seller would be requiring an additional premium to provide liquidity and hedge their own position. This is, in fact, corroborated by more recent research. “The risk-adjusted return spread for illiquid over liquid equity options is 3.4% per day for at-the-money calls and 2.5% for at-the-money puts” (Christoffersen et al, 2017). For this reason, it may be worthwhile to pay attention to the performance of the long side and short side of a vol arb portfolio separately and test the hypothesis with various volatility arbitrage spreads that use a different combination of long/short and puts/calls. Cao and Han touch on the possibility of volatility mispricing as the reason for the bias against idiosyncratic volatility but dismiss these variables after the results remain as they after controlling for a few volatility mispricing hypotheses laid out in other papers – market overreactions and large disparities between implied and historical volatility.

Another important consideration for choosing which options to select for a volatility arbitrage strategy is the stability of a stock’s volatility over time. An unpredictable volatility is one that cannot be explained and that will not follow a pattern, even if one across its broader category of equities. Ruan (2020) explores the impact of volatility-of-volatility on option equity returns, finding an economically and statistically significant negative relationship between delta-hedged option returns and the volatility-of-volatility. VOV is not a well-studied unit of measure so this is a newly identified relationship. With a portfolio construction approach similar to Goyal & Saretto (2009) and Cao & Han (2013), Ruan (2020) ranks equities into quintiles based on their VOV, as calculated by standard deviation of implied volatility / mean of implied volatility. He finds that a portfolio of options held until maturity (on average 50 days) that shorts the quintile with highest VOV and longs the quintile with the lowest VOV makes 0.16% on average between January 1996 and April 2016. Practically speaking, one potential variation upon the volatility arbitrage strategy would be to limit the top decile of the long candidates ranked by difference between historical and implied volatility to those also in the bottom quintile of VOV and the short candidates to those also in the top quintile of VOV. This could potentially help adjust for the set of stocks with little systematic pattern in volatility. Due to the already large scope of this project, this strategy will have to be sidelined for future research.

Other variables to consider in the risk/reward profile of a vol arb strategy is the maturity and strike of the contract. Longer maturities and options closer to being ATM are more sensitive to changes in volatility, while shorter maturities and OTM options are less reflective of volatility. Andersen et al (2016) study the characteristics of the increasingly prevalent "weeklies" – the shortest-dated instruments with a week-long maturity. They note that "short-maturity ATM options help pin down spot volatility, while the relative prices of deep OTM options assist in determining the intensity and distribution of jumps" (p.1336). The researchers take a semi-nonparametric approach to pricing options in order to take advantage of all the information on the option's surface but not be subject to minor misspecification errors, which can result in significant unpredictable effects. By specifically looking at how these shorter ATM maturities react to vol and jump events compared to longer and more OTM ones, the researchers identify a measure to price negative jump tail risk independent of volatility risk and create a way to easily identify "periods of heightened concerns about negative tail events that are not always 'signaled' by the level of market volatility and elude standard asset pricing models."

For the purposes of studying IV-HV arbitrage, isolating implied volatility while predicting and avoiding tail risks like the ones mentioned by in Bollerslev and Todorov (2011) would be the optimal approach, although the method proposed in the paper would not feasible for the function of my study due to the complexity of implementing parametric modeling to the entire range of stock options across decades. Rather, studying the effects of short tenor ATM options as a proxy for a cleaner volatility estimate and the effects of short tenor OTM options as a proxy for jump risk would be one possible approach to varying the risk/return profile. For instance, comparing the performance of a strangle to a straddle in situations of systematic volatility change versus situations of tail end risk would be an interesting implementation of these findings.

Park (2015) proposes an alternative method of measuring tail risk that does not follow the standard jump process framework that we see in Bollerslev and Todorov (2011) or Andersen et al (2016). It is not as sensitive to misspecification errors that can occur on an option's surface and does not require the complexity of a semi-nonparametric approach of Andersen et al (2016). He uses volatility-of-volatility as a measure of tail risk because "even a small change in the variable has a critical influence on the tails of return distributions" (p. 39). Park uses the VVIX index, which is a risk-neutral expectation of volatility of the 30-day forward VIX index. He finds that "a higher level of tail risk increases the current prices of tail risk hedges, lowering their subsequent returns over the next period" and that "the VVIX index is predictive of tail risk hedging returns with a negative sign over the next three to four weeks, implying that the tail risk hedging options become more expensive when the VVIX index is high" (p. 39). In the long term, a consistently high tail risk lowers This finding reinforces my justification for limiting the VOV as a potential adjustment strategy in making the returns more consistent but also indicates that doing so could potentially limit tail risk and therefore aid in making a volatility arbitrage strategy more resistant to drawdowns.

III. Research Strategy

One of the primary inputs in a traditional options pricing model is the volatility associated with the underlying security. The higher the volatility, the higher the likelihood of an option reaching or surpassing a given strike price within a maturity. This results in an increased time value for the option and a higher premium. Options sellers are compensated for taking on this risk in the form of a volatility risk premium. However, higher volatility is also associated with a higher downside risk for the buyer because the measure is symmetrical in both directions – the price could just as equally go down as it could go up. As such, the volatility implied by the option's price is a measure of investor uncertainty. Historical volatility is certain and documented while implied volatility is an estimate of future volatility, adjusted for the uncertainty risk. For this reason, implied volatility is usually significantly higher than historical volatility.

One options strategy known as volatility arbitrage assumes that the risk premium is unjustified – rather, future volatility is more closely associated with historical volatility than implied volatility. If volatility is consistent period after period, then given unchanging circumstances, it is logical that it would remain more similar to its historical averages. This is further justified by the trend of mean reversion – investors often overreact to news, leading to a spike then a drop in volatility. In the long-term, volatility smooths out to a historical average.

Evidence from papers such as Goyal & Saretto (2009) shows that historically, this assumption has held up. A volatility arbitrage strategy using straddles or delta-hedging resulted in statistically significant profits over the study period. However, there are a number of assumptions and qualifications that require further testing to validate. Most significantly, one trait of volatility arbitrage strategies that often appears in professional settings is a leptokurtic distribution with fat tails. Specifically, the majority of returns over many trading periods are small and positive. However, they are punctuated by large drawdowns from unexpected events. As such, one theory holds that the long-run return of any vol arb strategy is small or negative but appears positive because the time period studied is too short to account for these major drawdowns (Rennison & Pedersen, 2012). Rather than attempting to find a variant upon standard volatility arbitrage that is not subject to these large drawdowns, which I do not believe is possible, my goal is to modify it in such ways as to change the risk/return profile over the long-run to find efficient strategies to match investor preferences.

The first variation upon the strategy implemented by Goyal & Saretto (2009) will be a change from monthly to weekly option maturities. Weekly options, as evidenced by Andersen et al (2016), isolate volatility risk better than monthly options, which are subject to other various economic uncertainties. Weekly ATM options, as would be the case with straddles and short straddles are also more isolated from jump risk, which would help prevent against drawdowns. More generally, I believe weekly options have three other major advantages over longer maturities. For one, weekly options take better advantage of compounding effects. Since the majority of returns are small and positive, the fight is against time. Increasing the frequency of running the strategy allows for a grater cushion to be created before large drawdowns occur. Second, the driving principle of volatility arbitrage is to achieve a more accurate prediction of volatility than what is implied by the market. This is easier to do with weeklies because there is less time for an unexpected event to occur – this is a corollary of what is stated by Andersen et al (2016). Finally, shorter expirations attract significantly more interest than longer dated expirations. As a result, there is greater liquidity and a smaller bid-ask spread, which could be beneficial in some cases.

The second variation will focus specifically on modifications on liquidity. Goyal & Saretto's exceptional returns come from using the mid-price of the spread to buy and sell contracts. When using the bid while selling and the ask when buying, their returns become statistically insignificant, indicating that liquidity can make it or break it for this type of strategy. Cao & Han (2013) observe that delta-hedging produces poorer results with decreased liquidity, noting that options dealers charge a premium for providing this liquidity. From this, I believe it would be probable that in the case of long-short straddle portfolio, the long side would benefit from increased liquidity and the short side would benefit from decreased liquidity. As a result, putting additional restrictions on liquidity when evaluating portfolio candidates would help account for this observation.

Finally, I believe an unbalanced hedge, one that is biased towards the short side, would be a far more effective strategy in the long-run, given the type of distribution. A 50-50 split between the long and short side provides a balanced capture of the IV-HV difference, but the IV-HV difference is not equal. Rather, if mean-reversion theory holds, jumps in IV are followed by significant drops due to overreaction to news – more than would have been predicted by historical volatility. The reverse, however – sudden drops in volatility – do not happen anywhere near as often. No event can prompt it, and I cannot identify any reliable cause other than a reaction to a spike, which would have been already captured on the long side. Although an incomplete hedge would expose the strategy to higher drawdowns, since the majority of returns would see significant increases, this should offset the higher losses.

