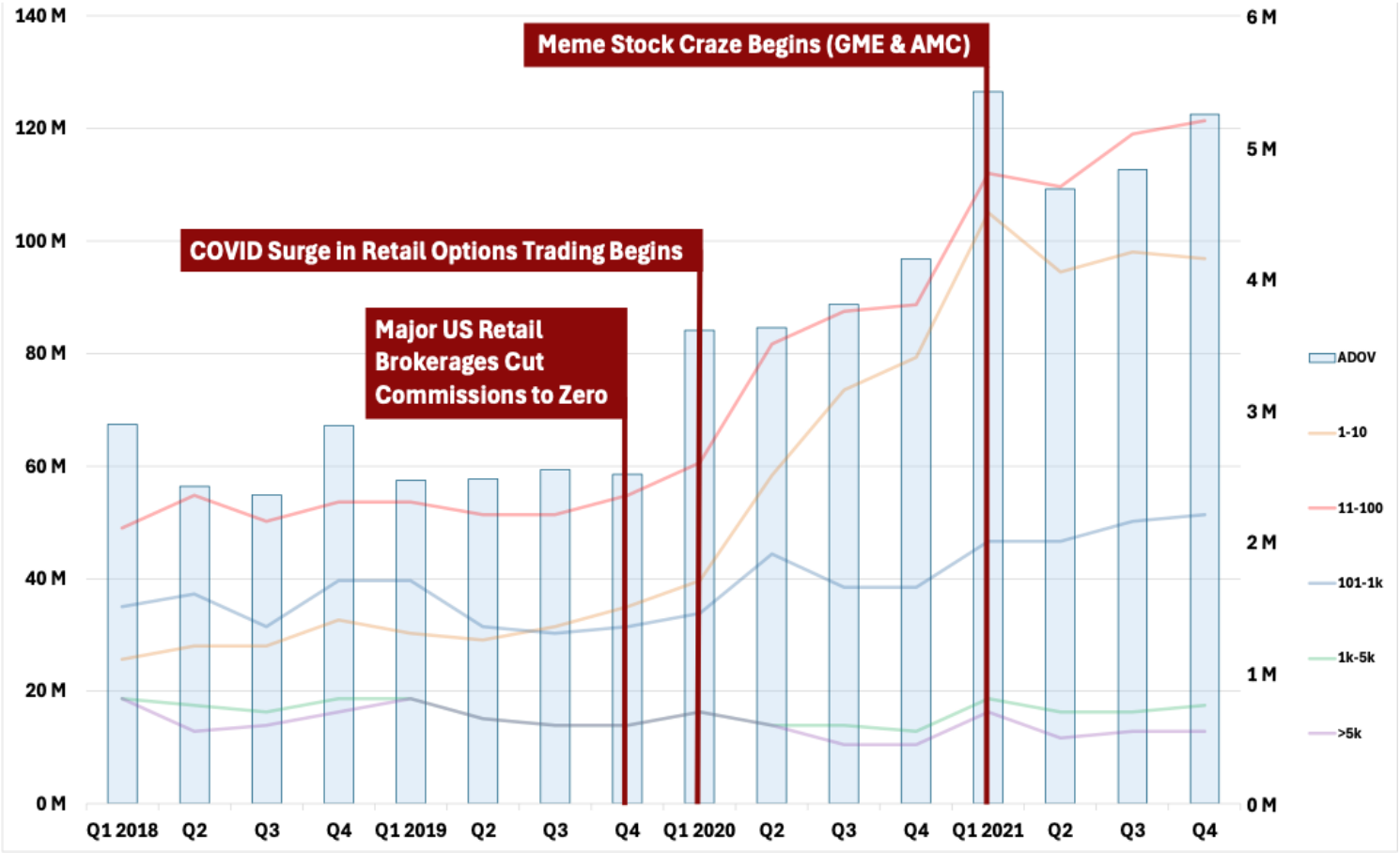
Retail Trading and Option Mispricing



FIN11 Term Paper

09.12.2024

Word Count (excl. abstract, refs., & app.): 6,328

Candidate Number: 89, 102, 120, 121, 162

Abstract

This paper explores the role of retail trading in the U.S. options market during 2020-2021 and its potential impact on option pricing inefficiencies. Retail activity surged due to zero-commission trading platforms, social media influence, and stimulus payments, resulting in significant changes in trading dynamics. Using a quantitative approach, we analysed volatility dislocations in single-stock (SS) options and developed a volatility-based trading strategy. Our methodology involved identifying inefficiencies through a Dislocation Score (DLS), comparing SS volatility risk premiums (VRPs) with sector exchange-traded funds (ETFs) as references.

We tested the strategy on the Consumer Discretionary sector, implementing delta-hedged short straddles to exploit volatility reversions. Initial results showed limited outperformance compared to a sector benchmark, but refining the strategy with an additional volatility filter improved the predictive power. The final strategy achieved significant outperformance during the examined period, uncovering an arbitrage opportunity capitalizing on retail activity and subsequent market-maker hedging behaviour, predicting volatility reversions. This research contributes to understanding how retail trading impacts options pricing and offers a framework for exploiting volatility-driven inefficiencies.

Table of Contents

Part 1: Introduction ***5***

1.1 Purpose and Hypothesis 5

1.2 What Is an Option? 5

Part 2: Increase In Retail Options Trading ***7***

2.1 Increase in US Options Market 7

2.2 Trends and Herd Behaviour 8

2.3 Short Until Expiration Date 10

2.4 Lot Sizes 12

2.5 Our Assumption and Timeline 13

Part 3: Inefficiencies In the Option Market ***15***

3.1 Volatility 15

3.2 Measuring Inefficiencies in the Option Market 16

3.2.1 VRPs – Another Probability Distribution 17

3.2.2 Identifying Retail Driven Option Mispricing 17

3.2.3 Using ETFs as Reference 18

3.2.4 Dislocation Score (DLS) 18

3.3 Applying Initial DLS 19

Part 4: Outline of Strategy ***21***

4.1 Delta Hedged Short Straddle 21

4.2 Initial Strategy and Adjustment 22

4.2.1 Reference Index 23

4.2.2 Adjusting for Differences in Volatility 24

4.3 Adjusting our Strategy 25

4.3.1 Market Makers and Gamma Squeeze 26

4.3.2 Final Strategy 27

4.3.3 Performance 29

4.3.4 Risk Adjusted Returns 29

4.3.5 Correlation and Return Attributes....30

Part 5: Limitations and Further Research....35

Part 6: Conclusion....37

Part 7: References....38

Part 8: Appendix....43

8.1 Black-Scholes Price Formula....43

8.2 Approximation to Option Prices....44

8.3 Code Explanation of Delta Hedged Straddles....44

8.4 Python Code....46

8.4.1 Delta Hedged Short Straddles....46

8.4.2 VIX Chart....53

8.4.3 Annual Option Volume....55

8.4.4 Regression Analysis....57

8.4.5 PnL Profile of a Straddle....58

Part 1: Introduction

When there are a significant number of irrational traders participating in the market, it might result in a deviation between fair value and price of the underlying asset. If correctly understood these deviations can be exploited for profit. Such an opportunity might have unfolded due to what seems like an increase in option retail trading during 2020.

1.1 Purpose and Hypothesis

The purpose of this paper is to explore how increased retail trading activity may have influenced the pricing of SS options in the U.S. market during the surge in retail participation from 2020 to 2021. Retail traders, often engaging in speculative behaviour and lacking extensive market expertise (Eaton et al., 2021), might contribute to inefficiencies in option pricing. By examining these potential inefficiencies, we aim to assess whether they could present opportunities for a volatility-based trading strategy.

Our hypothesis is that the influx of retail traders into the options market may have led to elevated VRPs in certain SS options, suggesting deviations from fair value. These dislocations might be driven by speculative order flows, disproportionate activity in specific stocks. By comparing SS options to sector-level benchmarks, it may be possible to identify and exploit such inefficiencies. Through our research, we aim to answer the following research question:

How could an investor effectively exploit retail-driven option mispricing from 2020 to 2021 using a volatility-based trading strategy?

To answer our research question, we will build a foundation attributing option volumes to retail activity. We aim to analyse how retail trading activity might have had impacts on the VRPs and option prices for certain sectors and stocks. Subsequently we will develop and backtest different strategies monetizing the activity.

1.2 What Is an Option?

An option is a financial derivative that provides the holder with the right, but not the obligation, to buy (i.e. call) or sell (i.e. put) an underlying asset at a predetermined strike price,

on or before a specific expiration date (Harris, p. 42). For this paper, we are looking exclusively at European style options, which only allows the holder to exercise the option on the date of expiration. Furthermore, options status relative to the market price is classified as in the money (ITM) if exercising yields a profit, at the money (ATM) when strike and market prices align, and out of the money (OTM) if unprofitable (Kenton, 2023).

If financial derivatives are weapons of mass destruction, options are weapons of mass precision. By using options, option strategies or structured options, the investor can express exactly how and when they believe the price of a security will develop. The investor may for example pay a premium to bet that a security moves to a specific price range, within a specific time range, in effect creating a disproportionate payoff if he is right (i.e., synthetic leverage). Equally, options can be used to reduce or hedge risks.

Options are often priced using the Black-Scholes model, which provides a theoretical framework for calculating the fair value of European options (see appendix).

Part 2: Increase In Retail Options Trading

A retail trader is an individual who buys and sells financial instruments, such as stocks, bonds, or derivatives, for their personal account rather than on behalf of an institution (Zucchi, 2023). Unlike professional traders, retail traders may lack advanced tools, research capabilities, or access to large-scale trading data, which can influence their trading decisions and strategies. Importantly, anyone over the age of 18 can, after a few clicks, join an online platform and start buying and selling, becoming a retail trader. In terms of strategy and competence retail traders are a diverse group – however, due to the low barriers to entry and lack of guardrails many becomes unprofitable, to the benefit of better-informed traders (Harris, p. 576).

2.1 Increase in US Options Market

Starting in 2020, the U.S. annual options market experienced a surge in option trading volume - doubling from 2019 to 2021.

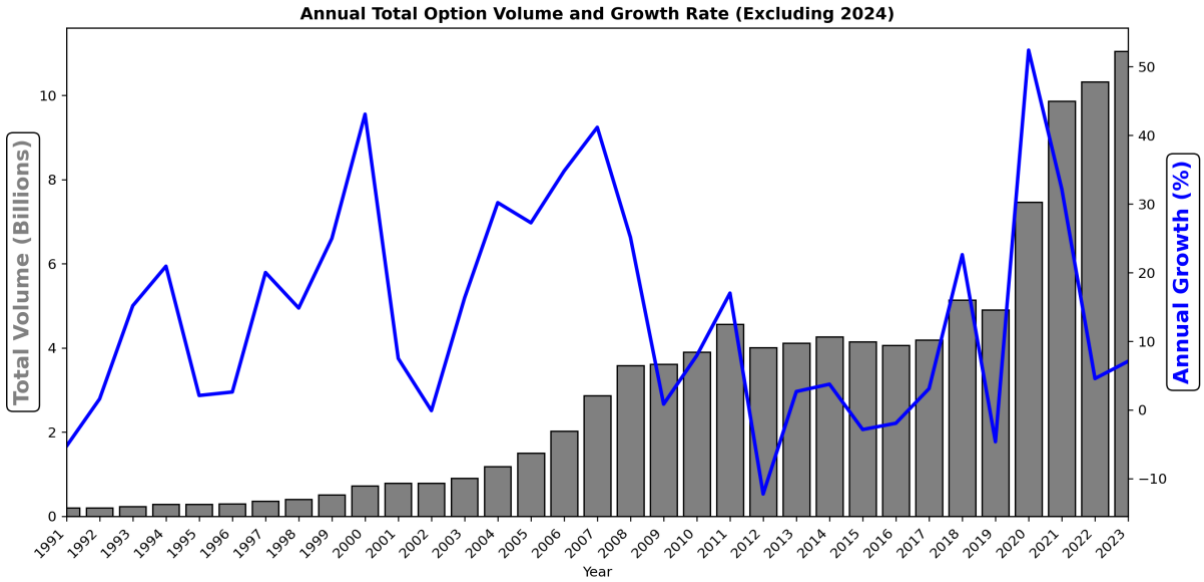


Figure 1: Annual total option volume and growth rate (excluding 2024). The chart highlights a consistent increase in option trading volume over time.

Data source: OCC (2024).

Meanwhile, volumes in trade index options remained relatively stable in this period (see Figure 2).

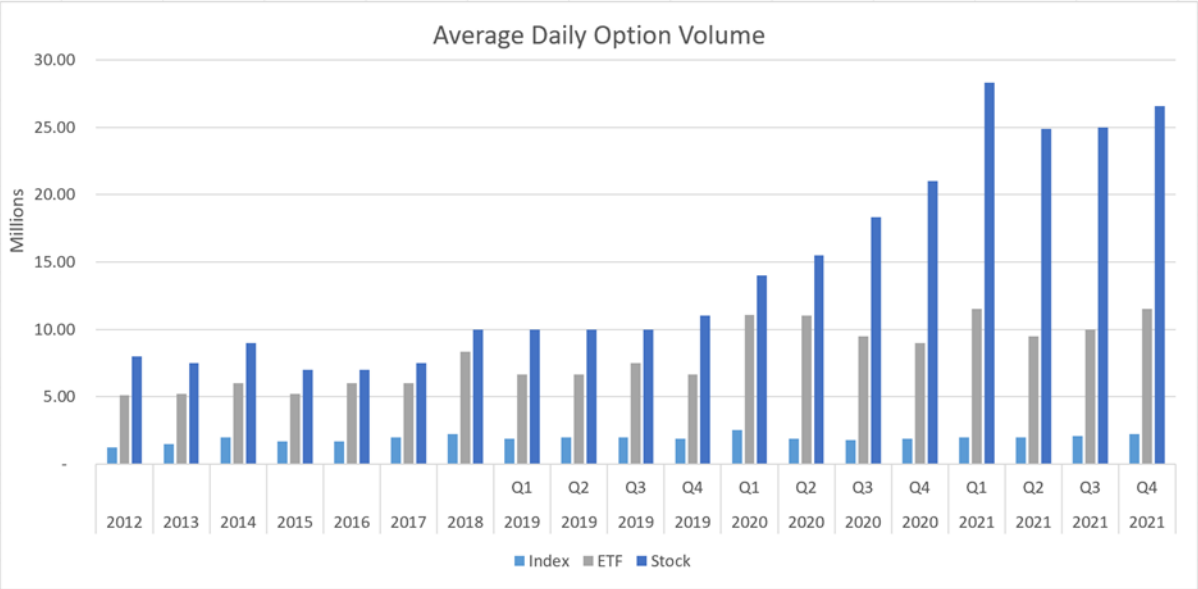


Figure 2: Average daily option volume (ADOV) from 2012 to Q4 2021.

