Trading Dispersion & Correlation

Duke University
Math 585

Evan Glas, Harry Wang, Lihong Wang, Muchang Bahng, Tony Wu

Abstract

We investigate an innovative trading strategy that combines dispersion trading on the Invesco QQQ Trust (QQQ) and Uniform Manifold Approximation and Projection (UMAP) for pairs selection in pairs trading, with regularized Markowitz portfolio optimization for optimal allocation between the two strategies in live trading. The primary objective of this research is to examine the effectiveness of this hybrid approach in generating superior risk-adjusted returns and portfolio diversification. Our empirical analysis establishes the potential of this innovative trading strategy to outperform conventional approaches in terms of risk-adjusted returns and portfolio diversification. The findings highlight the importance of employing advanced quantitative techniques in the development of robust and effective trading strategies for modern financial markets.

Contents

1 Dispersion Trading

1.1.1 Dispersion Trading & Implied Correlations

A dispersion trade refers to a trade in which one sells index volatility and buys volatility in index components, or vice versa. For the rest of the paper the term "dispersion position" shall refer to only the former trade and the reason for this will become clear. In some sense, a dispersion trade is of similar nature to an index arbitrage trade; instead of reconstructing an index or its corresponding ETFs using the component stocks, we attempt to reconstruct (or rather, approximate) and index option using the options of the component stocks.

Though it appears to the uninitiated that a dispersion trade expresses a market participant's opinion on index volatility, further considerations reveal that it in fact expresses an opinion on index correlation. In particular, index and component volatilities imply a correlation in the index**.** To see why this is the case, consider an index I that can be expressed as a linear combination of n stocks S_i where

$$I = \sum_{i=1}^{n} m_i S_i\,, \qquad m_i = number\ of\ shares\ in\ index$$

By abuse of notation, the instantaneous returns can then be written as

$$\frac{dI}{I} = \frac{1}{I}\sum_{i=1}^{n} m_i dS_i = \sum_{i=1}^{n} \frac{m_i S_i}{I}\frac{dS_i}{S_i} = \sum_{i=1}^{n} w_i \frac{dS_i}{S_i}, \qquad w_i = \frac{m_i S_i}{I}$$

The variance of the instantaneous index returns can then be written as

$$\sigma_I^2 = Var\left[\frac{dI}{I}\right] = Var\left[\sum_{i=1}^{n} w_i \frac{dS_i}{S_i}\right] = \sum_{i,j\le n} w_i w_j Cov\left[\frac{dS_i}{S_i}\frac{dS_j}{S_j}\right] = \sum_{i,j\le n} w_i w_j \sigma_i \sigma_j \rho_{ij}$$

The above equation suggests that index variance = avg stock variance × avg cross correlation. This can also be shown more rigorously, as was done in *(Bossu, 2006)*. Rearranging the equation shows that implied volatility from the stock and index options market must imply a correlation between the components of the index. Unless correlations are 1, the index variance must be lower than the average stock variance.

1.1.2 P&L of Delta-Hedged Dispersion Position

The delta-hedged portfolio consists of being short the option V and long δ of the stock S. Assuming some interest rate r and using first-order approximations for small changes in time and value, the $P\&L$ may be expressed as:

$$P\&L = \Delta V - \delta \Delta S + (\delta S - V)\, r\Delta t$$

ΔV corresponds to the price variation of the option, $\delta\Delta S$ to the stock price movements, of which we hold δ units, and $(\delta S - V)\ r\Delta t$ the risk-free return / margin cost on cash.

The value of the option assuming constant gamma and theta over some small time period can be written as the value captured by gamma and delta over the corresponding move of the underlying along in that time with the loss incurred by theta decay over that time:

$$\Delta V = \int (\delta + \Gamma S)\, dS + \int \theta\, dt = \delta\Delta S + \frac{1}{2}\Gamma(\Delta S)^2 + \theta\Delta t$$

Recall also the Black-Scholes PDE, which states that

$$\theta + \frac{1}{2}\sigma^2 S^2 \Gamma = rV - rS\delta$$

The RHS is essentially a risk-less position assuming instantaneous delta hedging. The equation therefore states that over any infinitesimal time interval the loss from theta and the gain from the gamma term must offset each other so that the result is a return at the riskless rate. Assuming that $r \approx 0$. This allows us to obtain the $P\&L$ expression as

$$P\&L = \frac{1}{2}\Gamma S^2\left(\left(\frac{\Delta S}{S}\right)^2 - \sigma^2\Delta t\right) \approx \theta\left(\left(\frac{\Delta S}{S\sigma\sqrt{\Delta t}}\right)^2 - 1\right) = \theta(z^2 - 1), \qquad z = \frac{\Delta S}{S\sigma\sqrt{\Delta t}}$$

The term z has a natural interpretation as the standardized move of the underlying on the considered period. The $P\&L$ of the delta-hedged index position can then be written as

$$\begin{aligned}
\theta_I(z_I^2 - 1) &= \theta_I\left(\left(\sum_{i=1}^{n} w_i z_i \frac{\sigma_i}{\sigma_I}\right)^2 - 1\right) \\
&= \theta_I\left(\sum_{i=1}^{n}\left(w_i z_i \frac{\sigma_i}{\sigma_I}\right)^2 + \sum_{i\neq j}^{n} w_i w_j z_i z_j \frac{\sigma_i\sigma_j}{\sigma_I} - 1\right) \\
&= \theta_I \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_I^2} w_i^2 (z_i - 1) + \theta_I \sum_{i\neq j}^{n} \frac{\sigma_i\sigma_j}{\sigma_I^2} w_i w_j \left(z_i z_j - \rho_{ij}\right)
\end{aligned}$$