In short, I will test the influence of changing the portfolio maturity, adding liquidity restrictions, and changing the hedge ratio compared to a monthly, unrestricted portfolio of long and short straddles.

IV. Data

The data is sourced from the OptionMetrics IvyDB US database through Wharton Research Data Services (WRDS). The *optionm_all* data library contains daily option prices for each year between 1996 and 2020. Each year is stored as a separate database within the library due to size limitations. The focus of the preliminary model was 2019 data, stored in the *opprcd2019* database. Within this dataset, the *date* (date of observation), *symbol* (unique identifier of each option), *secid* (unique identifier for the underlying security) *strike_price, best_bid, best_offer, open_interest, volume, impl_volatility, exdate* (expiration date), and *cp_flag* (whether the option is a call or a put) variables were collected. This database was joined with two other datasets, the *hvol2019* dataset and the *secprd* dataset. The former contains historical volatility data necessary for HV-IV calculations. The latter contains a code to decipher the *secid* variable into tickers. The datasets were joined to match rows based on the *secid* and *date* labels. Historical volatility data was limited to the most recent 365-day sample. In total, the dataset contained every ticker and every option of that ticker for the year 2019. It was composed of 233,867,903 observations across 14 variables. Because of the memory and size limitations, most of the code was initially tested on a much smaller sample, then applied to the larger dataset on the WRDS servers.

Data cleaning included these operations: eliminating observations where the bid price was greater than the ask price, where the bid price was equal to 0, where the difference between the bid and the ask was less than 0.05 for stocks closing under $3 or less than 0.10 for all other stocks (minimum tick size requirements), and where open interest was equal to 0. This was done to verify logical constraints. The ask must necessarily be greater than the bid to ensure supply and demand. There must be *some demand* for the option. And the option must meet institutional tick size requirements under Rule 612 of the SEC.

For simplicity and to match the process performed in Goyal and Saretto (2009), the midpoint between the bid and ask was created as an average of the two. Then, a subset of the dataset was created with only options expiring in exactly one month's time. Since straddles are at-the-money, this was further narrowed to options with strike prices closest to closing price of the stock on the given date. In the end, each stock only had a single pair of puts/calls for any given day. Goyal & Saretto take the average of the call and put IV's as the basis for the straddle's IV. I followed suit, calculating the IV for pairs of calls and puts per stock per month. The vast majority of monthly options expire on the third Friday of every month, so portfolios were constructed 30 days before the third Friday of each month.

Below are summary statistics for a sample of cleaned data that is for the full year of 2019, limited to at-the-money and 1 month to expiration. 2019 was chosen because it is a recent example of an unexceptional year. A single year was chosen to limit temporal differences in the strategy. Within the data for the single year, the summary statistics are for the full range of observations that would be sorted into deciles and long-short-portfolios.

Table 1: Summary Statistics

var	n	mean	sd	median	min	max	range	skew	kurtosis	se
strike price	95387	80.77	283.42	37.50	0.50	8725.00	8724.50	19.16	469.46	0.92
best bid	95387	2.32	5.88	1.17	0.01	197.80	197.79	16.64	397.45	0.02
best offer	95387	2.68	6.08	1.45	0.02	201.40	201.38	15.95	372.28	0.02
mean price	95387	2.50	5.97	1.33	0.02	199.60	199.59	16.33	385.92	0.02
open interest	95387	703.38	3868.61	53.00	1.00	253836.00	253835.00	19.86	654.12	12.53
volume	95387	90.36	649.36	2.00	0.00	37954.00	37954.00	21.26	637.79	2.10
impl vol	95387	0.39	0.26	0.32	0.02	2.92	2.90	2.46	10.39	0.00
his vol	95387	0.40	0.22	0.34	0.01	5.38	5.37	2.64	22.87	0.00
close	95387	80.78	283.41	37.45	0.32	8733.07	8732.75	19.16	469.50	0.92

Strike price, best bid, best offer, mean price, impl vol, his vol, close measured in dollars; open interest and volume measured by quantity of transactions.

There are a few things to note through these observations. First, the data shrunk by over 99%, from over 200 million observations to under 100 thousand, which is then further selected from based on deciles. This makes the data much easier to work with. It is also interesting that historical and implied volatilities have very similar summary metrics (~0.4 mean, ~0.33 med ~0.25 std) although historical volatility actually tends to have a higher max and a higher range. This is another piece of evidence to support the theory that on average and over time, implied volatility tends to revert back to historical levels. What does seem to be unusual is that historical volatility has a significantly higher kurtosis than implied volatility, indicating there are more tails or outliers. This could be in part due to the nature of implied volatility derivations. It could also be that investors often do not or cannot account for the potential of tail risks observed historically. The last observation would be on the range of open interest and volume. Volume has an automatic minimum at 0 while open interest has a defined minimum at 1 in order to allow at least some trading capacity. By contrast, the max values are in the tens or hundreds of thousands. This liquidity difference should be sufficient to test the liquidity hypothesis.

V. Empirical Specification

Each test was modeled through a similar methodology. At the beginning of a trading period, the difference between implied and historical volatility was calculated for every stock. Historical volatility is calculated as the standard deviation of realized daily stock returns over the last 365 days. This is done by first taking the continuously compounded return of day-to-day returns $R_n = \ln(\frac{C_n}{C_{n-1}})$ where C_n is the current closing price and C_{n-1} is the previous closing price. Then, an average of these returns is calculated with $R_{avg} = \frac{\sum_{i=1}^{n} R_i}{n}$ followed by the standard deviation from the mean $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(R_i - R_{avg})^2}{n-1}}$. Finally, the historical volatility is annualized. The implied volatility is the average of the ATM call and put implied volatilities. Both were sourced from OptionMetrics IV databases. The percent difference between the two was then sorted into deciles. These deciles formed the basis of portfolio creation. Option pairs in the first decile (largest negative difference between historical and implied volatility) were put in the short straddle portfolio. Options pairs in the 10^{th} decile (largest positive difference between historical and implied volatility) were put in the long straddle portfolio.

Table 2: First Decile (Short Straddle Portfolio)

vars	n	mean	sd	median	min	max	range	skew	kurtosis	se
strike price	9545	26.9	48.2	12.5	0.5	1160.0	1159.5	9.7	178.6	0.5
best bid	9545	1.2	2.0	0.7	0.0	45.2	45.2	7.7	105.1	0.0
best offer	9545	1.7	2.3	1.1	0.0	50.1	50.1	6.5	81.9	0.0
open interest	9545	690.1	3259.3	53.0	1.0	94447.0	94446.0	13.7	258.1	33.4
volume	9545	93.2	635.5	0.0	0.0	24263.0	24263.0	18.5	460.0	6.5
impl vol	9545	0.7	0.4	0.6	0.0	2.9	2.9	1.4	2.4	0.0
his vol	9545	0.5	0.2	0.4	0.0	2.0	2.0	1.1	2.3	0.0
close	9545	27.0	48.3	13.0	0.3	1162.5	1162.2	9.6	178.1	0.5
mean price	9545	1.5	2.1	0.9	0.0	47.7	47.6	7.2	97.0	0.0

Table 3: Last Decile (Long Straddle Portfolio)

vars	n	mean	sd	median	min	max	range	skew	kurtosis	se
strike price	9534	133.7	503.4	45.0	0.5	8250.0	8249.5	11.3	153.5	5.2
best bid	9534	3.2	9.0	1.3	0.0	143.7	143.7	9.8	121.2	0.1
best offer	9534	3.6	9.3	1.6	0.1	147.2	147.1	9.6	117.6	0.1
open interest	9534	629.6	3473.5	50.0	1.0	179019.0	179018.0	22.9	875.9	35.6
volume	9534	94.8	635.9	2.0	0.0	23842.0	23842.0	20.0	553.3	6.5
impl vol	9534	0.3	0.2	0.3	0.0	2.1	2.1	1.5	3.2	0.0
his vol	9534	0.5	0.4	0.4	0.0	5.4	5.3	3.2	24.7	0.0
close	9534	133.6	503.4	44.1	0.4	8241.9	8241.5	11.3	153.4	5.2
mean price	9534	3.4	9.2	1.5	0.0	145.4	145.3	9.7	119.6	0.1

Tables 2 and 3 compare the summary statistics of options that were placed within the first and tenth deciles. They indicate some initial differences between the options that will eventually make their way into the short and long portfolios. Only these deciles are used in the formation of long and short portfolios so the options in the remaining deciles are of no interest to us. There are a few notable differences other than their difference between implied and historical volatility. First, the tenth decile portfolio tends to have stocks with higher prices, reflected in a higher *strike price, close,* and *mean price*. Open interest and volume, surprisingly, seem to be about the same. This suggests there may not be a liquidity difference between them. It is also of note that implied volatility appears greater in the first decile, while historical volatility remains about the same. Both are quite similar to each other, even between deciles. However, the kurtosis for the historical volatility is much greater in the last decile compared to the first, indicating that tail risk could be higher in long portfolio than in the short.