Data source: CBOE, (Schwartz, 2022).

The year 2020 was volatile for the stock market, following economical outcry after the beginning of the pandemic. This might relate the new option volumes to hedging activities and normal speculative market behaviour following a volatile market. However, the massive and sustained increase in SS option volume indicates a fundamental change in option market participants. We have found multiple factors that support attributing the new volume to retail activity.

2.2 Trends and Herd Behaviour

As humans, we are heavily influenced by our peers and often let “herd behaviour” influence our decisions, especially when we are not experts in a field (Shiller, 2015, p. 14). This is prevalent amongst many retail traders who often participates and are influenced by recent trends, sometimes resulting in questionable investments. Building up to the increase in options trading there were many events that may have contributed to new retail investors and participation in the financial markets.

On 2nd October 2019, major retail brokerage firms introduced commission-free trading, allowing retail investors to execute trades without direct fees (Osterland, 2019). By eliminating cost barriers, zero-commission trading opened financial markets to a broader audience, enabling more people to participate actively. This is illustrated by the significant increase in user accounts at the retail brokerage Robinhood.

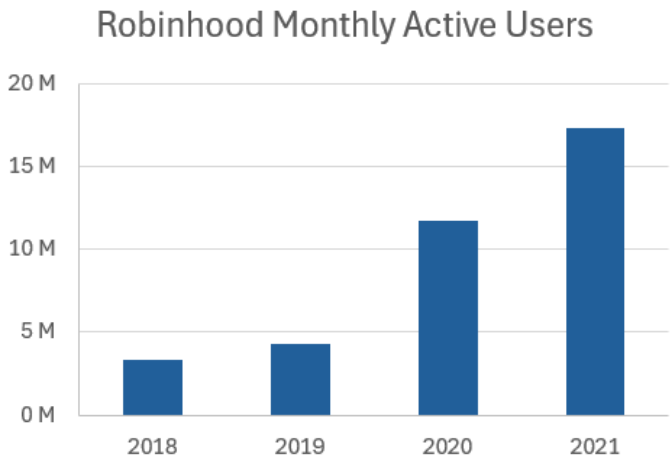


Figure 3: Robinhood monthly active users (MAU) from 2018 to 2021.

Data source: Robinhood Annual Report 2021 (2022) & Registration Statement (2021).

However, this model relies heavily on payment for order flow (PFOF), where brokerages route orders to market makers in exchange for compensation. While this enables "free" trading, it may compromise execution quality, as orders are often prioritized based on market maker payments rather than the best price for the investor. This practice can lead to slightly higher buy prices and marginally lower sell prices, representing a hidden cost to retail investors (Farley, 2022).

This shift to zero-commission attracted a wave of less experienced traders, primarily younger, risk-tolerant individuals with less capital who were focused on short-term gains (Eaton et al., 2021; Koh, 2021). For these new retail participants, minimizing trade costs is essential, making zero commissions a priority over securing the best possible price. As Harris notes, "*Retail traders pay more attention to low commissions than to good prices because you cannot buy anything that you cannot measure"* (Harris, p. 517).

The COVID-19 pandemic appears to have catalysed a resurgence in retail trading, with several contributing factors. Remote work arrangements afforded individuals more time to engage in financial markets, while government stimulus payments provided additional capital, some of which found its way into trading accounts. As MIT Professor Eric So described it:

"*Folks Were at Home looking for something to do*" (Vereckey, 2022). These conditions created a unique environment for individual speculative activity to flourish.

Additionally, social media platforms and online communities like Reddit's WallStreetBets amplified trading ideas and fostered a sense of collective action (Deloitte, 2021). Notably, 23% of Gen Z and 20% of Millennials reports trusting Reddit as an investment source (Caporal, 2023). This dynamic contributed to the "meme stock" phenomenon, where coordinated retail trading often leveraging options drove dramatic price surges in stocks such as GameStop (GME) and AMC. The heightened market volatility presented attractive opportunities for substantial short-term gains, drawing new participants eager to capitalize on rapid price movements. A noticeable shift in behaviour among younger investors, particularly millennials and Gen Z, suggests a growing preference for high-risk, short-term strategies shaped by social media narratives and the allure of quick profits. This shift diverges from the traditional long-term investment approach, reflecting how a combination of zero-commission trading, COVID-19 conditions, stimulus payments and social media influence transformed the retail trading landscape.

2.3 Short Until Expiration Date

Everybody wants to get rich quickly and live a life full of excitement and freedom, which is why some try their luck in the market. However, as Warren Buffet points out:

"The stock market is designed to transfer money from the active to the patient." - Warren Buffett, (Thomas, 2020)

Many of the new retail participants, unfortunately for their own good, do not have the patience and often operate with the "get rich quick" mentality (Hawkins et. al, 2022). Options with shorter expirations and OTM strikes have lower premiums because they are less likely to reach intrinsic value before expiration. Their reduced time value and low probability of significant price moves make them cheaper, appealing to those seeking quick, speculative bets. As the Head of Trading and Derivatives at Charles Schwab stated: "They like the fact that it is a binary event, either they are right, or they are wrong, and they figure that fairly quickly"

(Mazzola, 2024).

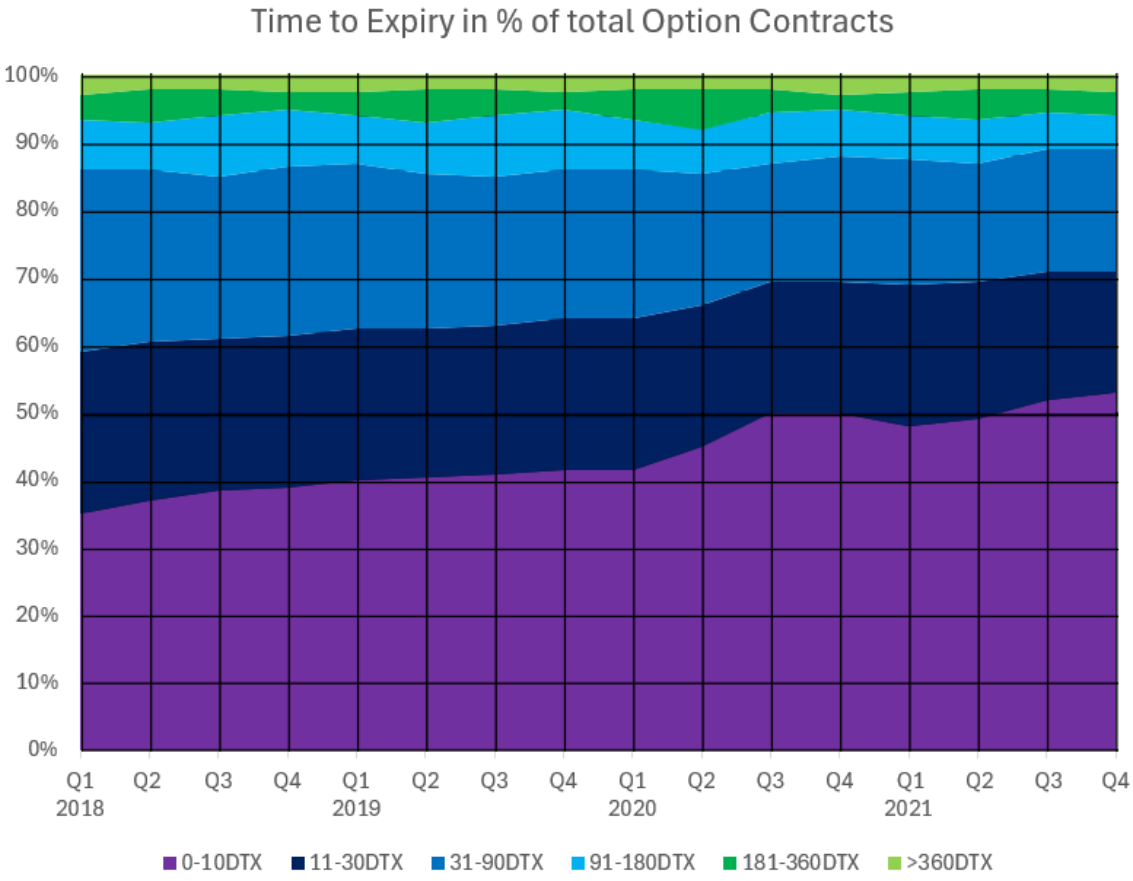


Figure 4: Time to expiry in % of total option contracts. The increase in SS options volume during this period was primarily driven by trading in short-dated options.

Data source: CBOE, (Schwartz, 2022).

At the end of 2020, almost 50% of the options traded had an expiration date between 0-10 days, and another 17% had an expiration date 11-30 days. It is important to note that short-dated options aren't only used by gamblers and are utilized by institutions for hedging and speculation in event driven strategies, like upcoming earning rapports. This trend alone does not stand as very strong evidence for retail driven options volume because of the uncertainty related to the practices of other market participants. Still, the development seems to be in line with the gambling nature of the retail investor and can be understood as an indication of increased retail participation in the option market.

2.4 Lot Sizes

Retail traders operate with less capital compared to other market participants. The average account size on Robinhood serves as a sufficient illustration of this, with the capital only being in the thousands of dollars.

Year	Average Account Size
2018	$2,533
2019	$2,772
2020	$5,038
2021	$4,317

Figure 5: Robinhood average account size (2018-2021).

Data source: Robinhood Annual Report 2021 (2022) & Registration Statement (2021).

If the wave of new retail traders actively traded options, their smaller account sizes would likely lead to trading in lower cash volumes. This would be evident in market data, particularly through smaller lot sizes.

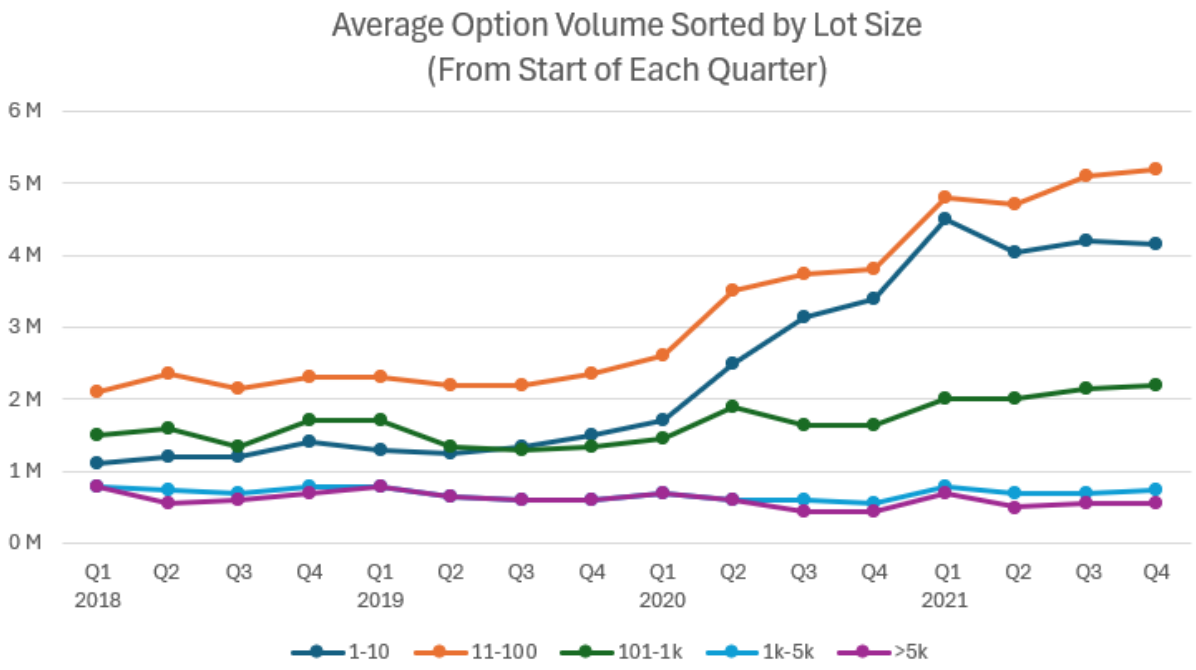


Figure 6: Average option volume (AOV) sorted by lot size from Q1 2018 – Q4 2021.

Data source: CBOE, (Schwartz, 2022).

Late 2019 to 2021 displays a steady upward trend in smaller lot sizes traded, especially those between 1-10 and 11-100. Given the fact that retail traders prefer smaller dollar-amount trades, the data provides further indication of retail activity. Additionally, the trend complements the upticks in "short until expiration" options traded, suggesting the increase in options volume are due to more speculative small sized option plays from retail investors.

2.5 Our Assumption and Timeline

It seems fair to assume that the changes in lot sizes and increase in short-dated options traded largely are due to retail investors, speculating in SS options. Looking at the timeline of events and developments with regards to the usual patterns of retail investors further supports this assumption.

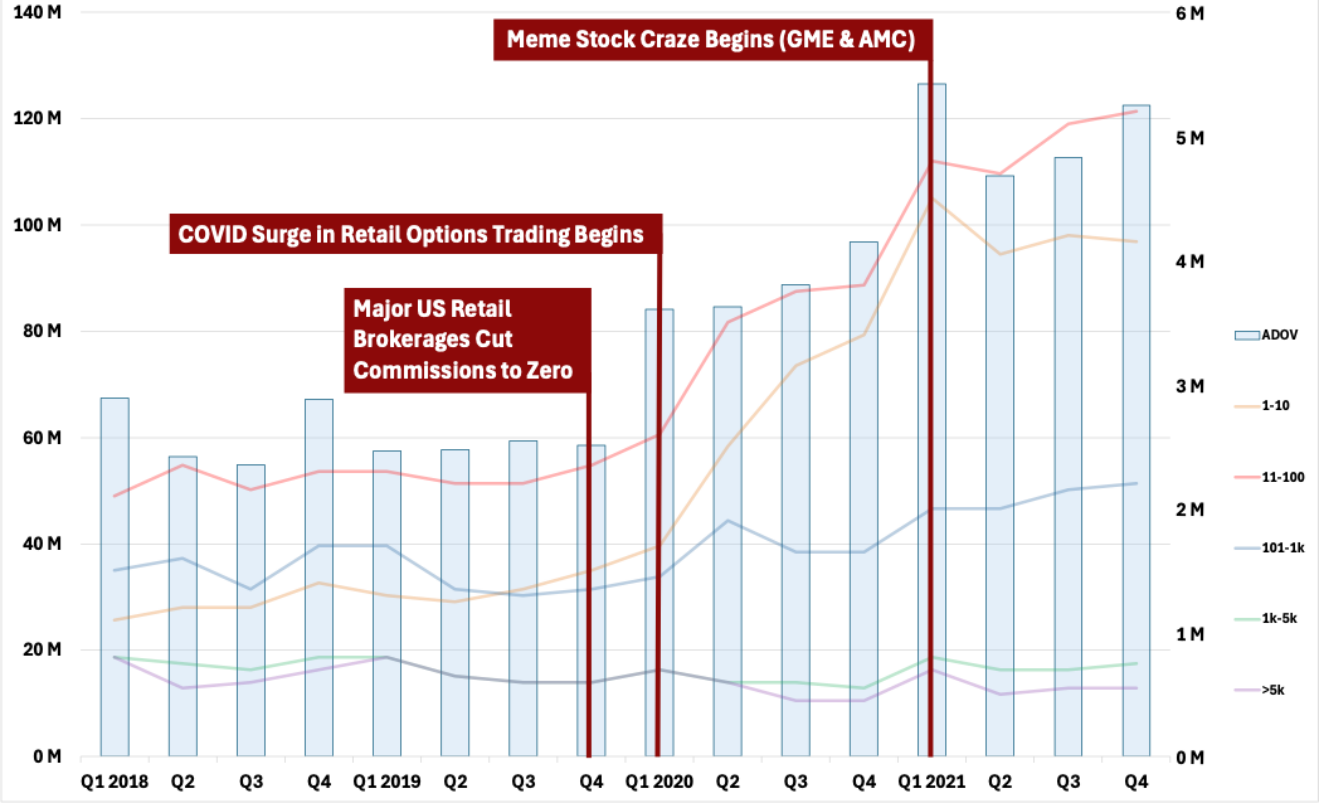


Figure 7: ADOV and lot size amid key market events from Q1 2018 to Q4 2021.

Data source: CBOE, OCC (2024).

We do however want to point out that our findings can potentially be attributed to other market participants, such as larger traders camouflaging themselves by executing smaller trades or automated trading (Schwartz, 2022). Regardless, we consider our research as strong indications of increased retail activity.

Part 3: Inefficiencies In the Option Market

Building on the assumption that increased retail activity has contributed to much of the SS option volume, this might have impacted the pricing of the options, potentially leading to exploitable inefficiencies. To explore this further, the following section establishes a theoretical framework surrounding options theory and inefficiencies. This framework will later be applied to identify and exploit potential inefficiencies in the options market caused by retail-driven activity.

3.1 Volatility

Volatility plays a crucial part in option theory. Options are priced based on the markets expectations of future volatility, which often referred to as implied volatility (IV) in the Black-Scholes model (Hayes, 2024). It is a measure of price fluctuation in an asset, regardless of direction. High volatility indicates large, rapid price changes, while low volatility suggests more stable, consistent movements.

Realized volatility (RV) is a term used when measuring these fluctuations in a set past time duration (e.g. last 30 days) and can be adjusted depending on what the data is to be used for. It is usually calculated using end of day (EoD) price data for the underlying asset. It can also be calculated using continues datasets, but some investors and researchers like us are focused on longer durations, which is one of the reasons why we use RV calculated based on EoD.

The standard formula for calculating RV aims to achieve both practical and theoretical goals. The first part of the equation catches the price movements from day to day.

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

R = *Return*

P = *Price of the underlying asset*

t = *Time*

The reason why logarithmic returns are used is because it treats both gains and losses symmetrically with regards to an increase or decrease in price. If a stock gains 100-dollar points

one day, and loses 100 the next, log returns treat the percentage change in the stock price the same.

A stock moving 5 dollars on Monday and keeps a stable price for the next four days of the week, is more volatile than a stock that increases evenly with 1 dollar every day. The formula reflects this by calculating the standard deviation of each return.

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R_i - \overline{R})^2}$$

$$\overline{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$$

From the equation, the stock with even increases would return zero, while the one with the rapid increase would return four. It factors in growth in the underlying asset and is why it is used for analysing fluctuations in longer time durations and comparing stocks with both negative and positive growth.

Lastly, the formula annualizes the volatility using the number of trading days as a measure for further standardizing the volatility. This is also important for being able to compare volatility with the output of the Black-Scholes pricing formula which usually returns implied volatility (IV) in an annualized matter. This is simply done by multiplying it with the square root of trading days in a year (252).

$$Annualized\ Volatility = \sigma \cdot \sqrt{T}$$

3.2 Measuring Inefficiencies in the Option Market

The difference between IV and RV is referred to as the volatility risk premium (VRP). A high VRP suggests that the option was overpriced relative to RV, while a low VRP indicates under-pricing (McGuane, 2018). It is important to note that the Black-Scholes model, while central to understanding fair option pricing through volatility, does not directly address whether an option will be profitable. For example, buying a call option with a strike of 100 and an IV of 30% will only be profitable if the underlying price of the asset at expiration exceeds 100, even if RV is exceptionally high.

The VRP metric is widely used to evaluate pricing inefficiencies in the options market because volatility inherently measures price fluctuations rather than direction (Hayes, 2024). The calculation of standard deviation eliminates the influence of directional returns, focusing solely on the magnitude of price movements. Similarly, IV derived from Black-Scholes reflects the expected deviation from an asset's mean return, irrespective of whether the price moves up or down. This symmetry ensures that volatility remains a neutral measure of risk, independent of expected returns.

3.2.1 VRPs – Another Probability Distribution

A positive VRP suggests that IV exceeded RV, often reflecting a risk premium demanded by investors. However, evaluating whether an option was truly mispriced requires understanding the market's probabilistic assumptions about future volatility.

In event-driven trading, investors may speculate on the outcome of a corporate deal. If the probability of success is only 1%, this expectation is likely priced into the asset's options and underlying stock. In such cases, VRP for unsuccessful outcomes might be slightly positive (i.e. overpriced), while for the rare successful case, VRP could appear extremely low (i.e. underpriced). This reflects market efficiency, as the average VRP across scenarios aligns with probabilities rather than deterministic outcomes.

In efficient markets, the expected VRP is not zero but typically positive due to the risk premium for bearing uncertainty. Analysing VRPs, along with underlying factors and market dynamics, can reveal pockets of inefficiencies where options may be systematically mispriced. Such opportunities often arise in scenarios where market expectations or risk perceptions diverge from reality.

3.2.2 Identifying Retail Driven Option Mispricing

In 2020 the new option volumes were mainly distributed over three sectors: consumer services, information technology and consumer discretionary (Deshpande et al., 2020). If option markets failed to adapt to these new volumes, stocks in these sectors would experience option mispricing reflected in elevated VRPs, especially on certain SS as this volume likely was not distributed equally throughout all the stocks in the sector.

VRPs reflects market efficiency based on a probability distribution and expected outcome. In theory this is impossible to obtain. To indicate mispricing in the underlying asset, it seems reasonable to use another asset and its option market a reference. If the two share similar characteristics in fundamentals and catalysts, they can potentially be utilized to isolate the effects of different retail option volumes and then estimate potential inefficiencies in option prices.

3.2.3 Using ETFs as Reference

We chose to use ETFs representing these sectors as benchmarks for analyzing each relevant stock. The idea is that a stock should to some extent be correlated with its sector ETF. Since the ETFs is a collection of many stocks within that sector, it seems reasonable to use a stock's sector ETF as a reference point to determine what should be a reasonable VRP. The sector ETF cannot be 100% correlated with the SS. Subsequently, IV and RV can deviate a lot for a stock within, and a stock can be impacted by factors only applying to them, resulting in an excessive VRP. This should be taken into consideration when comparing the two.

3.2.4 Dislocation Score (DLS)

To use an ETF as a tool for predicting elevated VRPs, it is important with a standardised approach to compare the ETF and the stock. There are multiple different time frames with regards to realized and RV, but because options usually are issued monthly, we decided to look at the 30 days VRP ratios. The VRP ratio can be defined as:

$$VRP\ Ratio = \frac{IV}{RV}$$

Our initial DLS aimed at deciphering elevated VRPs, regardless of option volumes and underlying volatility. This approach can potentially be sufficient in isolating price inefficiencies if retail traders pushed option prices resulting in sustained elevated VRPs.

We calculated the DLS by subtracting the ETF's VRP ratio from the stock's VRP ratio. A high score would by our logic indicate irrationally elevated VRPs, potentially as a result from retail activity.

$$DLS = VRP\ Ratio_{Stock} - VRP\ Ratio_{ETF}$$

However, the underlying VRP of a SS can be higher or lower than the ETFs due to the stock being expected to be more volatile. A high VRP does not inherently mean inefficiencies (e.g. earning reports). Therefore, we set a threshold of > 0.5, which we deemed as a sufficiently high DLS to determine the option price as potentially being inefficient. We believed this threshold would filter out most normal fluctuations while keeping many elevated VRPs, at least during certain market conditions. This however can be adjusted, and other modifications to the DLS may be necessary to isolate inefficiencies, something we did in our final strategy.

To test our initial DLS, we started looking at the stocks in the communications services sector using the ETF XLC. The ETF contains 25 companies with different weighting and includes some big companies like GOOGL and META. As part of the sample data, we decided on using stocks with less weighting in the ETF, and analysed the seven stocks: CHTR, CMCSA, DIS, T, TMUS, TTWO, VZ.

3.3 Applying Initial DLS

We tested our DLS by investigating the stocks between the timeframe 7th August 2020 to 22nd November 2020. The DLS is a metric showing potential inefficiencies. However, to integrate the metric in a volatility-based option strategy, it is essential to look at the predicative property of the metric. If the DLS displays an inefficiency (i.e. > 0.5), it deems the option to be mispriced. We analyse the predicative property by measuring the subsequent VRP ratio after an inefficiency displayed by the DLS. This uses IV as an income and the RV as a cost , in line with Black-Scholes volatility-based valuation of options.

$$\boldsymbol{Premium\ collected\ (PC)} = 1\ for\ every\ inefficiency\ displayed\ by\ DLS$$

$$\boldsymbol{VRP\ Ratio} = \frac{IV}{RV}$$

$$\boldsymbol{Cost\ of\ volatility\ (Cov)} = \frac{PC}{VRP\ Ratio} = \frac{1}{VRP\ Ratio}$$

$$\boldsymbol{Return} = PC - Cov = 1 - Cov\ [represents\ subsequent\ return]$$

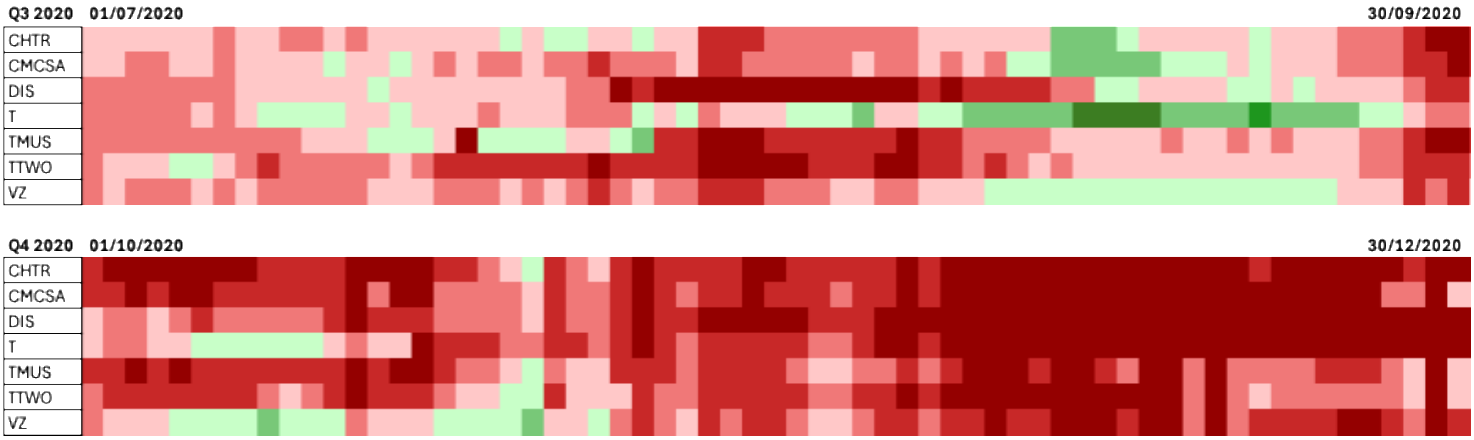


Figure 8: Heatmap of DLS for CHTR, CMCSA, DIS, T, TMUS, TTWO, VZ from Q3 2020 – Q4 2020. Green means a DLS greater than 0.5 while red means smaller than 0.5.

The heatmaps indicate that during this period, the DLS predominantly favoured two specific stock and their options: AT&T (T) and Verizon (VZ). Further analysis of how the individual stocks performed with the DLS based VRP returns indicates an ability to isolate inefficiencies in the stocks VRPs, at least for that specific period.

Ticker	Number of ineff.	Cost of vol.	Return
TTWO	2	2.05	-0.05
CHTR	6	6.44	-0.44
CMCSA	11	10.53	0.47
VZ	30	19.87	10.13
DIS	5	3.67	1.33
TMUS	1	0.68	0.32
T	33	22.05	10.95

Figure 9: Table of inefficiencies identified, the associated cost of volatility, and the return for each selected ticker. Highlighted rows represent stocks with notably higher inefficiencies and corresponding returns.

On these two stocks the RV during the period 7th August 2020 to 22nd November 2020 was only about 66% of the IV after an inefficiency displayed by the DLS approach. They were also displayed as inefficient significantly more frequent than the other stocks further emphasising the potential accuracy of the approach. The DLS seems to potentially hold some predicative value as a metric for future inefficiencies.

Part 4: Outline of Strategy

During the period 7th August 2020 to 5th October 2020 there seems to have taken place inefficiencies in certain SS options, potentially driven by heightened retail trading activity. In our strategies we aim to isolate stocks experiencing elevated VRPs, going short volatility to profit from lower RV relative to IV. We decided to look at the Consumer Discretionary sector instead of Communication Services, due to it containing more stocks in the S&P500 as well as being one of the three sectors gaining the most option volume in 2020 (Deshpande et al., 2020).

4.1 Delta Hedged Short Straddle

A short straddle is an options strategy that involves selling both a call option and a put option with the same strike price and expiration date. This strategy is typically used when an investor expects minimal movement in the underlying asset price. The trader profits when the directional move fall short of the dollar value received by the premium collected (Chen, 2024).