The return for long straddles was calculated as the difference between the net premium at which the put and calls were sold at expiration and the price paid initially, divided by the initial net premiums. The return for short straddles was calculated as the difference between the initial net premium received and the final net premium paid out, divided by the initial net premium received.

Each variant / test of the baseline monthly strategy has an additional step in the calculation process. The weekly strategy requires a change from portfolios created at the beginning of about a 30-day period and sold at the end of it to portfolios created at the beginning of a 5-day period and sold at the end of it. The liquidity strategy requires an analysis of the bid/ask spread, as well as the open interest values. For this, the assumption that the mid-price of the bid and ask is generally attainable is relaxed. Instead, I assume that the purchase and sale of options is done at an unfavorable price. If buying a put or call, the purchase price is closer to the ask. If selling, the purchase is closer to the bid. Specifically, I look at a 25% bias towards to the unfavorable side. For example, the price of purchasing a long straddle would be (0.25*bid + 0.75*ask) while the price of selling it would be (0.75*bid + 0.25*ask). Liquidity requirements for the options are created by ranking them into deciles based on OI, choosing all long portfolio options from the last decile, and choosing all short portfolio options from the first decile. Finally, for the hedging test, I will optimize the Sharpe ratio and the total CAGR in two separate tests for each of the other strategies by changing the weights of the long and short portfolios to determine the optimal hedge.

One of the analyses of risk and return requires additional explanation. To compare the risk and return of the market to the risk and return of the options portfolio, I decided to apply the Capital Asset Pricing Model (CAPM) and derive the Capital Market Line (CML). The CAPM formula is $ER_i = r_f + \beta_i(ER_m - r_f)$ where ER_i is the expected return of investment, r_f is the risk-free rate, β_i is the beta of the investment, and $(ER_m - r_f)$is the market risk premium. To apply this analysis, I first calculated the returns of the S&P 500 index and the VIX index for the same time and duration as the options portfolio. The S&P is the standard set of risky equity assets while the VIX is calculated from market volatility. From these set of returns, I calculated the covariance between the options strategy and the two indexes and the variance of the two indexes. The entire set of monthly returns for the six-year period was used. These numbers gave me the beta of the strategy relative to each index. $ER_m$was set as the average annual return of the S&P and the VIX between 2015 and 2020. I assumed the risk-free rate to be the current 10-year treasury. The CAPM formula was then applied to find ER_i. Using beta provides an indicator of systematic risk.

In contrast, the Capital Market Line Equation, derived from CAPM, gives an estimate of total risk in the market. It compares the market and portfolio returns at a level of standard deviation to estimate its efficiency. The CML formula is s $R_p = r_f + \frac{R_T - r_f}{\sigma_T}\sigma_p$ where R_p is the portfolio return, r_f is the risk-free rate, R_T is the market return, σ_p is the standard deviation of portfolio returns, and σ_T is the standard deviation of market returns.

Results

Monthly / Baseline Strategy

Year	Long Avg Monthly Return	Short Avg Monthly Return	L/S Avg Monthly Return
2015	9.6%	9.1%	9.3%
2016	-3.6%	7.0%	1.7%
2017	-1.6%	-2.3%	-1.9%
2018	17.8%	-1.7%	8.1%
2019	8.1%	-3.4%	2.4%
2020	49.4%	-9.8%	19.8%
Total	13.3%	-0.2%	6.6%

Table 4: Average Monthly Returns; long, short, and combined (L/S) portfolios. Return values are expressed as the average of each portfolio's total monthly returns for each year. The "total" row averages the return for all months for the 6-year period.

Year	Long CAGR	Short CAGR	L/S CAGR
2015	129.1%	145.5%	159.2%
2016	-37.1%	96.0%	17.8%
2017	-17.3%	-23.5%	-20.0%
2018	356.7%	-32.9%	128.2%
2019	117.9%	-35.6%	27.1%
2020	676.4%	-351.7%	372.1%
Total	112.5%	-166.8%	79.5%

Table 5: Compound Average Growth Rate (CAGR); monthly strategy; long, short, and combined (L/S) portfolios. Yearly CAGR is calculated as percent return assuming all profits/losses at the end of each month are reinvested at the beginning of the following month. The "total" column CAGR is calculated through the formula $CAGR = \left(\frac{V_{final}}{V_{begin}}\right)^{\frac{1}{t}} - 1$ *where* V_{final} *is the final value,* V_{begin} *is the beginning value and t is time in years. The profits/losses at the end of each year are reinvested at the beginning of the following year. The value of the portfolio at end of 2020 is* V_{final}.

Moving forward, the combined or hedged portfolio consisting of a 50% allocation in the short portfolio and a 50% allocation in the long portfolio will be referred to as the Long/Short (L/S) portfolio. In table 4, return values are expressed as the average of each portfolio's total monthly returns for each year. The "total" row averages the return for all months for the 6-year period. In table 5, yearly CAGR is calculated as percent return assuming all profits/losses at the end of each month are reinvested at the beginning of the following month. The percent difference between the value of the final month and the original starting amount is the CAGR. The "total" column CAGR is calculated through the formula $\mathrm{CAGR} = \left(\frac{V_{final}}{V_{begin}}\right)^{\frac{1}{t}} - 1$ where V_{final} is the final value, V_{begin} is the beginning value and *t* is time in years. The profits/losses at the end of each year are reinvested at the beginning of the following year. The value of the portfolio at end of 2020 is V_{final}.

The first set of results discussed will be the returns on the baseline / monthly portfolios over the period. On average, the strategy returned 6.6% per month, ranging from an average of -1.9% per month in 2017 to +19.8% per month in 2020. Reinvesting all profits from the beginning of each year until the end of each year resulted in a total CAGR of 79.5%. Individual annual returns ranged from -20.0% in 2017 to +372.1% in 2020.

Within the time range, the return was primarily driven by the long portfolio which significantly outperformed the short portfolio. Overall, the long portfolio averaged 13.3% per month while the short portfolio averaged almost no movement, falling by 0.2% per month. This translated to a 112.5% return for the long portfolio and a -166.7% return for the short portfolio. 2020 and 2018 were the best years for the long portfolio, returning 676.4% and 356.7% respectively, while 2016 was the worst year, falling by 37.1%. 2015 and 2016 were the best years for the short portfolio, returning 145.5% and 96.0% respectively, while the worst year was 2020 where it fell by 351.7%. This disparity is actually quite unusual given the fact that volatility was low for the majority of the bull market. The theory would be that constant low volatility would benefit the short vol portfolio more than the long. Part of the reason for the difference could be the nature of the underlying strategies. Long straddles have limited risk and infinite profit potential while short straddles have unlimited risk and limited profit potentials.

The results were also heavily skewed by the COVID crisis and the resulting stock market volatility in 2020. That year, the long portfolio returned 49.4% per month and grew by 676.4% while the short portfolio returned -9.8% per month and fell by 351.7% overall. Excluding 2020 brings the long, short, and combined portfolio average monthly returns to 6.1%, 1.8%, and 3.9% respectively. As a result, the average CAGR for each portfolio changes to 109.9%, 29.9%, and 62.4%, respectively. This is a decrease in the long and combined portfolio returns but a significant increase in the short portfolio returns.

The table below provides the minimum, maximum, first quartile, third quartile, and median returns, yearly and as a whole. Except for 2017, the annual median returns for the L/S portfolio are positive. Between 2015 and 2020, the median monthly return was 3.6%. One interesting feature is the fact that the scale of minimum and maximum returns varies between portfolios. The minimums of the short portfolio tend to be absolutely greater than the maximums while the maximums of the long portfolio tend to be equal or greater than the minimums. The L/S portfolio does not have as strong of a pattern.