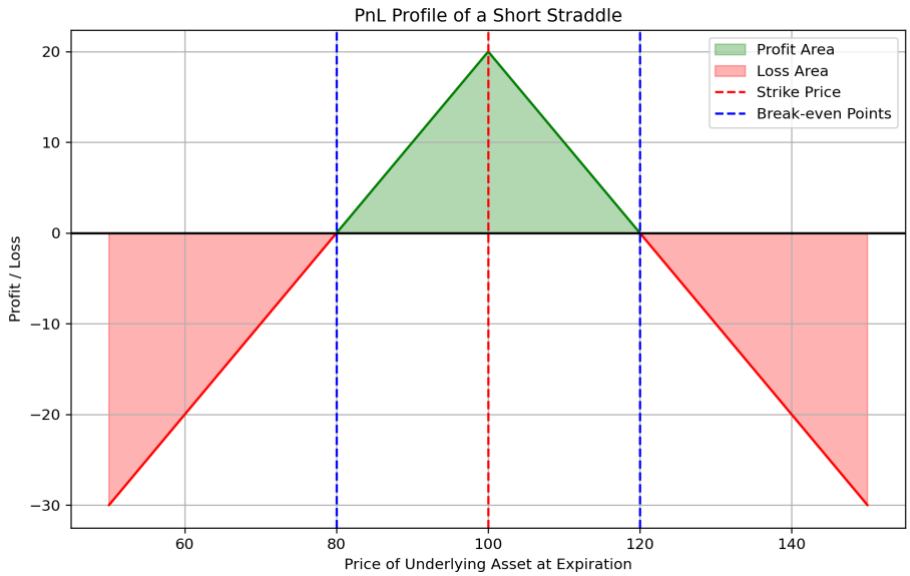


Figure 10: The figure shows the Profit and Loss (PnL) profile of a short straddle options strategy with an IV of 0.2 and a strike price of 100.

Delta hedging a short straddle aims to reduce directional risk and maintain a delta-neutral position. This is achieved by buying or selling the underlying asset as the net delta of

the straddle position changes. In theory, the strategy will now be profitable as long as IV is greater than RV (Gavin, 2020). However, it is impossible to constantly buy and sell the underlying asset staying delta neutral. Due to factors such as transaction costs, traders limit the frequency of their delta hedging. In our strategy, we will sell ATM short straddles and delta hedge the stocks daily at EoD to profit from elevated VRPs.

4.2 Initial Strategy and Adjustment

We back tested multiple strategies over the course of 10 years. Initially, at the beginning of each month, we applied our DLS to identify the 5 stocks with the highest one-month DLS, and back tested selling daily delta-hedged short straddles to capitalize on elevated IVs. Selling the top 5 performing stocks is just one of many approaches, and fewer stocks could potentially introduce additional risk as well as higher returns, depending on a specific company. Furthermore, we decided to weigh the straddles sold based on the underlying stocks DLS divided on the sum of each of the five companies DLS in a given month:

$$w_i = \frac{DLS_i}{\sum_{j=1}^{5} DLS_j}$$

This introduces less diversification influencing returns but was a measure to be more selective in which straddles are being sold.

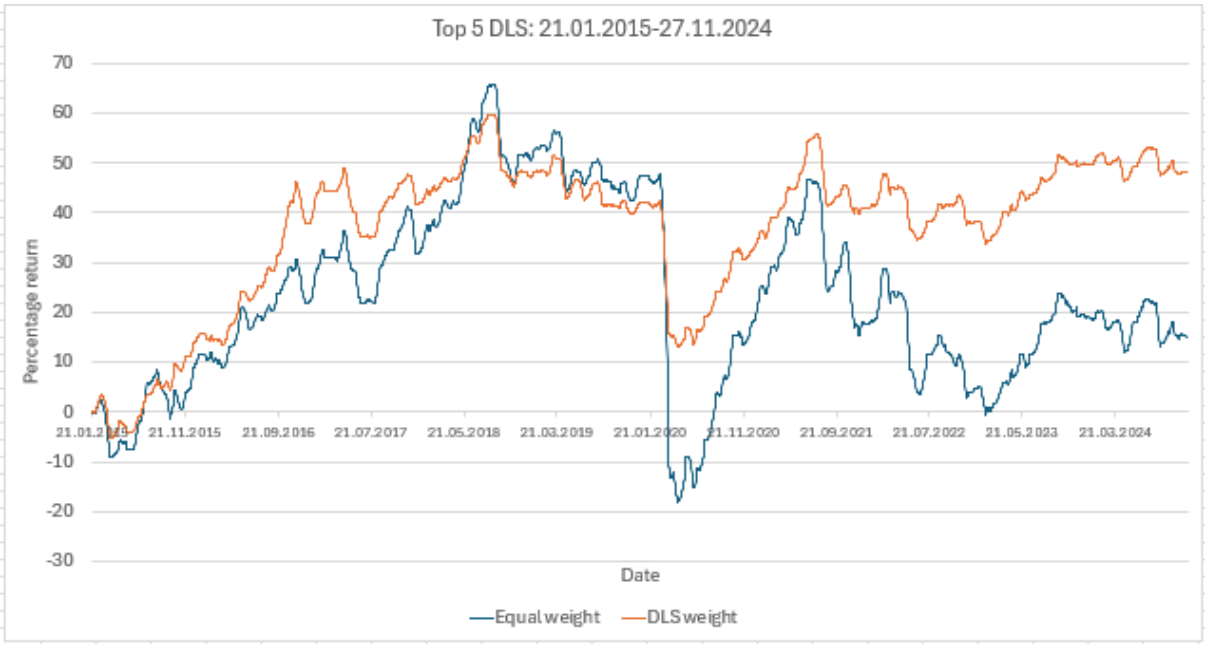


Figure 11: Weighted and unweighted returns for initial strategy for top 5 stocks from 21st January 2015 to 27th November 2024.

4.2.1 Reference Index

To measure if a strategy isolates inefficiencies rather than just following the general market trend, we compare our results to the returns of selling straddles on the sector ETF; XLY (Consumer Discretionary ETF).



Figure 12: XLY and weighted returns.

The difference in volatility influences the returns and initially makes it hard to compare the performances.

4.2.2 Adjusting for Differences in Volatility

There are many different approaches to adjust for volatility, and there is not necessarily a correct answer. Aiming to analyse the volatility adjusted expected returns of our strategy compared to the sector, we adjusted our strategies PnL based on the HV relationship between our strategy and returns on XLY straddles. This may appear counterintuitive, as the 5 stocks in our strategy were updated monthly. However, when we looked at the data for which stocks end up with the highest DLS, they tend to stay top performers up to a year. Therefore, we deemed it as sufficient for a comparison between the two when evaluating if our initial strategy exploited inefficiencies. We calculated a one-month-volatility-ratio dividing the top 5 DLS stocks rolling 30 days volatility by the 30 days XLY rolling volatility.

$$Volatility\ Ratio = \frac{\sigma_{Top\ 5}(30)}{\sigma_{XLY}(30)}$$

Further, we multiplied this ratio with the DLS graph monthly, establishing how much of our stocks' straddles should be sold.

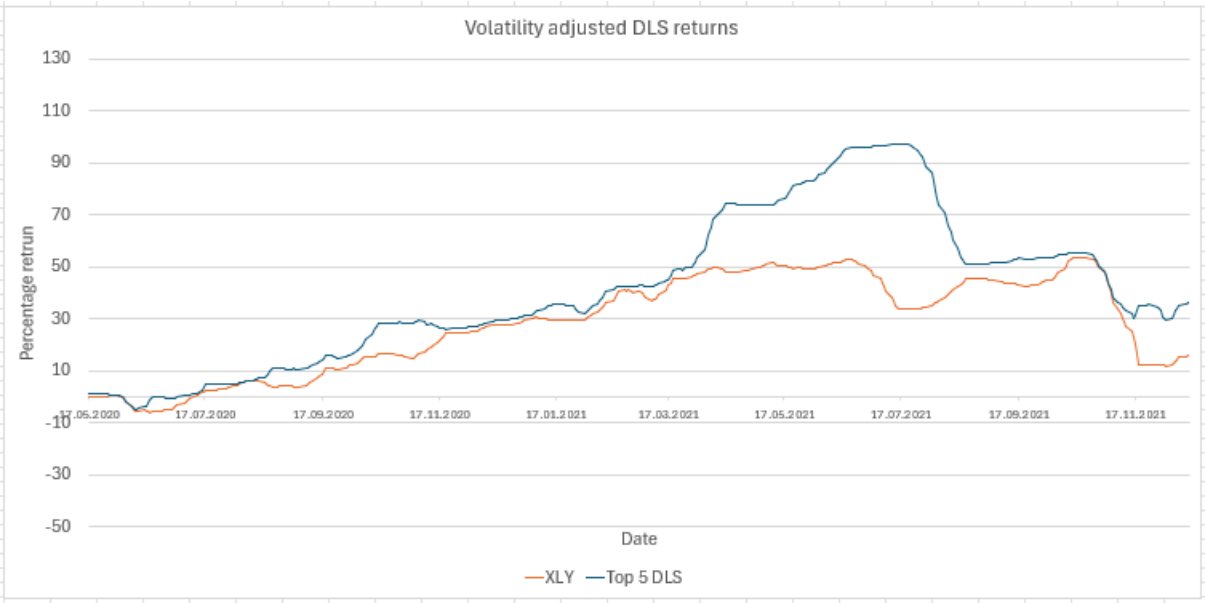


Figure 13: Volatility adjusted DLS strategy returns and XLY returns.

The strategy was mostly parallel to the reference index, and did not significantly outperform XLY. The excess returns further illustrate the in-line performance.

$$Excess\,return = R_{DLS} - R_{XLY}$$



Figure 14: Excess returns between our initial DLS strategy and XLY.

This was a time where retail traders were prevalent in the market, suggesting that the companies picked out with the DLS approach did not predictably exploit any inefficiencies. The options were mostly fair priced on the companies picked out, at least relative to the sector.

4.3 Adjusting our Strategy

To improve our strategy, we analysed other factors that might predict elevated VRPs. Our initial DLS was not consistently predicative of inefficiencies by itself. This means that the market priced our options efficiently, and the strategy did not yield sufficient excess returns. However, there is one potential way the influx of retail traders might have created price inefficiencies.

4.3.1 Market Makers and Gamma Squeeze

During the increase in option trading, SS trading volume increased significantly as well. Additionally, the stability of stock trading volume was extremely volatile following the pandemic.

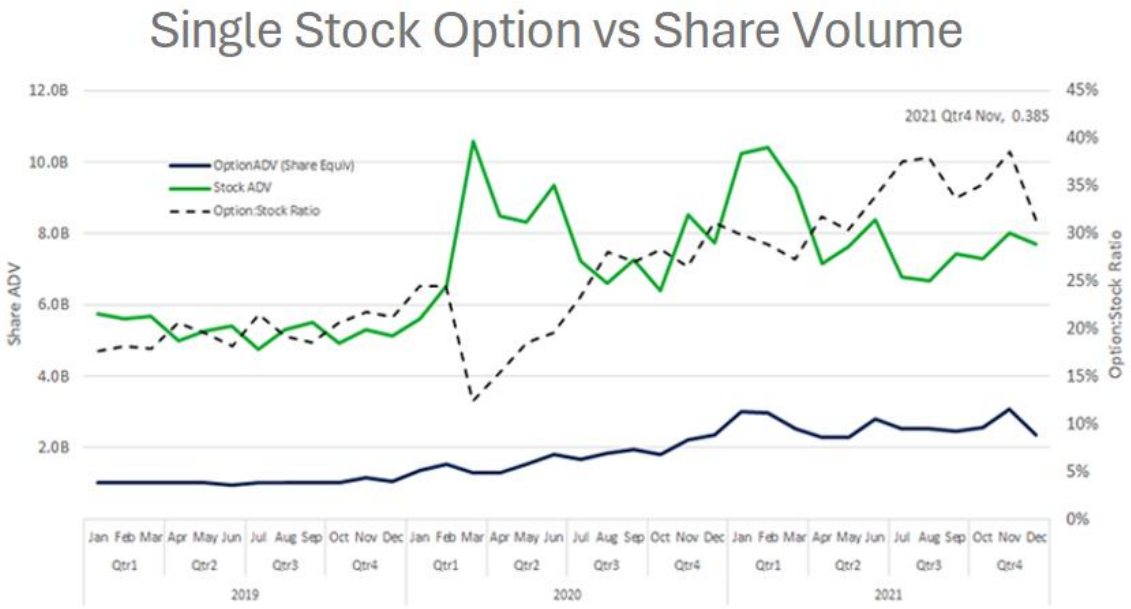


Figure 15: SS option vs share volume from 2019 to 2021.

Source: CBOE, (Schwartz, 2022)

When market makers trade options, they receive an order flow of buy and sell orders. If they sell both a put and call option, they receive a relatively neutral net delta making their positions easier to hedge. This makes their positions and hedging activity similar to our strategy of selling straddles and delta hedging.

When the order flow is not balanced, they end up with one-sided option positions making their net delta higher. If they sold a lot of calls, they now must buy a lot of the underlying asset to achieve delta-neutrality. Subsequently when the options expire, they sell their position as the options reaches expiration (Rocco, 2024).

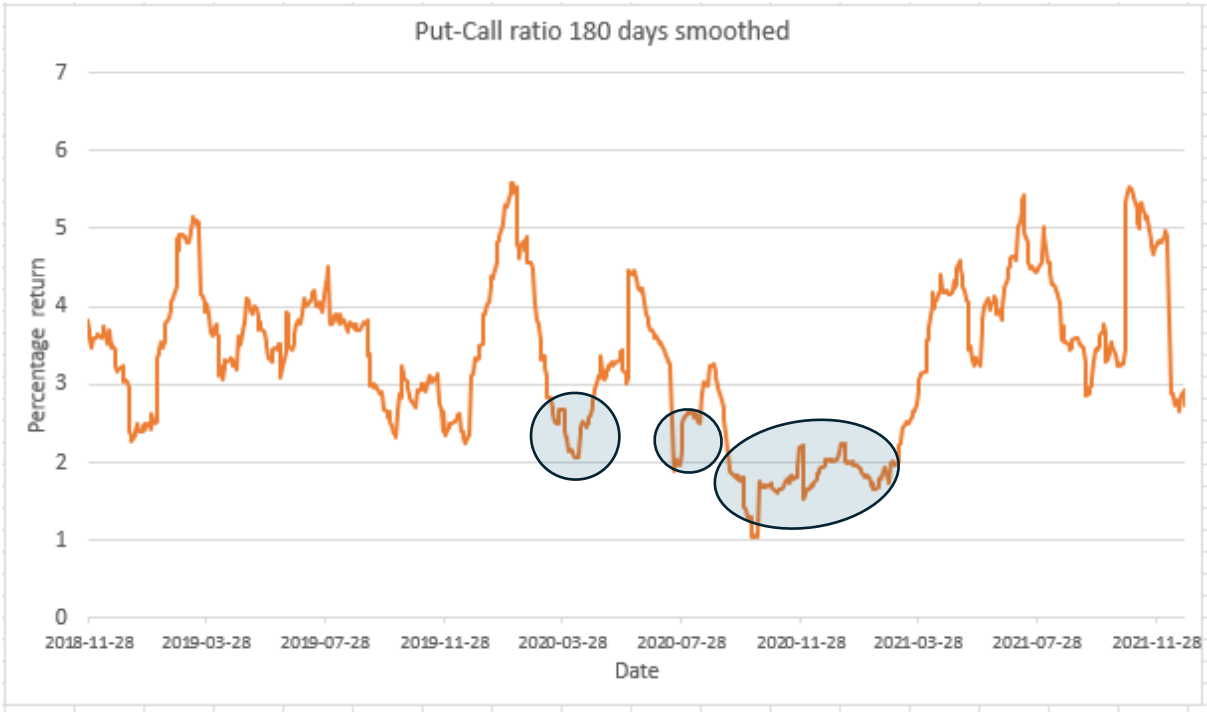


Figure 16: Put-Call ratio 180 days smoothed.

Data Source: VolVue (2024)

The SPY experienced multiple periods of low put/call volume ratios, suggesting an imbalanced order flow. High option volume combined with low put/call ratio may have resulted in temporary volatility dislocations. Market makers might have been buying and selling the underlying stocks to an extent that it significantly influenced the volatility of the underlying stock price. The closer these options are to expiration date, the more sensitive their positions become to price fluctuations. This forces them to adjust their positions frequently, resulting in a gamma squeeze (Rocco, 2024).

If the option volume received from retail traders or other market participants were significant enough to impact the volatility of the underlying stock price, the volatility should revert after an expiration date. If the market failed to price in the subsequent predictable volatility reversion, it would be an exploitable market inefficiency.

4.3.2 Final Strategy

The initial DLS itself did not find stocks with exploitable inefficiencies. However, it is still an important piece of information as it shows high VRPs relative to the sector. To single out stocks with elevated RV, potentially from hedging activity, we implemented an additional filter

to our strategy. We still use the sector ETF XLY as a reference by subtracting XLY's 30 days RV from the underlying stocks 30 days RV.

$$DLS = VRP\ Ratio_{Stock} - VRP\ Ratio_{ETF}$$

Additional

$$RV_{Stock} - RV_{ETF} > 20\%$$

Instead of continuing with our top 5 DLS approach, we decided on setting thresholds for the variables instead. This can be done in many ways, but we decided to keep our threshold of 0.5 for the DLS, and filter for stocks with RVs of 20% after subtracting the sector volatility.

If the straddles are sold on stocks with fluctuations and higher RVs, the distribution of returns should still make the strategy generally perform in line with the sector, until the high RVs reverts due to less hedging from market makers.



Figure 17: DLS volatility filtered strategy returns and XLY returns. Every time a straddle is sold on a SS, one is sold on XLY, making a fair comparison.

4.3.3 Performance

We will focus on analysing performance between the timeframe 20th April 2020 to 17th December 2021.

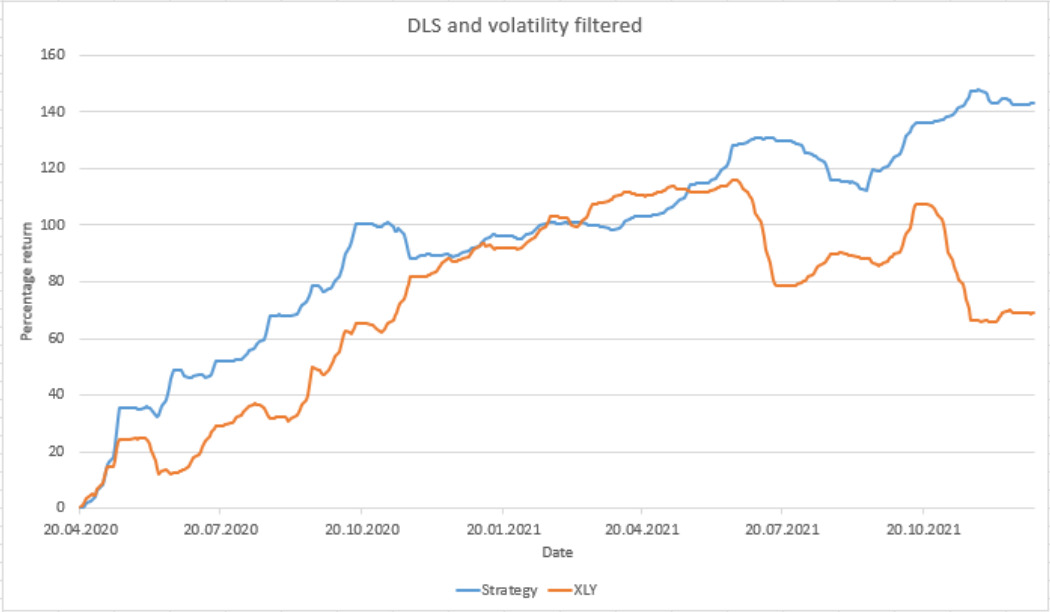


Figure 18: DLS volatility filtered strategy returns and XLY returns in the timeframe 20th April 2020 to 17th December 2021.

For the duration, the strategy performed a total of 141% while XLY returned 69%, making our strategy outperform the reference index by just over 100%.

4.3.4 Risk Adjusted Returns

Examining risk-adjusted returns is crucial, especially in derivative-based strategies as they often involve leveraging positions (Chen, 2023). Derivatives can amplify gains and losses due to increased exposure in the market. We calculated the risk adjusted returns using the Sharpe and Sortino ratio.

$$Sharpe\ Ratio = \frac{R_{portfolio} - R_{risk\ free}}{\sigma_{portfolio}}$$

$$Sortino\ Ratio = \frac{R_{portfolio} - R_{risk\ free}}{\sigma_{downside}}$$

Ratio	DLS Volatility Filtered	XLY
Sharpe	0.38	0.15
Sortino	0.63	0.16

Figure 19: Comparison of performance ratios. The data indicates superior risk-adjusted performance for the DLS volatility filtered strategy.

Both strategies returned low Sharpe ratios indicating that the strategies involved significant risk. However, in derivative strategies where one might exploit short-term inefficiencies, these metrics are not always good at reflecting actual risk taken, as events that could have contributed to loss and volatility never happened. This is usually referred to as tail risk (Fernando, 2024).

The relationship between the ratios is often more interesting. Figure 19 shows that our strategy did significantly better with the Sortino ratio than with the Sharpe ratio. This can potentially be attributed to underlying stock volatility quickly reverting back to normal levels after a gamma squeeze, resulting in rapid profit for our strategy, and a higher standard deviation. These quick returns are not reflected in the Sortino ratio as it only covers downside deviation, indicating that our strategy profits from inefficiencies that are corrected quickly.

4.3.5 Correlation and Return Attributes

Our strategy aimed at exploiting volatility reversion by looking at companies with higher RV and VRP ratios compared to the sector. This can generate returns due to leveraging market trends rather than actually exploiting inefficiencies. The VIX index displays big gaps between RV, meaning shorting volatility generally was profitability across the market.

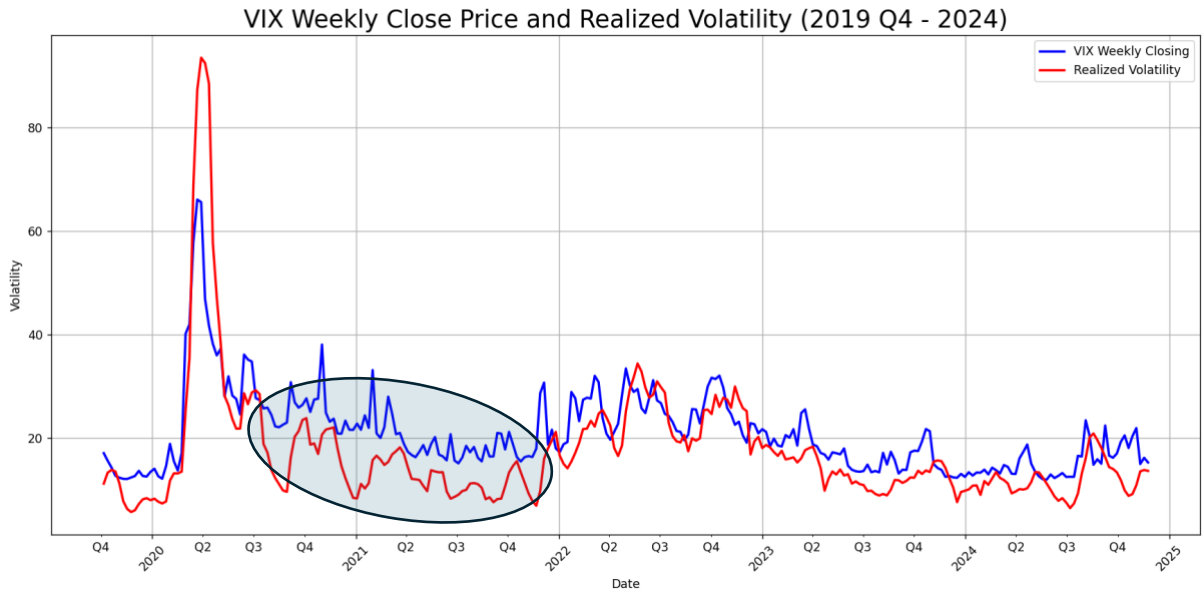


Figure 20: VIX weekly closing prices and RV from Q4 2019 to 2024. Notable spikes in volatility, such as during early 2020, reflect significant market turbulence, with both measures aligning closely in trend but differing in magnitude at various intervals.

Data Source: CBOE (VIX) & Investing.com (2024)

To further analyse the correlation of returns we completed a regression analysis between daily returns for XLY and the filtered volatility DLS strategy.

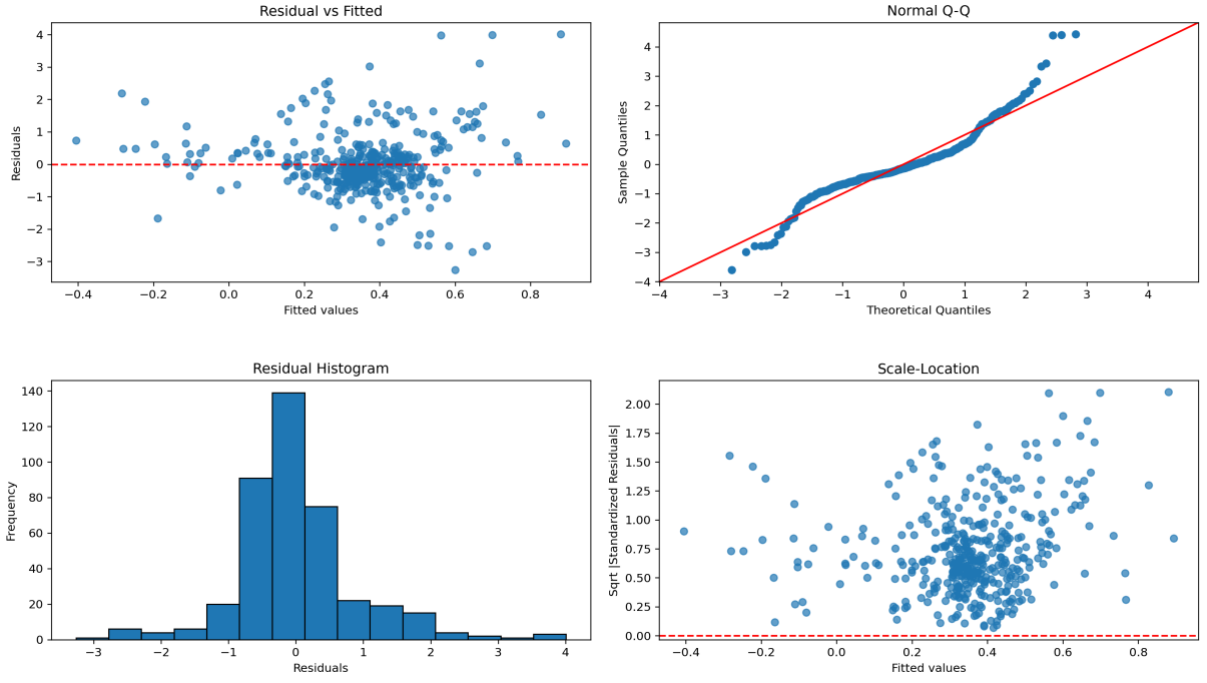


Figure 21: Diagnostic plots for linear regression between DLS and XLY returns.

```
Correlation Coefficient between DLS Returns and XLY_Returns: 0.1874
                            OLS Regression Results
==============================================================================
Dep. Variable:            DLS_Returns   R-squared:                       0.035
Model:                            OLS   Adj. R-squared:                  0.033
Method:                 Least Squares   F-statistic:                     14.77
Date:                Thu, 05 Dec 2024   Prob (F-statistic):           0.000141
Time:                        16:49:18   Log-Likelihood:                -538.70
No. Observations:                 408   AIC:                             1081.
Df Residuals:                     406   BIC:                             1089.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.3248      0.045      7.142      0.000       0.235       0.414
XLY_Returns    0.1532      0.040      3.844      0.000       0.075       0.232
==============================================================================
Omnibus:                       72.682   Durbin-Watson:                   0.423
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              263.857
Skew:                           0.752   Prob(JB):                     5.06e-58
Kurtosis:                       6.641   Cond. No.                         1.22
==============================================================================
```

Figure 22: OLS regression summary for DLS and XLY returns.

The regression analysis reveals a weak positive relationship between DLS and XLY returns, with a correlation coefficient of 0.187 and an R-squared value of 3.5%. This indicates that only a small portion of the variability in DLS returns is explained by sector-wide movements captured by XLY, suggesting that the filtered DLS strategy performance is largely independent. The regression coefficient of 0.154 further highlights the low sensitivity of DLS to changes in the broader sector, as a 1% change in XLY corresponds to only a 0.154% change in DLS on average. The significant variability in residuals implies that DLS returns are driven by factors not directly tied to XLY.

As displayed in a higher Sortino ratio compared to Sharpe ratio, our strategy seems to have certain days yielding most of the profits, in contrast to XLY with more equally distributed returns. This is also shown in both the distribution and residuals displaying outliers, supporting the theory that our returns are due to quick volatility reversions likely as a result from market makers hedging activity and subsequent adjustments in option prices.