Year	Quartile	Long	Short	L/S
2015	Min	-14.6%	-15.4%	-0.2%
	1	-2.7%	4.5%	4.5%
	Med	1.7%	7.4%	6.0%
	3	9.6%	16.6%	11.5%
	Max	60.2%	28.0%	27.3%
2016	Min	-20.0%	-24.7%	-9.4%
	1	-12.5%	2.3%	-3.4%
	Med	-4.4%	11.9%	1.2%
	3	5.2%	14.1%	5.8%
	Max	12.5%	18.9%	15.7%
2017	Min	-7.1%	-11.2%	-8.3%
	1	-5.6%	-6.2%	-4.8%
	Med	-0.9%	-3.3%	-1.7%
	3	1.3%	1.6%	0.9%
	Max	5.3%	7.8%	3.3%
2018	Min	-7.4%	-43.1%	-1.6%
	1	-4.2%	-9.4%	4.2%
	Med	5.7%	2.0%	6.4%
	3	23.4%	9.0%	9.2%
	Max	82.3%	22.2%	31.5%
2019	Min	-16.6%	-19.2%	-6.4%
	1	0.7%	-10.5%	-1.9%
	Med	10.2%	-7.2%	2.8%
	3	16.9%	3.9%	6.4%
	Max	27.8%	11.3%	10.3%
2020	Min	-22.6%	-208.3%	-10.9%
	1	-11.2%	-4.2%	0.4%
	Med	3.4%	12.1%	8.5%
	3	28.2%	20.1%	19.1%
	Max	491.5%	30.9%	141.6%
Total	Min	-22.6%	-208.3%	-10.9%
	1	-5.3%	-5.6%	-1.1%
	Med	2.1%	3.7%	3.6%
	3	10.8%	12.2%	8.5%
	Max	491.5%	30.9%	141.6%

Table 6: Quartile statistics of monthly strategy; all portfolios

Overall, the returns of the L/S portfolio showcase the benefits of hedging. This is especially evident in the yearly minimums. For example, the minimum average monthly return in 2015 was -14.6% and -15.4% in the long and short portfolios, respectively, However, the combined portfolio only had a minimum of -0.2%. Similarly, in 2018 the minimum returns were -7.4% and -43.1% for the long and short but only -1.6% for the combined. The maximum returns per year, however, were not as equally reduced, coming in at a return approximate to one of the other portfolios or close to average of the two.

Without adjusting for risk, the returns do seem abnormally large. The long-term monthly average of the S&P 500 is 0.66% / month, giving a hedged vol arb strategy, which returned an average of 6.6% per month, about 10 x the return of the market. However, this brings up the question of riskiness. Is the high return justified by its corresponding level of risk or are the returns still high even adjusting for the level of risk?

The next section discusses the risk and distribution of returns over time. The standard deviation of monthly returns was multiplied by the square root of 12 to annualize. Table 7 shows the annualized standard deviation of the portfolios for each year in the study.

Year	Long SD	Short SD	L/S SD
2015	72.7%	38.1%	28.1%
2016	35.7%	40.4%	24.2%
2017	14.3%	18.6%	12.7%
2018	99.4%	61.9%	28.3%
2019	44.3%	34.4%	19.4%
2020	490.2%	223.7%	139.5%
Total	216.5%	101.2%	65.5%

Table 7: Standard deviation of returns, monthly strategy

The long portfolio tended to have similar or higher standard deviation than the short portfolio, indicating a possibility that its stronger returns came with a higher risk profile. The L/S portfolio showcased the true function of hedging. Every year, the standard deviation of returns was significantly reduced by hedging the two portfolios together. For instance, in 2018 the long portfolio has a SD of 99.4% and the short portfolio had a SD of 61.9%, but the L/S portfolio only had a SD of 28.3%.

The figures below show the distribution of all monthly returns, 2015-2020, for each portfolio. The red bar indicates the bin with the sample mean. We can highlight a few things. From a first glance, it appears the long distribution is positively skewed (right-tailed) while the short distribution is negatively skewed (left-tailed). The L/S portfolio is not as strongly skewed as the two parent portfolios but still has a clear positive skew. The tighter clustering of returns of the L/S portfolio around the mean is a reflection of its lower standard deviation compared to the component portfolios.

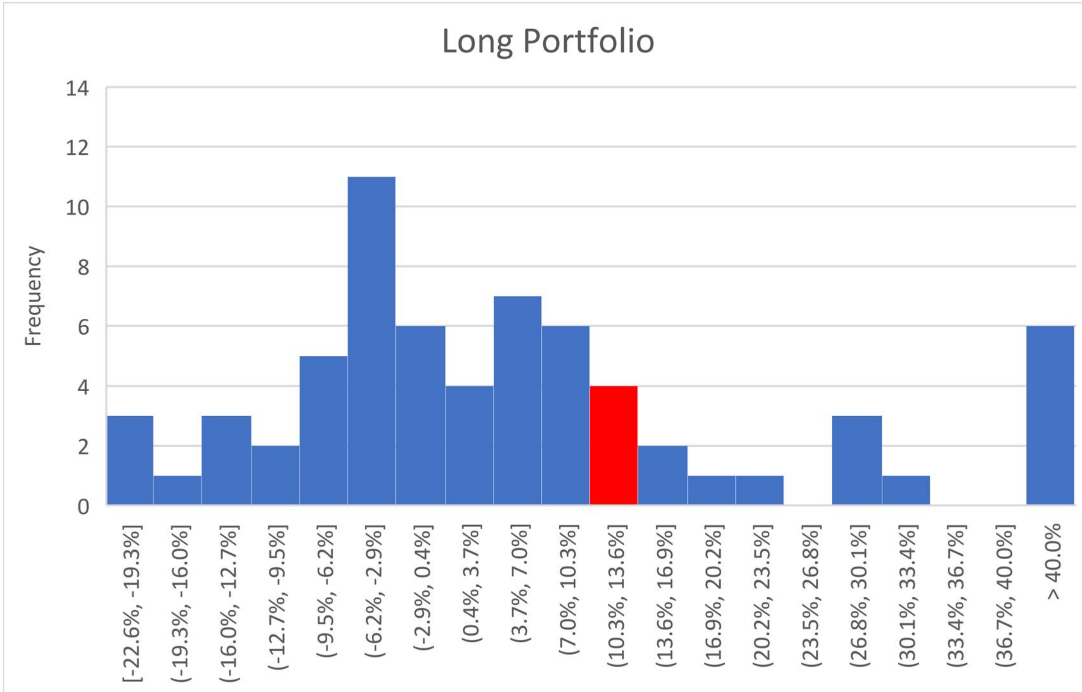

Figure 1: Distribution of all monthly returns, long portfolio

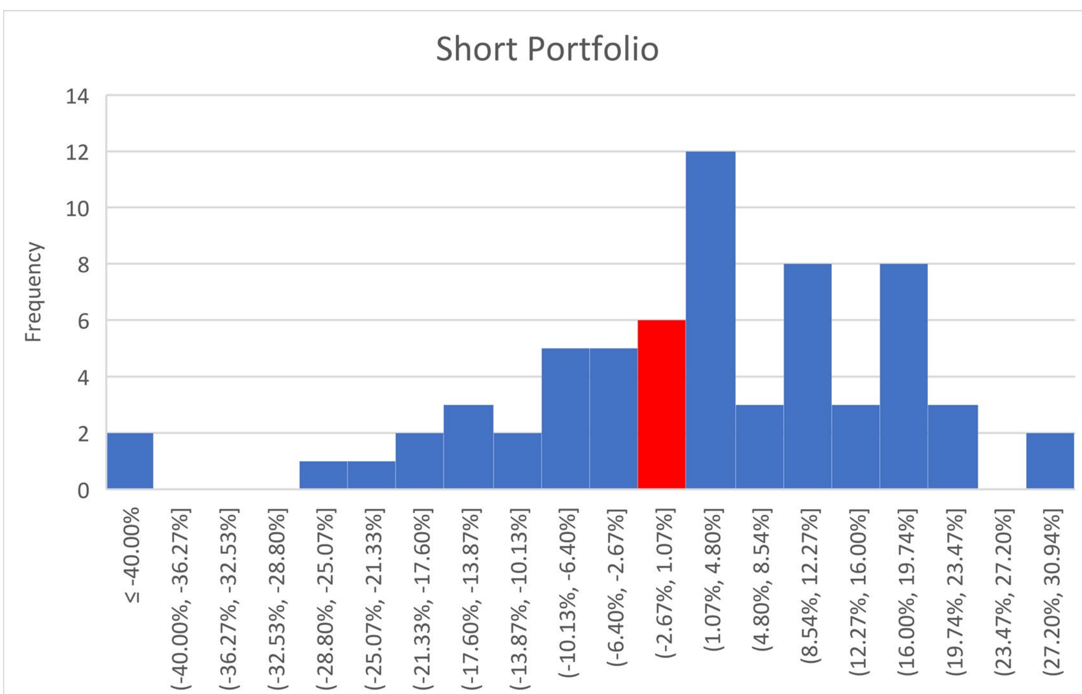

Figure 2: Distribution of all monthly returns; short portfolio

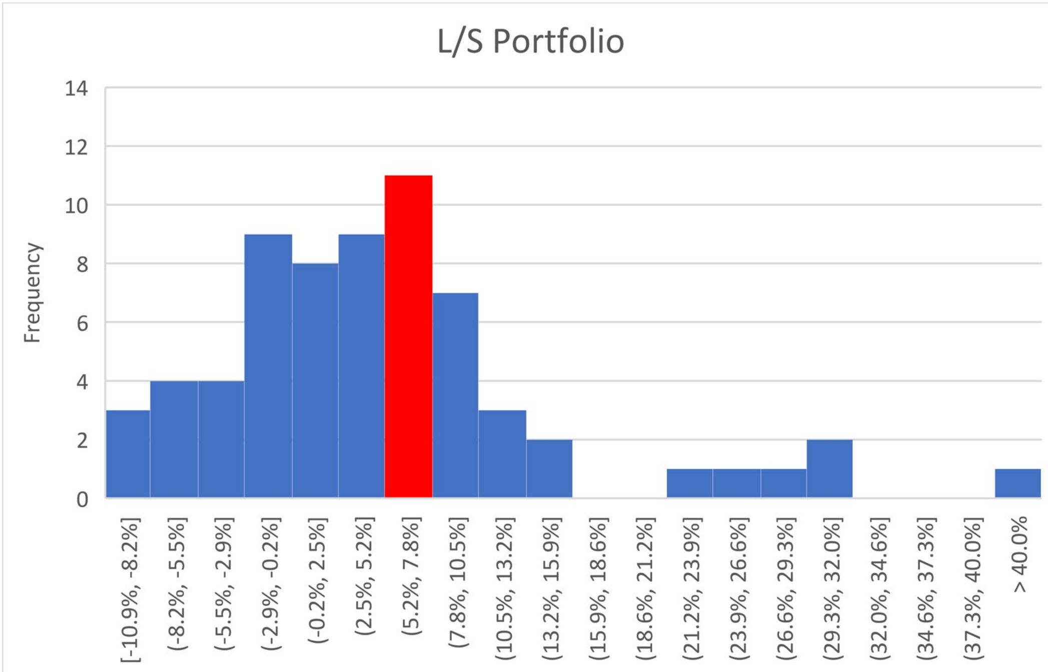

Figure 1: Distribution of all monthly returns, L/S portfolio

The tables below show the yearly and total skew and kurtosis of the monthly returns. These confirm that the long portfolio and short portfolio are skewed in opposite directions and that the L/S portfolio follows the skew of the long portfolio more than the short. The kurtosis of a normal distribution is 3. Since the kurtosis in the table rarely crosses 3, individually, most years for both portfolios are platykurtic – the strategy produces few returns that are extreme. However, on an aggregate scale that looks at kurtosis from the entire set of returns 2015-2020, the distribution is highly leptokurtic with significant fat tails that could results in a greater chance of extreme events. This supports the theory that vol arb has high returns that are often distorted by large extreme movements. However, given the fact that the L/S portfolio is right-tailed with no returns below three standard deviations of the mean, the evidence does not support the hypothesis that these returns have significant *negative* fat tails.

Year	Long Skew	Short Skew	L/S Skew
2015	1.60	-0.54	1.31
2016	0.08	-1.95	0.47
2017	0.29	0.23	-0.17
2018	1.45	-1.03	2.27
2019	-0.52	0.08	-0.26
2020	3.18	-3.00	2.83
Total	6.96	-5.60	5.69

Table 8: Skewness of all monthly returns; all portfolios

Year	Long Kurtosis	Short Kurtosis	L/S Kurtosis
2015	2.13	1.13	0.92
2016	-1.34	4.50	-0.10
2017	-1.41	-0.49	-1.01
2018	1.11	1.22	6.54
2019	-0.33	-1.34	-1.19
2020	10.31	9.42	8.62
Total	52.98	39.20	39.60

Table 9: Kurtosis of all monthly returns; all portfolios

Finally, how does the risk and return of these portfolios compare to the risk and return of the broader market? To do this, I turned to the CAPM model, comparing the expected return of the portfolio to its actual return. The beta of the L/S portfolio compared to the S&P 500 for the same time period was -2.34. Compared to the VIX, it was 0.036. The risk-free rate was set as the 10-year treasury which yielded 1.72% as of time of writing. The S&P returned 11.19% per year over the last six years while the VIX grew by an average of 16.69%. The CAPM model calculations are below.

$$R_i = R_f + \beta_i(ER_m - R_f)$$

$$S\&P\ 500\text{: } R_i = 0.0172 - 2.34(0.1119 - 0.0172) = -20.5\%$$

$$VIX\text{: } R_i = 0.0172 + 0.036(0.1669 - 0.0172) = 2.3\%$$

Because the CAPM assumes that higher returns have to be associated with higher betas, the L/S portfolio overperformed for its level of systematic risk. Since its beta against the S&P was high and negative, it means it had significantly higher volatility and moved against the market. And since its beta against the VIX was almost zero, it moved with similar magnitude to the VIX but had no correlation with its movements.

Next, the standard deviation of market returns was calculated as the SD of returns over the same investment periods as the options portfolio, annualized. This was set as the input into the CML, from which the Sharpe ratio was then derived.

$$R_p = r_f + \frac{R_T - r_f}{\sigma_T}\sigma_p$$

$$S\&P: R_p = 0.0172 + \frac{0.1119 - 0.0172}{0.1917} 0.6551 = 34.1\%$$

$$VIX: R_p = 0.0172 + \frac{0.1669 - 0.0172}{1.797} 0.6551 = 7.2\%$$

Compared to the expected portfolio return given the level of risk in the market, the options portfolio still outperformed, but the bar for expected return was also set much higher. The actual CAGR for the L/S portfolio was 79.5% for the period, doubling the expected return. The Sharpe ratio is the slope of the CML, providing an easy way to compare risk and return. The Sharpe ratio is captured by the formula $S(x) = \frac{(r_x - R_f)}{SD(rx)}$ where r_x is the average rate of return for the investment and R_f is the risk-free rate. Below is the Sharpe ratio calculation for the L/S portfolio.

$$S(L/S) = \frac{(.795 - 0.0172)}{0.6551} = 1.19$$

In this context, the Sharpe ratio can be used to judge this as a good investment since it is above 1, but it is not considered very good or excellent because it is not above 2 or 3. It is interesting to note that the high standard deviation of the long portfolio, despite its higher returns, makes it a poor investment in isolation. Its Sharpe ratio is only about 0.5, which means its return is overpriced for its level of risk.

$$S(L) = \frac{(1.125 - 0.0172)}{2.165} = 0.512$$

However, these estimates are distorted by the fact that returns are heavily skewed and so standard deviation of the portfolio is not a perfectly accurate measure of risk. CAPM and, by proxy, the Sharpe ratio assumes returns are normally distributed. In reality, since returns are positively skewed, the standard deviation is likely an overestimate of the true riskiness of the portfolio. However, adjusting for higher moments would still not likely make an isolated long portfolio a viable investment.

Weeklies Strategy

For the sake of brevity, the remaining tests will have a condensed analysis consisting of CAGR, standard deviation, distribution, skew & kurtosis, and Sharpe ratio.

The second variant of options arbitrage is weeklies, modifying the length of a portfolio from 1 month to 1 week. Hypothetically, weeklies isolate volatility risk better than monthly strategies. Options were ranked into deciles and portfolios were created and opened on every Monday of the year, expiring the upcoming Friday.

Year	Long Avg Weekly Return	Short Avg Weekly Return	L/S Avg Weekly Return
2015	14.0%	1.9%	8.0%
2016	8.3%	-2.7%	2.8%
2017	8.5%	-2.7%	2.9%
2018	16.5%	-13.8%	1.3%
2019	4.1%	-12.4%	-4.2%
2020	11.7%	-1.6%	5.0%
Total	10.5%	-5.3%	2.6%

Table 10: Average Weekly Returns; all portfolios. Return values are expressed as the average of each portfolio's total weekly returns for each year. The "total" row averages the return for all weeks for the 6-year period.