The index outperformed our strategy between the middle of the third quarter 2020 and the middle of the second quarter 2021, even when our strategy achieved positive results.



Figure 23: Plot of excess returns between DLS volatility filtered strategy returns and XLY returns.

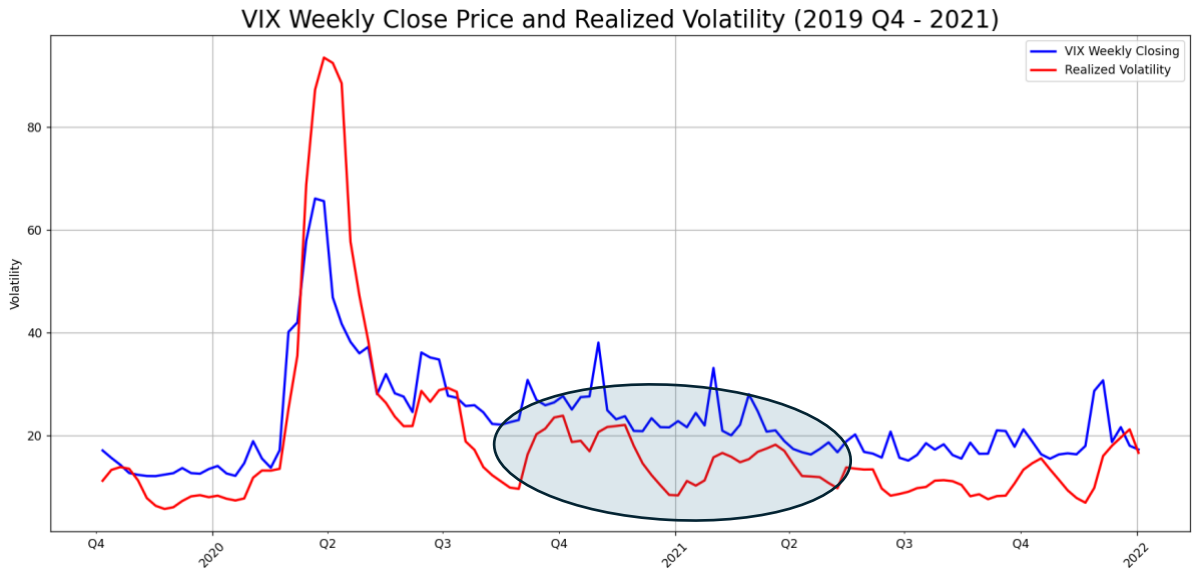


Figure 24: Comparison of VIX weekly closing prices and RV from Q4 2019 to the end of 2021. Highlighted section emphasizes a period in 2021 where RV consistently trailed the VIX, indicating a divergence between market expectations and actual observed volatility.

The VIX index unfolds a possible explanation. This was a time where RV generally was higher, potentially explaining why our strategy returned less when betting on volatility

reversions. XLY contains resilient stocks, and it might have experienced less volatility due to companies like Google, Meta and others.

After these two quarters XLY becomes unprofitable, while our strategy keeps singling out the right stocks and makes significant profits shorting volatility, until returns stagnate around the second quarter of 2022.



Figure 25: DLS volatility filtered strategy returns and XLY returns with highlighted periods of under and outperformance.

Part 5: Limitations and Further Research

Deriving sufficient data for our research was difficult as the schools Bloomberg terminal and free-to-use applications such as Yahoo Finance have limited data about historical option chains, prices and volumes. As a result, we have approximated option prices and deltas by using IV data from VolVue (approach described in Appendix), which might have made our PnL and DLS calculations inaccurate at times. However, we have done this in a way that ensures decisions were never based on future data. Our results should be accurate enough to capture trends and be representative of the respective changes in option prices and premiums.

We did not include transaction costs in our back testing, something that might significantly have impacted strategy returns if accounted for. Additionally, we did not use information about the liquidity of the options sold. Further research considering volumes and transaction costs could give valuable insight into the profitability and scalability of the current strategy, and potential adjustments required to be applicable in an actual market.

This paper was inspired by a Barclays report published 14th September 2020. The paper presented returns from a proprietary strategy that involved shorting volatilities in response to heightened retail trading activity (Deshpande et al., 2020). We do not exactly know what their methodology entailed, but we do know they selectively sold delta hedged straddles on 50 stocks and applied the strategy on SPX as a reference for returns.

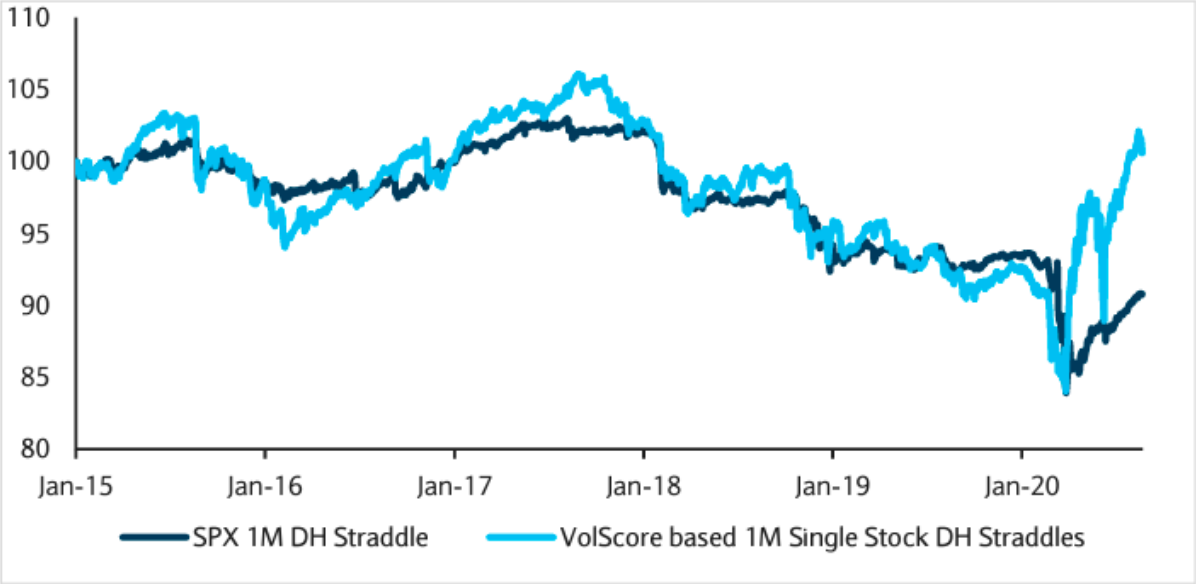


Figure 26: Barclays trading strategy returns for January 2020.

Source: Deshpande et al. (2020)

Even though we only applied our strategy to the consumer discretionary sector, our results seem comparable to theirs. Further research could explore the applicability of our strategy to stocks in other sectors, such as technology, which feature numerous companies with high option trading volumes.

We believe that identifying, analysing, and understanding the factors driving volatility, such as market makers' hedging activities, holds potential to reveal predictable patterns. Further research may determine whether and how these insights can be effectively leveraged for profit in other market situations.

There is currently what seems to be another "boom" in cryptocurrency and "meme coins" market. Figure 27 indicates signs of Bitcoin trending upward, likely representing retail interest in the asset and crypto in general. Additionally, CBOE has enabled options on Bitcoin ETFs, enabling investors to profit from pure volatility-based strategies (CBOE, 2024). Understanding of retail market dynamics, trends and subsequent effects on volatility, may present an opportunity to profit from the current trend.



Figure 27: Google Trends worldwide frequency chart for search word "Bitcoin".

Source: Google Trends (2024)

Part 6: Conclusion

This study examined the increase and impact of retail participation in the U.S. options market. There appears to have been a significant influx in retail activity, driven by market trends and the introduction of zero-commission options trading on multiple platforms. The development and application of the DLS metric suggest that options were likely not systematically mispriced directly by retail traders bidding up option prices, indicating that our initial hypothesis may not hold.

Our final strategy, implemented through delta-hedged short straddles, outperformed the sector benchmark over the examined period, delivering both high returns and superior risk-adjusted performance. While the initial strategy lacked predictive power, incorporating an additional filter to account for relative volatility produced notable improvements. To address our research question:

How could an investor effectively exploit retail-driven option mispricing from 2020 to 2021 using a volatility-based trading strategy?

Our findings suggest that retail-driven market activity indirectly impacted the volatility of underlying asset prices through market-maker hedging activity. Their influence led to predictable volatility reversions that were not fully priced into the market, revealing an arbitrage opportunity that remained effective until 2022.

Part 7: References

Business Insider. (2021, January 27). *How Reddit day traders are using the platform to upend the stock market and make money in the process.* Retrieved from https://markets.businessinsider.com/news/stocks/reddit-day-traders-wallstreetbets-gamestop-gme-rally-upending-stock-market-2021-1-1030009280

Caporal, J. (2023, November 27). *Gen Z and Millennial Investors: Ranking the Most Used, Trusted Investing Tools*. The Motley Fool. Retrieved from https://www.fool.com/research/gen-z-millennial-investors-tools/

Cboe Global Markets. (n.d.). *VIX historical data*. Retrieved December 8, 2024, from https://www.cboe.com/tradable_products/vix/vix_historical_data/

Cboe Global Markets. (2024, November 22). *Cboe to launch the first cash-settled options product related to spot Bitcoin, beginning Monday, December 2*. Retrieved from https://ir.cboe.com/news/news-details/2024/Cboe-to-Launch-the-First-Cash-Settled-Options-Product-Related-to-Spot-Bitcoin-Beginning-Monday-December-2/default.aspx

Chen, J. (2023, September 13). *Understanding Risk-adjusted Return and Measurement Methods.* Investopedia. Retrieved from https://www.investopedia.com/terms/r/riskadjustedreturn.asp

Chen, J. (2024, August 5). *What Is a Straddle Options Strategy and How Is It Created.* Investopedia. Retrieved from https://www.investopedia.com/terms/s/straddle.asp

Comcast. (2020, June 17). *ComScore finds increased consumer interest in brokerage and investment accounts.* Retrieved from https://www.comscore.com/Insights/Press-Releases/2020/6/Comscore-Finds-Increased-Consumer-Interest-in-Brokerage-and-Investment-Accounts

Deloitte. (2021). *The rise of newly empowered retail investors.* Retrieved from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/financial-services/us-the-rise-of-newly-empowered-retail-investors-2021.pdf

Deshpande, M. S., Sen, A., Pu, B., Gao, S., & Krauklis, E. (2020, September 14). *Impact of retail options trading.* U.S. Equity Derivatives Strategy, Barclays. Retrieved from https://amarketplaceofideas.com/wp-content/uploads/2021/08/Barclays_US_Equity_Derivatives_Strategy_Impact_of_Retail_Options_Trading.pdf

Eaton, G. W., Green, T. C., Roseman, B. S., & Wu, Y. (2021). *Zero-commission individual traders and their impact on market quality.* Stern Microstructure Meeting 2021. Retrieved from https://pages.stern.nyu.edu/~jhasbrou/SternMicroMtg/Old/SternMicroMtg2021/Papers/Zero-commission%20individual%20traders.pdf

Farley, A. (2022, September 30). *Is there a Cath to Free stock trading: What's the catch?* Investopedia. Retrieved from https://www.investopedia.com/investing/free-stock-trading-whats-catch/

Fernando, J. (2024, August 4). *Sharpe ratio: Definition, formula, and examples*. Investopedia. Retrieved December 8, 2024, from https://www.investopedia.com/terms/s/sharperatio.asp

Gavin. (2020, September 12). *The ultimate guide to the short straddle strategy*. Options Trading IQ. Retrieved from https://optionstradingiq.com/short-straddle

Google. (n.d.). *Bitcoin.* Google Trends. Retrieved from https://trends.google.com/trends/explore?date=today%205-y&q=bitcoin

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners.* Oxford University Press.

Hayes, A. (2024, July 3). *Volatility: Meaning in finance and how it works with stocks.* Investopedia. Retrieved from https://www.investopedia.com/terms/v/volatility.asp

Hayes, A. (2024, July 11). *Black-Scholes model: What It Is, How It Works, and Options Formula.* Investopedia. Retrieved from https://www.investopedia.com/terms/b/blackscholes.asp

Hawkins, L. E., Versace, C., Abssy, M. (2022, March 31). *Gen Z: What to Know About the Next Generation of Investors*. Nasdaq. Retrieved from https://www.nasdaq.com/articles/gen-z%3A-what-to-know-about-the-next-generation-of-investors

Investing.com. (n.d.). *Realized volatility: Historical data*. Retrieved December 8, 2024, from https://www.investing.com/indices/realized-volatility-historical-data

Kenton, W. (2023, July 5). *Moneyness Definition and Intrinsic Value of Options*. Investopedia. Retrieved December 8, 2024, from https://www.investopedia.com/terms/m/moneyness.asp

Koh, J. (2021, July 16). *Young investors have a huge stomach for risk right now, data suggests.* Yahoo Finance. Retrieved from https://finance.yahoo.com/news/young-investors-have-a-huge-stomach-for-risk-right-now-data-suggests-194739891.html

Mazzola, J. (2024). *Retail trader options flow and the increasing use of index options.* Nasdaq. Retrieved from https://www.nasdaq.com/videos/retail-trader-options-flow-and-increasing-use-index-options

McGuane, J. (2018, May). *What are volatility risk premium strategies?* Retrieved from https://www.marquetteassociates.com/wp-content/uploads/2018/05/What-are-Volatility-Risk-Premium-Strategies-1.pdf

U.S. Securities and Exchange Commission. (2021). *Form S-1 registration statement: Robinhood Markets*. Retrieved from https://www.sec.gov/Archives/edgar/data/1783879/000162828021013318/robinhoods-1.htm

Osterland, A. (2019, November 7). *As behemoth brokerage firms go zero-commission on trades, advisors are concerned.* CNBC. Retrieved from https://www.cnbc.com/2019/11/06/as-brokerage-firms-go-to-zero-commission-on-trades-advisors-worry.html

Peter G. Peterson Foundation. (2021, March 15). *What to know about all three rounds of coronavirus stimulus checks.* Retrieved from https://www.pgpf.org/article/what-to-know-about-all-three-rounds-of-coronavirus-stimulus-checks/

Rocco, A. (2024, June 7). *Gamma squeeze: Understand the GameStop play.* Yahoo Finance. Retrieved from https://finance.yahoo.com/news/gamma-squeeze-understand-gamestop-play-173600502.html

Robinhood Markets. (2022). *Annual report 2021*. Retrieved from https://s28.q4cdn.com/948876185/files/doc_financials/2021/ar/HOOD-2021-Annual-Report.pdf

Schwartz, H. (2022, March 8). *Option flow 2021 – Retail rising.* Chicago Board Options Exchange (CBOE). Retrieved from https://www.cboe.com/insights/posts/option-flow-2021-retail-rising/

Shiller, R. J. (2015). *Irrational exuberance* (3rd ed.). Princeton University Press.