Year	Long CAGR	Short CAGR	L/S CAGR
2015	3905.79%	28.59%	1591.49%
2016	181.40%	-94.97%	114.27%
2017	1785.51%	-84.48%	161.74%
2018	39500.34%	-99.99%	27.06%
2019	134.65%	-100.01%	-92.76%
2020	3214.7%	-96.6%	520.9%
Total	1907.6%	N/A	94.5%

Table 11: Compound Average Growth Rate; weekly strategy; all portfolios. Yearly CAGR is calculated as percent return assuming all profits/losses at the end of each week are reinvested at the beginning of the following week. The "total" column CAGR is calculated through the formula $CAGR = \left(\frac{V_{final}}{V_{begin}}\right)^{\frac{1}{t}} - 1$ *where* V_{final} *is the final value,* V_{begin} *is the beginning value and t is time in years. The profits/losses at the end of each year are reinvested at the beginning of the following year. The value of the portfolio at end of 2020 is* V_{final}.

From a first glance, it does appear that the patterns observed in the monthly portfolios were magnified in the weekly portfolios. The long portfolio CAGR has some unusually high returns for the time period including an almost 40,000% return in 2018 and a 4,000% return in 2015. What is surprising is that *the long portfolio did not have a single year where it did not at least double its investment.* On the other hand, the short portfolio lost almost all of its investment every single year. As a result of these two extremes, the L/S portfolio had significant fluctuations in its return, ranging from a 92.8% drop in 2019 to a 1591.5% gain in 2015.Most of these high returns or losses are simply the result of more periods for compounding. A few weeks of median (5%) returns followed by a few abnormally high returns (>100%) will compound to large numbers very quickly. A table of summary statistics including min, 1Q, med, 3Q, and max is found in the appendix.

The next section considers the risk as measured by standard deviation, skew, and kurtosis. Below are histograms representing the distribution of all returns, 2015-2020, for the long, short, and L/S portfolios. The red bin contains the mean of returns. By the law of large numbers, because of a larger sample size, the distribution looks much more fitted to the characteristics described in the monthly portfolios. The long portfolio is clearly right tailed while the short portfolio is clearly left tailed. The combined portfolio has fat tails on both sides as a result but not does not have a strong skew. It is also much more clustered around the mean, indicating a lower standard deviation.

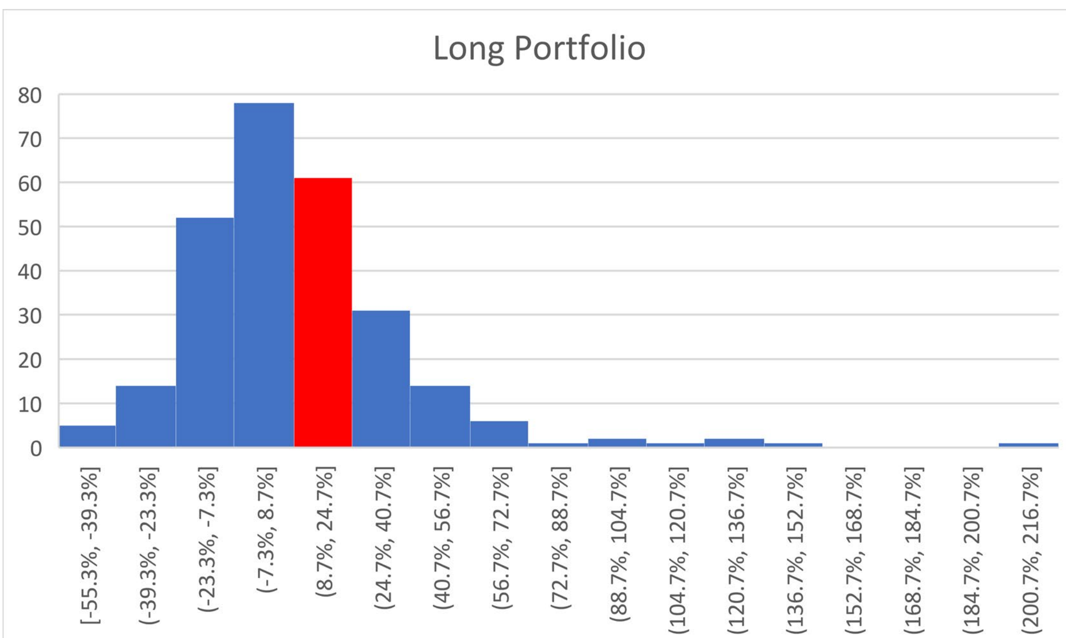

Figure 2: Distribution of all weekly returns; long portfolio

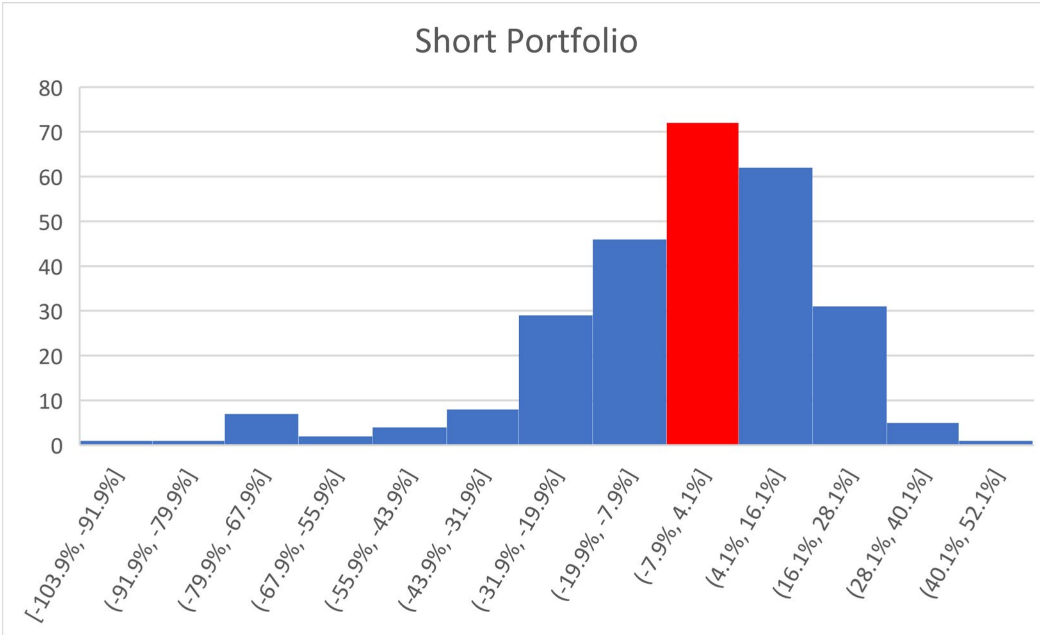

Figure 3: Distribution of all weekly returns; short portfolio

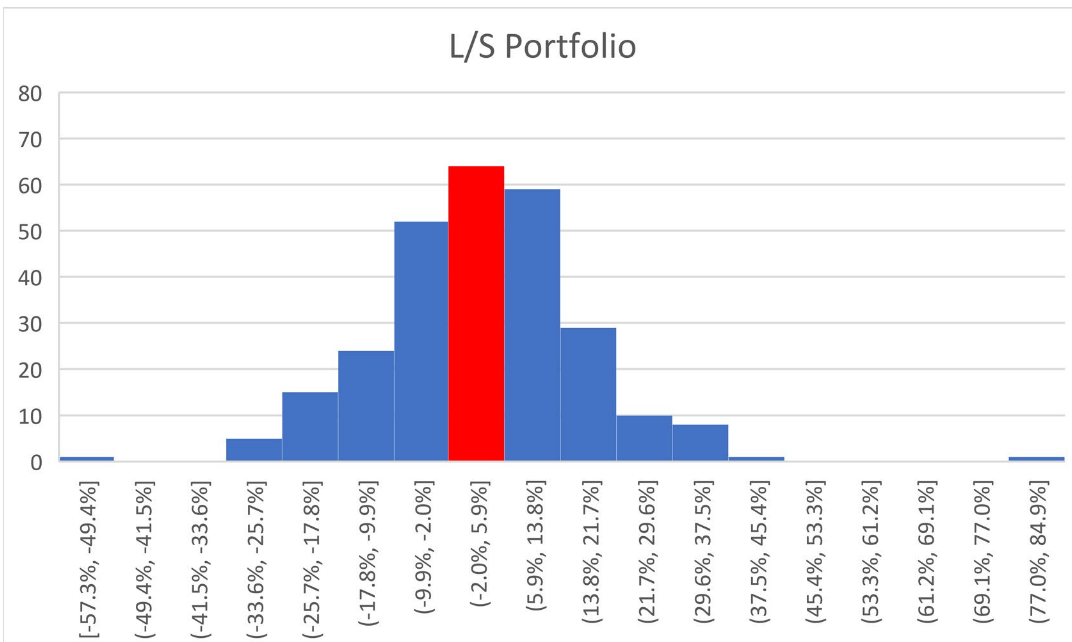

Figure 4: Distribution of all weekly returns; L/S portfolio

These findings are confirmed by the indicators below, showing annualized SD, skewness, and kurtosis. As is expected, there is a higher standard deviation of returns per year than one would find in the monthly portfolio. And again, the combined L/S portfolio has significantly lower SD than either one of the component portfolios are a result of hedging.