The Options Clearing Corporation. (n.d.). *Historical volume statistics*. Retrieved December 8, 2024, from https://www.theocc.com/market-data/market-data-reports/volume-and-open-interest/historical-volume-statistics

Thomas, K. (2020, December 9). *Warren Buffett: Patience will translate to massive returns.* Yahoo Finance. Retrieved from https://ca.finance.yahoo.com/news/warren-buffett-patience-translate-massive-234539853.html

Vereckey, B. (2022, August 4). *Retail traders lose big in options markets, research shows.* MIT Sloan School of Management. Retrieved from https://mitsloan.mit.edu/ideas-made-to-matter/retail-investors-lose-big-options-markets-research-shows

VolVue. (2024). *Historical data of IV and RV.* Retrieved from https://volvue.com/

Zucchi, K. (2023). *Institutional traders vs. retail traders: What's the difference?* Investopedia. Retrieved from https://www.investopedia.com/articles/active-trading/030515/what-difference-between-institutional-traders-and-retail-traders.asp

Part 8: Appendix

8.1 Black-Scholes Price Formula

Theoretical option pricing is intrinsically linked to the principle of arbitrage-free pricing. Arbitrage-free pricing ensures that the value of financial instruments, like options, aligns with the absence of risk-free profit opportunities in efficient markets. The Black-Scholes pricing model builds on this foundation for European options (Hayes, 2024). The model uses mainly five inputs to determine the fair value of an option. The option's strike price (K), the underlying asset's current price (S_0), the risk-free interest rate (r), and the time to expiration (T) and lastly the future volatility (σ referred to as implied volatility (IV)). The only one of these that isn't known and needs to be estimated is the IV (Hayes, 2024).

$$C = S_0 \times N(d_1) - K \times e^{-rT} \times N(d_2)$$

Black-Scholes formula for a European call option.

$$P = K \times e^{-rT} \times N(-d_2) - S_0 \times N(-d_1)$$

Black-Scholes formula for a European put option.

$$d_1 = \frac{ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right) \times T}{\sigma \times \sqrt{T}} \quad d_2 = d_1 - \sigma \times \sqrt{T}$$

Definition of d_1 and d_2.

Definition of variables:

C = Price of the call option.

P = Price of the put option.

d_1 = Represents the standardized measure of how far the current asset price (S_0) is above the strike price K, adjusted for the expected growth rate and volatility over the option's life.

d_2 = Reflects the probability threshold for the option expiring in-the-money considering the time value and volatility.

S_0 = Current price of the underlying asset.

K = Strike price of the option.

T = Time to expiration (in years).

r = Risk-free interest rate (annualized).

σ = Volatility of the underlying asset's returns (annualized).

$N(d)$ = Cumulative distribution function of the standard normal distribution.

When the price of a call or a put already is known, IV is directly derived from the Black-Scholes option pricing formula. It understands the pricing as the market's anticipation of future volatility, making it easy to compare the IV with the subsequent RV.

8.2 Approximation to Option Prices

We utilised IV data from VolVue to calculate the option prices each day and a risk-free interest rate of 2%. Our strike prices for the options were derived from rounding down the share price at time of option sale to the closest round number. From 30-20: days remaining, we used the 30 days IV, 20-10 days remaining: 20 days IV, 10-0 days remaining: 10 days IV. VolVue has data separate for both puts and calls.

Our approach ensures that the PnL never is based on future data. However, approximations may have resulted our strategy, as strike prices, IVs and risk-free interest rates may not always align perfectly with reality.

8.3 Code Explanation of Delta Hedged Straddles

Below is the code to run delta-hedged straddles on selected tickers. The code requires the following input data: IV, stock prices for the underlying ticker, and the sell and exercise dates for the options. The output is a csv file containing daily PnL of straddles for different tickers.

In the code, d1 is calculated using the Black-Scholes formula as:

$$d1 = \frac{\ln(S/K) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

S is the stock price, K is the strike price, T is the time to expiration (in years), or is the risk-free rate, and σ (sigma) is the annualized volatility. The formula combines the log of the stock-to-strike price ratio, the risk-free rate adjusted for volatility $\left(r+\frac{\sigma^2}{2}\right)$ over time and scales it by volatility and the square root of time $\left(\sigma\sqrt{T}\right)$. This essentially measures how far the current stock price is from the strike price in terms of standard deviations.

For the delta calculation, several parameters needed to be estimated. The IV (σ) for both call and put options is derived from market data and adjusted based on the time to expiration (e.g., 30-day, 20-day, or 10-day volatilities). The time to expiration (T) is calculated as the number of days from the current date to the option's expiration, converted into years ($T = Days/365$). Additionally, the risk-free interest rate (r) is assumed to be a constant value, reflecting prevailing short-term treasury yields and providing a practical benchmark for discounting the option's strike price to present value. These inputs directly influence the calculation of d1, which determines delta as $\phi(d1)$ for calls and $\phi(d1)-1$ for puts, measuring the sensitivity of the option price to changes in the underlying stock price.

The hedgePnL is calculated as the profit or loss resulting from adjusting the hedge position in the underlying stock to offset the options' delta exposure. Each day, the net delta is determined based on the combined deltas of the call and put options, adjusted for the number of contracts. The hedge position is updated daily by buying or selling shares of the stock to match the new delta. The daily hedgePnL is then calculated as the change in the value of the hedge, considering the stock price difference between consecutive days. This is expressed as:

$$-\Delta_{prev}\cdot S_{prev}+\Delta_{prev}\cdot S_{current}$$

Δ_{prev} is the previous day's delta, S_{prev} and $S_{current}$ are the stock prices on consecutive days. The cumulative hedgePnL reflects the total profit or cost from maintaining the hedge over the period.

The dailyPnL represents the profit or loss for the options position on a given day, factoring in option premiums, hedge adjustments, and buyback costs. It begins with the initial premiums collected from selling the options. The daily hedgePnL, which reflects the profit or

cost from adjusting the hedge to offset delta exposure, is added. The cost of repurchasing the options at their current market value (i.e. buyback cost) is then subtracted. The result is normalized by dividing the net PnL by the initial premiums, yielding a percentage measure of daily performance relative to the original premium.

8.4 Python Code

8.4.1 Delta Hedged Short Straddles

```
import numpy as np
import pandas as pd
import datetime as dt
import matplotlib.pyplot as plt
from scipy.stats import norm
import os

# Define the final output file path where the aggregated results for all
tickers will be saved
final_output_path = "Final_Output/FinalCsvNormalizedSPY.csv"


# Load volatility data for options
volatilites = pd.read_csv(f'Volatility_Data/Volatility_DataSPY.csv',
delimiter=',')

# List of tickers for which calculations will be performed
tickers = [
"SPY"
]



# Loop through each ticker to calculate daily PnL and deltas
for ticker in tickers:

    dailyPnL = {}
    dailyDelta = {}

    # Function to calculate the d1 term for Black-Scholes formula
    def calculate_d1(S, K, T, r, sigma):
        """
        Calculate the d1 term used in the Black-Scholes formula.

        Args:
        - S: Current stock price
```

```
        - K: Strike price of the option
        - T: Time to expiration (in years)
        - r: Risk-free interest rate
        - sigma: Volatility of the underlying asset (annualized)

        Returns:
        - d1: The calculated d1 value
        """

        return (np.log(round(S) / K) + (r + 0.5 * sigma**2) * T) / (sigma *
np.sqrt(T))

    # Function to calculate call and put deltas using Black-Scholes formula
    def calculate_deltas(S, K, T, r, sigma_call, sigma_put):
        """
        Calculate the delta values for call and put options.

        Args:
        - S: Current stock price
        - K: Strike price of the option
        - T: Time to expiration (in years)
        - r: Risk-free interest rate
        - sigma_call: Volatility of the call option
        - sigma_put: Volatility of the put option

        Returns:
        - call_delta: Delta for the call option
        - put_delta: Delta for the put option
        """

        d1_call = calculate_d1(round(S), K, T, r, sigma_call)
        d1_put = calculate_d1(round(S), K, T, r, sigma_put)
        call_delta = norm.cdf(d1_call)
        put_delta = norm.cdf(d1_put) - 1
        return call_delta, put_delta

    # Function to calculate option price using Black-Scholes formula
    def calculate_option_price(S, K, T, r, sigma, option_type="call"):

        """
        Calculate the price of an option using the Black-Scholes formula.

        Args:
        - S: Current stock price
        - K: Strike price of the option
        - T: Time to expiration (in years)
        - r: Risk-free interest rate
        - sigma: Volatility of the underlying asset (annualized)
```

```
        - option_type: Type of option ("call" or "put")

        Returns:
        - Option price (float)
        """

        if isinstance(T, pd.Timedelta): # Ensure T is a float
            T = T.total_seconds() / (365 * 24 * 60 * 60)

        d1 = calculate_d1(S, K, T, r, sigma)
        d2 = d1 - sigma * np.sqrt(T)
        if option_type == "call":
            return round(S) * norm.cdf(d1) - K * np.exp(-r * T) * norm.cdf(d2)
        elif option_type == "put":
            return K * np.exp(-r * T) * norm.cdf(-d2) - round(S) * norm.cdf(-
d1)

    # Function to process hedging calculations for a specific ticker
    def process_hedging(file_name, r, sell_date, expiration_date,
contracts=1):

        """
        Processes hedging calculations for a specific stock ticker between the
sell and expiration dates.

        Args:
        - file_name: Path to the stock price data CSV file.
        - r: Risk-free interest rate.
        - sell_date: The date the option position is initiated (sell date).
        - expiration_date: The date the option expires.
        - contracts: Number of option contracts sold (default is 1).

        Performs:
        - Loads stock price data and filters it for the given time range.
        - Matches implied volatilities for the ticker based on sell and
expiration dates.
        - Calculates option premiums and deltas.
        - Tracks daily changes in PnL (Profit and Loss) and deltas for
hedging.

        Raises:
        - ValueError: If no stock price data is available between the sell and
expiration dates.
        """

        # Load the stock price data
        stock_prices = pd.read_csv(file_name)
```

```
        # Parse the Date column to remove timezone and standardize
        stock_prices['Date'] = pd.to_datetime(stock_prices['Date'],
utc=True).dt.tz_convert(None)

       # Format the Date to YYYY-MM-DD without changing it to a string
        stock_prices['Date'] = pd.to_datetime(stock_prices['Date'].dt.date)

        # Keep only the Date column and the column corresponding to the ticker
        stock_prices = stock_prices[['Date', ticker]]

        # Rename the ticker column for consistency
        stock_prices.rename(columns={ticker: "StockPrice"}, inplace=True)

        # Sort values by Date
        stock_prices = stock_prices.sort_values(by="Date")

        # Get sell date and expiration date
        sell_date = pd.Timestamp(sell_date)
        expiration_date = pd.Timestamp(expiration_date)

        # Filter data between sell date and expiration date
        stock_prices = stock_prices[(stock_prices['Date'] >= sell_date) &
(stock_prices['Date'] <= expiration_date)].reset_index(drop=True)
        if stock_prices.empty:
            raise ValueError("No data available between sell date and
expiration date.")

        # Initialize option parameters
        strike_price = stock_prices.loc[0, "StockPrice"]
        S = strike_price
        K = strike_price


        volatilites['Date'] = pd.to_datetime(volatilites['Date'])  # Ensure
'Date' is a datetime object
        sell_date = pd.to_datetime(sell_date)  # Convert sell_date to datetime

        # Match volatility data for the given ticker and sell date
        try:
            firstRow = volatilites[(volatilites['Date'] == sell_date) &
(volatilites['Ticker'] == f"{ticker}")]
            iv_call = firstRow['30-Day Call IV'].iloc[0] / 100
            iv_put = firstRow['30-Day Put IV'].iloc[0] / 100
        except IndexError:
            print("something fishy")
            return

        expiration_date = pd.Timestamp(expiration_date)
        sell_date = pd.Timestamp(sell_date)
```

```
        T_initial = (expiration_date - sell_date).days / 365

        # Calculate option premiums
        try:
            call_premium = calculate_option_price(S, K, T_initial, r, iv_call,
option_type="call")
            put_premium = calculate_option_price(S, K, T_initial, r, iv_put,
option_type="put")
            total_call_premium = call_premium * contracts
            total_put_premium = put_premium * contracts
            total_premium = total_call_premium + total_put_premium
        except ValueError:
            print("Something Smelly")
            return


        # Calculate DaysTillExpiration
        stock_prices["DaysTillExpiration"] = (expiration_date -
stock_prices["Date"]).dt.days

        # Initialize variables
        hedge_shares = []
        net_deltas = []

        # Ensure the Date columns are datetime and match formats
        stock_prices['Date'] = pd.to_datetime(stock_prices['Date'])

        for i, row in stock_prices.iterrows():
            S = row["StockPrice"]
            T = row["DaysTillExpiration"] / 365

            #Adjust volatility for remaining time
            try:
                if 20 < row["DaysTillExpiration"] < 30:
                    volrow =  volatilites.loc[(row["Date"] ==
volatilites["Date"]) & (volatilites['Ticker'] == f"{ticker}") ]
                    iv_call = volrow["30-Day Call IV"].iloc[0] / 100
                    iv_put = volrow["30-Day Put IV"].iloc[0] / 100
                elif 10 < row["DaysTillExpiration"] <= 20:
                    volrow =  volatilites.loc[(row["Date"] ==
volatilites["Date"]) & (volatilites['Ticker'] == f"{ticker}") ]
                    iv_call = volrow["20-Day Call IV"].iloc[0] / 100
                    iv_put = volrow["20-Day Put IV"].iloc[0] / 100
                elif 0 < row["DaysTillExpiration"] <= 10:
                    volrow =  volatilites.loc[(row["Date"] ==
volatilites["Date"]) & (volatilites['Ticker'] == f"{ticker}") ]
                    iv_call = volrow["10-Day Call IV"].iloc[0] / 100
                    iv_put = volrow["10-Day Put IV"].iloc[0] / 100
```

```
            except IndexError:
                print("Something Stinky")

            # Calculate delta and hedge shares
            if row["DaysTillExpiration"] > 0:
                call_delta, put_delta = calculate_deltas(S, strike_price, T,
r, iv_call, iv_put)
                net_delta = -call_delta - put_delta
                hedge_shares_needed = -net_delta * contracts
                net_deltas.append(net_delta)
                hedge_shares.append(hedge_shares_needed)

            # Calculate buyback cost
            call_buyBack = calculate_option_price(S, K, T, r, iv_call,
option_type="call")
            put_buyBack = calculate_option_price(S, K, T, r, iv_put,
option_type="put")
            total_buyBack = call_buyBack + put_buyBack

            # If its the first day, the hedgePnL = 0, else calculate the
hedgePnL
            if i != 0:
                hedgePnL += - stock_prices["StockPrice"].iloc[i-1]  *
(net_deltas[i-1]) + stock_prices["StockPrice"].iloc[i] * (net_deltas[i-1])
            else:
                hedgePnL = 0


            # Store the date and PnL to parse into dataframe
            if row["Date"] not in dailyPnL:
                dailyPnL[row["Date"]] = 0
                dailyPnL[row["Date"]] += (total_premium + hedgePnL -
(total_buyBack))/total_premium # Initialize the key if it doesn't exist
            else:
                dailyPnL[row["Date"]] += (total_premium + hedgePnL -
(total_buyBack))/total_premium

            dailyDelta[row["Date"]] = net_delta


    # Main function to loop through sell and expiration dates
    def main(stock_path, output_name, SellDates):
        """
        This function processes hedging data for multiple sell and expiration
dates,
        calculates daily PnL and delta for each ticker, and saves the results
to a CSV file.

        Args:
```

```
        - stock_path: Path to the stock price data CSV file.
        - output_name: Path where the results CSV for the current ticker will
be saved.
        - SellDates: Path to the CSV file containing sell and expiration
dates.
        """

        # Load sell and expiration dates from the CSV file
        SellDates = pd.read_csv(SellDates)


        # Loop through each pair of sell and expiration dates
        for i in range(len(SellDates)):
            sell_date = SellDates.loc[i, 'SellTrading Dates']
            expiration_date = SellDates.loc[i, 'Excersice Dates']

            # Process hedging calculations for the current sell and expiration
date
            process_hedging(stock_path, 0.02, sell_date, expiration_date)

        # Create a DataFrame to store the daily PnL and Delta values
        results_data = pd.DataFrame({
        "Date": list(dailyPnL.keys()),
        "DailyPnL": list(dailyPnL.values()),
        "DailyDelta": list(dailyDelta.values())
        })

        # Save the results to the output CSV file
        results_data.to_csv(output_name, index=False)

    # Use the main function
    combined_results = main(f"Ticker_Data/Stock_DataSPY.csv",
f"Output_Data/{ticker}_results.csv", "NewHedge/SellExcersieDates.csv")

    # Create or update the FinalCsv file with aggregated data for all tickers
    if os.path.exists(final_output_path):
        # If the final output file exists, load it
        final_data = pd.read_csv(final_output_path)
    else:
        # If it doesn't exist, create a new DataFrame with dates as the
initial column
        final_data = pd.DataFrame({"Date": list(dailyPnL.keys())})

    # Ensure the 'Date' column is properly formatted for consistency
    final_data["Date"] = pd.to_datetime(final_data["Date"])

    # Create a DataFrame for the current tickers' data
    current_ticker_data = pd.DataFrame({
        "Date": list(dailyPnL.keys()),  # Dates from the current ticker
```

```
        f"{ticker}": list(dailyPnL.values()) # Corresponding PnL values
    })

    # Merge the current ticker's data with the final output
    final_data = pd.merge(final_data, current_ticker_data, on="Date",
how="outer")

    # Save the updated final DataFrame to the output file
    final_data.to_csv(final_output_path, index=False)
    print(f"Final CSV updated with data for ticker: {ticker}")
```

8.4.2 VIX Chart

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as mdates

# File paths to CSV files
vix_file_path = "VIX_History.csv"
vol_file_path = "Realized Volatility Historical Data.csv"

# Load the VIX data
vix_data = pd.read_csv(vix_file_path)
vix_data.columns = vix_data.columns.str.strip()  # Clean column names
vix_data['DATE'] = pd.to_datetime(vix_data['DATE'])  # Convert DATE to
datetime

# Filter for the date range
vix_data = vix_data[(vix_data['DATE'] >= '2019-10-01') & (vix_data['DATE'] <=
'2021-12-31')]
vix_data.set_index('DATE', inplace=True)
vix_weekly_data = vix_data.resample('W').agg({'CLOSE': 'last'})  # Resample
weekly VIX close price

# Load the realized volatility data
vol_data = pd.read_csv(vol_file_path)
vol_data.columns = vol_data.columns.str.strip()  # Clean column names
vol_data['Date'] = pd.to_datetime(vol_data['Date'])  # Convert Date to
datetime

# Filter for the date range
vol_data = vol_data[(vol_data['Date'] >= '2019-10-01') & (vol_data['Date'] <=
'2021-12-31')]
vol_data.set_index('Date', inplace=True)

# Resample weekly Realized Volatility
```

```
vol_weekly_data = vol_data.resample('W').agg({'Price': 'last'})  # Weekly
closing of Realized Volatility

# Plot
fig, ax = plt.subplots(figsize=(14, 7))

# Plot VIX Weekly Closing
ax.plot(vix_weekly_data.index, vix_weekly_data['CLOSE'], label='VIX Weekly
Closing', color='blue', linestyle='-', linewidth=2)  # Raw data, solid line

# Plot Realized Volatility as solid line
ax.plot(vol_weekly_data.index, vol_weekly_data['Price'], label='Realized
Volatility', color='red', linestyle='-', linewidth=2)  # Raw data, solid line

# Set labels and title
ax.set_xlabel('Date')
ax.set_ylabel('Volatility')
ax.set_title('VIX Weekly Close Price and Realized Volatility (2019 Q4 -
2021)', fontsize=20)
ax.tick_params(axis='y', labelcolor='black')

# Format the x-axis to show quarters and years
locator = mdates.MonthLocator(bymonth=[1, 4, 7, 10])  # Locate the first month
of each quarter

# Customizing quarter formatting
def quarter_format(x, pos):
    dt = mdates.num2date(x)  # Convert to datetime
    year = dt.year
    quarter = (dt.month - 1) // 3 + 1
    return f"Q{quarter}"  # Only show Q1, Q2, Q3, Q4

# Set the major locator and formatter
ax.xaxis.set_major_locator(mdates.YearLocator())  # Set ticks for each year
ax.xaxis.set_minor_locator(locator)  # Set minor ticks for each quarter
ax.xaxis.set_major_formatter(mdates.DateFormatter('%Y'))  # Show only the year
on the major ticks
ax.xaxis.set_minor_formatter(plt.FuncFormatter(quarter_format))  # Show Q1,
Q2, Q3, Q4 on minor ticks

# Rotate the x-axis labels for better readability
plt.xticks(rotation=45)

# Add a legend to distinguish the lines
ax.legend()

# Ensure everything fits without overlapping
fig.tight_layout()
```

```
# Show the plot
plt.grid(True)
plt.show()
```

8.4.3 Annual Option Volume

```
import pandas as pd
import matplotlib.pyplot as plt


# Load the Excel file containing option volume data
file_path = 'Option volume data.xlsx'
excel_data = pd.ExcelFile(file_path)

# Load the data from the first sheet
data = excel_data.parse('Sheet1')

# Rename columns for clarity and filter necessary columns
data.columns = ['Year', 'Equity', 'Non-Equity', 'Debt', 'Futures', 'Stock
Loan', 'Average Daily Volume',
                'OCC Total', 'Options', 'Futures_Open', 'Stock Loan_Open',
'Equity Options_Open',
                'Futures_Open_Interest', 'Stock Loan_Value']

# Filter only Year, Equity, and Non-Equity columns and drop rows with non-
numeric year (header rows)
data_filtered = data[['Year', 'Equity', 'Non-Equity']].dropna()
data_filtered = data_filtered[data_filtered['Year'].apply(lambda x:
str(x).isdigit())]
data_filtered = data_filtered.astype({'Year': int, 'Equity': int, 'Non-
Equity': int})

# Remove the year 2024
data_filtered = data_filtered[data_filtered['Year'] != 2024]

# Calculate total volume in billions
data_filtered['Total Volume (Billions)'] = (data_filtered['Equity'] +
data_filtered['Non-Equity']) / 1e9

# Correct annual growth rate with forward shift (-1), excluding rows where
shift produces NaN
data_filtered['Annual Growth (%)'] = (
    (data_filtered['Total Volume (Billions)'] - data_filtered['Total Volume
(Billions)'].shift(-1)) /
    data_filtered['Total Volume (Billions)'].shift(-1)
) * 100
```

```
# Remove NaN values in Annual Growth (%) caused by the last row
data_filtered = data_filtered.dropna(subset=['Annual Growth (%)'])

# Create a mask for plotting annual growth only from 1991 onwards
growth_filtered = data_filtered[data_filtered['Year'] >= 1991]

# Plotting
fig, ax1 = plt.subplots(figsize=(12, 6))

# Bar plot for total volume in billions with dark blue color (all years)
bars = ax1.bar(data_filtered['Year'], data_filtered['Total Volume
(Billions)'], color='grey', label='Total Volume (Billions)', edgecolor =
"black")
ax1.set_xlabel('Year')
ax1.set_ylabel('Total Volume (Billions)', color='grey',
               fontsize=15,
               fontweight='bold',
               bbox=dict(facecolor='white', edgecolor='black',
boxstyle='round'))

ax1.set_xlim(1973, data_filtered['Year'].max())


# Set x-axis ticks for every year
ax1.set_xticks(data_filtered['Year'])
ax1.set_xticklabels(data_filtered['Year'], rotation=45, ha='right')

# Line plot for annual growth rate on the right y-axis, in red and dashed
(1991 onwards)
ax2 = ax1.twinx()
ax2.plot(growth_filtered['Year'], growth_filtered['Annual Growth (%)'],
color='blue', linestyle='-', label='Annual Growth (%)', linewidth=2.5)
ax2.set_ylabel('Annual Growth (%)', color='blue',
               fontsize=15,
               fontweight='bold',
               bbox=dict(facecolor='white', edgecolor='black',
boxstyle='round'))
ax2.set_xlim(left=1991)
# Adding titles and layout adjustments
plt.title('Annual Total Option Volume and Growth Rate (Excluding 2024)',
fontweight='bold')
fig.tight_layout()
plt.show()

# Print for validation
print(data_filtered[['Year', 'Total Volume (Billions)']])
print(growth_filtered[['Year', 'Annual Growth (%)']])
```

8.4.4 Regression Analysis

```
import pandas as pd
import statsmodels.api as sm
import numpy as np
import matplotlib.pyplot as plt

# Load the data from the returns file
file_path = 'ReturnsToEvaluate.xlsx'
data = pd.read_excel(file_path)

# Rename columns for clarity
data.columns = ['DLS_Returns', 'XLY_Returns']

# Skip rows where DLS_Returns is 0 as this is the start of a new straddle
data = data[data['DLS_Returns'] != 0]

# Calculate and display the correlation coefficient
correlation_coefficient = data['DLS_Returns'].corr(data['XLY_Returns'])
print(f"Correlation Coefficient between DLS Returns and XLY_Returns:
{correlation_coefficient:.4f}")

# Add a constant to the predictor for the intercept in the model
X = data[['XLY_Returns']]
X = sm.add_constant(X)
y = data['DLS_Returns']

# Fit the Ordinary Least Squares (OLS) regression model
model = sm.OLS(y, X).fit()

# Print the summary of the model
print(model.summary())

# Diagnostic Plots
# Predicted values and residuals
fitted_values = model.fittedvalues
residuals = model.resid
standardized_residuals = (residuals - np.mean(residuals)) / np.std(residuals)

# Residual vs Fitted Plot
plt.figure(figsize=(12, 6))

# Residual vs Fitted
plt.subplot(2, 2, 1)
plt.scatter(fitted_values, residuals, alpha=0.7)
plt.axhline(y=0, color='r', linestyle='--')
plt.xlabel('Fitted values')
```

```
plt.ylabel('Residuals')
plt.title('Residual vs Fitted')

# Q-Q Plot
plt.subplot(2, 2, 2)
sm.qqplot(residuals, line='45', fit=True, ax=plt.gca())
plt.title('Normal Q-Q')

# Residual Histogram
plt.subplot(2, 2, 3)
plt.hist(residuals, bins=15, edgecolor='black')
plt.xlabel('Residuals')
plt.ylabel('Frequency')
plt.title('Residual Histogram')

# Scale-Location Plot (Standardized Residuals vs Fitted)
plt.subplot(2, 2, 4)
plt.scatter(fitted_values, np.sqrt(np.abs(standardized_residuals)), alpha=0.7)
plt.axhline(y=0, color='r', linestyle='--')
plt.xlabel('Fitted values')
plt.ylabel('Sqrt |Standardized Residuals|')
plt.title('Scale-Location')

# Display the plots
plt.tight_layout()
plt.show()
```

8.4.5 PnL Profile of a Straddle

```
import numpy as np
import matplotlib.pyplot as plt

# Parameters
strike_price = 100  # Strike price of both call and put
premium_received = 10  # Premium received for each option

# Price range for the underlying asset
prices = np.linspace(50, 150, 400)

# Calculating PnL for the short straddle
# PnL = Premium received - max(0, price - strike_price) - max(0, strike_price
- price)
call_pnl = premium_received - np.maximum(0, prices - strike_price)
put_pnl = premium_received - np.maximum(0, strike_price - prices)
total_pnl = call_pnl + put_pnl

# Break-even points (fixed at 80 and 120)
```

```
breakeven_lower = 80
breakeven_upper = 120

# Plotting the PnL profile
plt.figure(figsize=(10, 6))
for i in range(len(prices) - 1):
    if total_pnl[i] > 0:
        plt.plot(prices[i:i+2], total_pnl[i:i+2], color='g')
    else:
        plt.plot(prices[i:i+2], total_pnl[i:i+2], color='r')
plt.fill_between(prices, total_pnl, where=(total_pnl > 0), color='green',
alpha=0.3, label='Profit Area')
plt.fill_between(prices, total_pnl, where=(total_pnl < 0), color='red',
alpha=0.3, label='Loss Area')
plt.axhline(0, color='black', linestyle='-')
plt.axvline(strike_price, color='red', linestyle='--', label='Strike Price')
plt.axvline(breakeven_lower, color='blue', linestyle='--', label='Break-even
Points')
plt.axvline(breakeven_upper, color='blue', linestyle='--')

plt.title('PnL Profile of a Short Straddle')
plt.xlabel('Price of Underlying Asset at Expiration')
plt.ylabel('Profit / Loss')
plt.legend()
plt.grid(True)
plt.show()
```