Year	Long SD	Short SD	L/S SD
2015	290.2%	111.6%	122.2%
2016	261.5%	158.0%	107.3%
2017	142.0%	115.2%	87.2%
2018	203.3%	179.5%	92.9%
2019	156.6%	157.7%	108.0%
2020	214.2%	191.0%	91.0%
Total	218.9%	160.9%	105.5%

Table 12: Standard deviation of returns, weekly strategy

As evidenced by the histogram, the long portfolio is skewed to the right while the short portfolio is skewed to the left. The major difference between this and the monthly portfolio is that returns in the L/S portfolio are not highly skewed right. The hypothesis being, the larger the sample size, the more the long and short portfolios average out.

Year	Long Skew	Short Skew	L/S Skew
2015	3.01	-1.05	2.11
2016	0.81	-1.36	0.18
2017	0.58	0.23	0.23
2018	2.70	-0.81	0.57
2019	0.47	-0.66	0.31
2020	1.53	-1.54	-0.26
Total	2.05	-1.27	0.46

Table 13: Skewness of all weekly returns; all portfolios

The long portfolio is much more highly kurtotic than the short portfolio, which appears to have very little fat-tails at all (a normal distribution has a kurtosis of 3). The L/S portfolio, as a result, has only slight kurtosis.

Year	Long Kurtosis	Short Kurtosis	L/S Kurtosis
2015	13.04	1.24	8.18
2016	1.53	3.01	-0.39
2017	0.06	0.45	0.59
2018	9.18	1.26	2.05
2019	-0.95	-0.22	-0.98
2020	3.68	2.00	0.16
Total	9.13	2.74	3.89

Table 14: Kurtosis of all weekly returns; all portfolios

The higher moments of the distribution of L/S portfolio returns are therefore quite normal, a stark contrast to the sizable skewness and kurtosis of the monthly L/S portfolio. This gives the following CAPM analysis more validity as some of the bias is removed. However, as seen by the returns and standard deviation of the weekly portfolios, if the long portfolio has strong enough return for its level of risk, it could be the case that hedging is unnecessary. In that case, the skew and kurtosis would play a role in determining the strategy's risk profile.

$$S(x) = \frac{(r_x - R_f)}{SD(rx)}$$

$$\text{L/S: } S\left(\frac{L}{S}\right) = \frac{(.945 - .0172)}{.1055} = .88$$

$$\text{Long: } S(L) = \frac{(19.07 - .0172)}{2.19} = 8.71$$

The Sharpe ratio of the L/S portfolio is 0.88, making the level of return unjustified for the amount of risk undertaken. However, the long portfolio, despite having twice as much risk as the L/S portfolio, has a Sharpe ratio of 8.71, making it an excellent investment given its level of risk. This is, of course, still biased because of the leptokurtic distribution of long portfolio returns, possibly underestimating the true riskiness of the investment.

Hedge Ratio

The third variation on this form of vol arb explores the impact of changing the hedge ratio. To do so, I solved optimization problems to maximize the Sharpe ratio, maximize the total 6-year CAGR, and minimize the annualized standard deviation of the L/S portfolio in separate tests by changing the weights of the long and short portfolios. The only constraint was that the sum of the weightings had to be 1 (could not use leverage).

For the monthly strategy, the optimal weighting to maximize the Sharpe Ratio was 0.33 in the long portfolio and 0.67 in the short portfolio. This more than halved the annualized standard deviation from 65.5% to 29.6% but only reduced the CAGR from 79.5% to 62.8%. As a result, the Sharpe Ratio increased from 1.19 to 2.06. A maximization of CAGR yielded a full 100% allocation to the long portfolio. The Sharpe ratio was reduced to 0.43, however, because the SD increased to 216.5%. An optimal minimization of SD resulted in a 0.69 allocation in the short portfolio and a 0.31 allocation in the long. This reduced standard deviation to 28.9% but decreased the CAGR to 59.7%. As a result, the optimal Sharpe ratio is very close to the minimum standard deviation (minimum variance) portfolio.

For the weekly strategy, the optimal weighting to maximize the Sharpe ratio was a full 100% allocation to the long portfolio, increasing the ratio from 0.88 to 8.71. This also maximized the CAGR. Minimizing standard deviation resulted in a 0.61 allocation to the short portfolio and a 0.39 allocation to the long portfolio. This reduced SD from 105.5% to 99.7% but reduced the CAGR from 94.5% to -8.4% and the Sharpe ratio from 0.88 to -0.10. This barely decreased the risk but greatly decreased the return, making it an inviable strategy.

Liquidity Analysis

The final test relaxes the assumption that one can buy and sell options at the average of the bid and ask. It first compares the returns of a liquidity-adjusted portfolio with a mid-price assumption and the returns of the same portfolio with unfavorable market conditions. The same process is then done with the original baseline strategy. The goal is to compare how much of a difference unfavorable purchasing/selling price makes on returns and how much a liquidity-adjusted portfolio can offset these differences.

The hypothesis in creating liquidity-adjusted portfolios is that the long portfolio would benefit from higher liquidity while the short portfolio would benefit from lower liquidity. Tables 15 and 16 below show the results of the analysis, short portfolio on top, long portfolio on bottom. The first column shows the average monthly return for a liquidity-adjusted portfolio using the average of the bid and ask. The second column shows the same average monthly return for a liquidity-adjusted portfolio using unfavorable liquidity conditions when buying and selling. The third column provides the percentage difference between the average and the biased columns. The next three columns repeat this process but for the baseline strategy instead of the liquidity-adjusted portfolio. The final column provides a multiple between the differences. For example, in the 2015 short portfolio, there was a 42x multiple between the difference in the baseline results and the difference in the liquidity-adjusted results. This means the liquidity-adjusted portfolio experienced only 1/42 of the decrease between the average and the unfavorable conditions that the baseline portfolio experienced.

From the “short baseline bias” column in the short portfolio, it is easy to see that without a liquidity adjustment, the bid-ask spread has a significant impact on the performance of the portfolio, returning a monthly average of -35.0% over the time period compared to 3.1% using the midpoint. Choosing straddles from a portfolio of highly illiquid securities reduces the average return to -2.0% (using the mid-price). However, it also stifles the negative effective of a poor bid-ask spread as the percent change is not nearly as large as with the baseline portfolio.

Year	Short Liquidity Average	Short Liquidity Bias	Average-Bias Difference (Liquid)	Short Baseline Average	Short Baseline Bias	Average-Bias Difference (Baseline)	Difference Multiple
2015	10.6%	9.8%	-7.7%	9.1%	-20.6%	-326.5%	42.32
2016	5.3%	5.5%	3.4%	7.0%	-26.0%	-469.2%	-137.37
2017	-11.9%	-12.0%	-0.3%	-2.3%	-31.7%	-1304.6%	3814.77
2018	-17.4%	-19.6%	-12.7%	-1.7%	-37.7%	-2176.7%	171.03
2019	-21.0%	-23.7%	-13.0%	-3.4%	-41.7%	-1125.0%	86.46
2020	22.5%	-97.2%	-532.4%	9.7%	-52.1%	-637.7%	1.20
Average	-2.0%	-22.9%	-93.8%	3.1%	-35.0%	-1006.6%	663.07

Table 15: Comparison of mid-price and 25% mark between baseline and liquidity-adjusted portfolios; short portfolio. Return values are expressed as the average of each portfolio's total monthly returns for each year.

The patterns in the long portfolio are not as clearly defined as in the short portfolio. While the liquidity-adjusted portfolio did stifle some of the decrease from the mid-price to the 25% mark, both performed so poorly, it is not worth justifying this change.

Year	Long Liquidity Average	Long Liquidity Bias	Average-Bias Difference (Liquid)	Long Baseline Average	Long Baseline Bias	Average-Bias Difference (Baseline)	Difference Multiple
2015	6.2%	-8.3%	-233.9%	9.6%	-10.0%	-204.7%	0.88
2016	-9.5%	-24.0%	-353.9%	-3.6%	-25.0%	-799.5%	2.26
2017	-4.9%	-18.1%	-272.3%	-1.8%	-21.8%	-1114.0%	4.09
2018	7.1%	-10.0%	-241.3%	17.8%	-6.7%	-137.5%	0.57
2019	-3.6%	-18.0%	-397.0%	8.1%	-14.3%	-275.9%	0.70
2020	42.4%	22.3%	-47.6%	49.42%	18.75%	-62.1%	1.31
Average	6.3%	-9.4%	-257.7%	13.3%	-9.8%	-432.3%	1.63

Table 16: Comparison of mid-price and 25% mark between baseline and liquidity-adjusted portfolios; long portfolio. Return values are expressed as the average of each portfolio's total monthly returns for each year

VI. Conclusion

Given the analysis of risk and return with perspectives given on maturity, hedging, and liquidity, what is there to make of options volatility arbitrage as an investment strategy and its implications for the efficiency of the options market?

The analysis for the monthly portfolio provides a good summary for the vol arb strategy as a whole. On the surface, the returns are enticing. A 79.5% CAGR over the last 6 years would make any investor happy. However, digging below the surface brings out potential risks. For one, there is a high standard deviation and high chance of extreme events given the level of kurtosis. Adjusting for standard deviation using the Sharpe Ratio classifies the L/S monthly portfolio as a barely passable investment. There is a high return, but that return is justified given the high level of risk.

There is another way to look at it, however. First, the beta of the portfolio is high and negative against the S&P and 0 against the VIX. This means the strategy can be used to reduce correlations as part of a broader portfolio, allowing it to be used as a diversification tool. Secondly, the 50/50 hedge can be lifted to create more optimal returns. As we saw in the results of the optimization problem, increasing the weight of the short portfolio to about 2/3 slightly reduces the CAGR but greatly reduces the risk, making the L/S strategy a great investment based on its Sharpe Ratio of greater than 2. The performance of the component long and short portfolios also raises interesting questions. The short portfolio had significantly worse CAGR than the long portfolio, but it provided the hedging to reduce the standard of the overall portfolio and improve its ratio of risk to return. As a result, despite all desire to just use the long portfolio for the monthly strategy, it is still optimal to hedge it correctly.

The initial hypothesis for the hedge strategy was that a short vol portfolio would be more effective than a long vol portfolio and that “a smaller proportion long hedge (not 50-50) would be more effective than a 50-50 hedge, resulting in higher returns but likely more significant drawdowns.” The evidence indicates that a short vol monthly strategy is not more effective than a long vol strategy as returns are significantly smaller, but that a smaller long hedge *is* more effective than a 50-50 hedge. However, this is not because of higher returns but because of a higher return/risk ratio.

The weekly strategy, despite following a similar distribution of returns to the monthly portfolio, resulted in a different conclusion. For the weekly strategy, the long portfolio outperformed the short portfolio to the point where, despite its function in risk reduction, the optimal allocation was to avoid it completely. While it is difficult to make a broad generalization with only 6 recent years of data, the long weekly portfolio seems to be an outperformer, generating extremely high returns for its level of risk. These returns are, however, still biased by the higher moments in the distribution.

Relating back to the original hypothesis, the evidence points in favor of weeklies as having more consistent returns. The distribution of returns for a weekly L/S strategy is much more normal than for a monthly strategy with less skew and kurtosis. However, consistency does not translate to profitability as the combined weekly portfolio assumed a higher ratio of risk to return than the combined monthly portfolio. An isolated long weekly portfolio, however, is both more consistent and more profitable than a combined monthly portfolio as its distribution has less skew, less kurtosis, and no years of negative return. Although more research needs to be conducted, this provides some evidence that weeklies are better suited for isolating volatility risk from jump risk.

The strong return-generating ability of the long portfolio also generates evidence to weaken my initial hypothesis that IV more often than not falls towards HV rather than rise to meet it. It also raises the question why the long portfolio is an outperformer to such an extent. One possibility is that any option in the long portfolio has limited risk but unlimited profit potential while options in the short portfolio have unlimited risk but limited profit potential. Future testing would require a different strategy that does not share the limitations of a short straddle.

Conclusions arising from liquidity requirements/adjustments are unconvincing. There is evidence to support that creating liquidity restrictions on the short side might reduce the impact of a large bid/ask spread but placing this restriction also reduces the number of viable candidates and might hurt the overall return. No conclusion can be drawn on the long side as both the mid-price return and the 25% mark have similarly high negative return.

Finally, the strategy thrived in one of the largest stock market crises in the last decade, having its best year in the study period be 2020. While it is a single moment, it is a strong example of a black swan event. The strategy’s strong performance in this time could be an indicator of its resistance in times of crises but additional research needs to be conducted. Overall, I think with proper risk management, both the weekly and monthly strategies can be highly effective parts of a portfolio of an investor with a tolerance for above average risk.

Although there does seem to be some alpha-generating potential, I do not think there is enough strong evidence to support the theory that the market is not efficient. Risk is priced into many forms with this strategy that are difficult to untangle.

Sources / References

Andersen, T. G., Fusari, N., & Todorov, V. (2017). Short-Term Market Risks Implied by Weekly Options. *The Journal of Finance*, *72*(3), 1335–1386. https://doi.org/10.1111/jofi.12486

Cao, J., & Han, B. (2013). Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, *108*(1), 231–249. https://doi.org/10.1016/j.jfineco.2012.11.010

Christoffersen, P., Goyenko, R., Jacobs, K., & Karoui, M. (2017). Illiquidity Premia in the Equity Options Market. *The Review of Financial Studies*, *31*(3), 811–851. https://doi.org/10.1093/rfs/hhx113

Goyal, A., & Saretto, A. (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*, *94*(2), 310–326. https://doi.org/10.1016/j.jfineco.2009.01.001

Gu, M., Kang, W., & Xu, B. (2018). Limits of arbitrage and idiosyncratic volatility: Evidence from China stock market. *Journal of Banking & Finance*, *86*, 240–258. https://doi.org/10.1016/j.jbankfin.2015.08.016

Jarrow, R. (2013). Option Pricing and Market Efficiency. *The Journal of Portfolio Management*, *40*(1), 88–94. https://doi.org/10.3905/jpm.2013.40.1.088

McGee, R. J., & McGroarty, F. (2017). The risk premium that never was: A fair value explanation of the volatility spread. *European Journal of Operational Research*, *262*(1), 370–380. https://doi.org/10.1016/j.ejor.2017.03.070

Park, Y.-H. (2015). Volatility-of-volatility and tail risk hedging returns. *Journal of Financial Markets*, *26*, 38–63. https://doi.org/10.1016/j.finmar.2015.05.003

Rennison, G. R., & Pedersen, N. P. (2012, September). The Volatility Risk Premium. PIMCO Analytics. https://www.pimco.com/en-us/insights/viewpoints/research/the-volatility-risk-premium

Ruan, X. (2020). Volatility-of-volatility and the cross-section of option returns. *Journal of Financial Markets*, *48*, 100492. https://doi.org/10.1016/j.finmar.2019.03.002

Appendix

Year	Quartile	Long	Short	L/S
2015	Min	-40.6%	-45.0%	-18.9%
	1	-10.7%	-4.5%	-2.0%
	Med	4.2%	6.0%	5.4%
	3	24.0%	10.3%	14.5%
	Max	212.6%	31.1%	83.8%
2016	Min	-55.3%	-78.4%	-27.8%
	1	-12.4%	-11.1%	-5.9%
	Med	7.8%	2.0%	1.3%
	3	25.6%	9.9%	14.7%
	Max	121.4%	38.0%	37.0%
2017	Min	-25.4%	-40.0%	-22.6%
	1	-7.8%	-15.1%	-7.6%
	Med	6.1%	-3.9%	5.0%
	3	21.5%	6.3%	9.2%
	Max	59.4%	35.0%	35.4%
2018	Min	-14.3%	-81.8%	-25.7%
	1	0.9%	-23.0%	-6.3%
	Med	7.0%	-9.5%	2.7%
	3	22.7%	1.0%	7.2%
	Max	143.9%	42.1%	45.1%
2019	Min	-35.1%	-103.9%	-57.3%
	1	-11.9%	-21.7%	-9.8%
	Med	1.3%	-9.3%	-2.6%
	3	17.7%	1.1%	6.7%
	Max	66.7%	25.9%	27.8%
2020	Min	-30.5%	-76.8%	-29.6%
	1	-5.9%	-11.7%	-2.6%
	Med	4.5%	3.4%	3.0%
	3	25.9%	15.9%	13.0%
	Max	123.7%	28.6%	28.1%
Total	Min	-55.3%	-103.9%	-57.3%
	1	-9.1%	-16.1%	-6.2%
	Med	5.2%	-1.8%	2.4%
	3	23.0%	8.2%	10.5%
	Max	212.6%	42.1%	83.8%

Table 17: Quartile statistics of weekly strategy; all portfolios