Given that stock i's delta-hedged options position has $P\&L$ approximated by $\theta_i(z_i^2 - 1)$ we can finally derive the $P\&L$ of the Vega-hedged dispersion trade position as

$$P\&L = \sum_{i=1}^{n}\left(\theta_i + \theta_I \frac{\sigma_i^2}{\sigma_I^2} w_i^2\right)(z_i - 1) + \theta_I \sum_{i\neq j}^{n} \frac{\sigma_i\sigma_j}{\sigma_I^2} w_i w_j \left(z_i z_j - \rho_{ij}\right)$$

This above equation has a financial interpretation where the first term represents the $P\&L$ from the realize single-stock and index movements against implied volatilities, while the second term represents the $P\&L$ from the realize cross-market movement against implied correlations. Because delta-hedging in perfectly efficient and liquid markets has zero expected value, this expression can also be interpreted as the expected $P\&L$ from the position. Instantaneous delta hedging reduces the variance of $P\&L$ to zero, while in real financial markets, frequent delta hedging incurs slippage costs but is risk-reducing.

In the above calculations, we used the following two decompositions

$$z_I := \frac{\Delta I}{I\sigma_I\sqrt{\Delta t}} = \left(\sigma_I\sqrt{\Delta t}\right)^{-1}\frac{\sum_{i=1}^{n} m_i \Delta S_i}{\sum_{i=1}^{n} m_i S_i} = \sigma_I^{-1}\sum_{i=1}^{n}\frac{w_i S_i}{\sum_{j=1}^{n} w_j S_j}\sigma_i\frac{\Delta S_i}{\sigma_i S_i\sqrt{\Delta t}} = \sum_{i=1}^{n} w_i\frac{\sigma_i}{\sigma_I} z_i$$

$$\sigma_I^2 = \sum_{i,j\le n} w_i w_j \sigma_i \sigma_j \rho_{ij} = \sum_{i=1}^{n}(w_i\sigma_i)^2 z_i + \sum_{i\neq j} w_i w_j \sigma_i \sigma_j \rho_{ij}$$

1.2.1 Liquidity and Execution

In prior calculations on the weights of the options position in the index, we did not specify what these weights or their corresponding stocks actually were. The two basic questions:

1) Which stocks to pick?

2) How to weigh them?

Have yet to be answered. While there does not exist a definite formula for structuring a good dispersion trade, we can still list some reasonable considerations and justify the choices that we make in the trade that is presented in this paper. First, as the options market for some stocks may lack liquidity, we should not acquire a position on all components of the index but instead settle for a “dirty dispersion” position where we are partially hedged with only a subset of the index components. The subset of the index that we choose should be representative of the index in two ways:

1) Subset average correlation should approximate index average correlation.

2) Subset average volatility should approximate index component average volatility.

This is since the $P\&L$ from the dispersion trade decomposes into $P\&L$ from volatility and $P\&L$ from cross-correlation as we have previously shown. There is no guarantee, however, that a reasonably small subset of the index can actually fit these two characteristics. In practice, taking the assets that forms a large percentage of the index ETF’s Net Asset Value (NAV) appears to be a viable avenue to establish the “dirty dispersion” position. What is then desirable to us is a highly liquid index ETF with a liquid options market, that takes the majority of its NAV from the equity of a few large-cap companies that are themselves highly liquid. After selecting the assets that we wish to execute a dispersion trade on, we are faced with two main weighting strategies:

1) Vega-neutral: Protected against volatility but exposed to directionality.

2) Gamma-neutral: Protected against directionality but exposed to volatility.

Typically, dispersion trades are guided by a human trader who trades into positions that are low on both Vega and Gamma, as well as theta and other Greeks. The automation of this process can be a potential future endeavor but for our naïve implementation of dispersion trading we have to prioritize one of the two.

The main argument for Vega-neutrality is that a dispersion position, even if it has zero Vega, is in fact short volatility. Mathematically, a dispersion position has convex Vega and so Vega hedges only work for small movements in volatility. Economically, equity correlations crash upwards to 1 as stock markets crash downwards and volatility explodes (since returns are usually positive). Either way, we see that we are likely to lose money on our dispersion position when volatility is unusually high. The main argument for Gamma-neutrality is that it reduces the need to delta hedge as often, and therefore reduces overall slippage incurred in crossing the bid-ask spread. Assuming an extremely liquid underlying, the need for gamma neutrality reduces.

Therefore, it is reasonable for us to opt for Vega neutrality especially since the shape of the Vega curve is convex. To see this fact, considering the following:

1) If stock volatilities are constant, and index volatilities increase, implied correlation must be increasing. Equivalently, as correlation increases, we get short Vega as the index volatility short grows faster than the stock volatility longs.

2) The reverse also holds that we get net long Vega as volatility falls.

Having determined that we will prize Vega neutrality over Gamma neutrality, it is time for us to consider candidate indices that actually have sufficient liquidity to support our strategy. We believe that Invesco QQQ (Hereby called QQQ) is one such candidate. QQQ is an ETF that tracks the Nasdaq-100, an index comprised of 100 of the largest stocks that have their primary listing on the Nasdaq Exchange. The index is weighted by capitalization so the equity of larger companies have a bigger influence, which may allow us to set up a safer and more efficient dirty dispersion position. The following table lists the ten largest QQQ components at the end of 2020.

Top Ten NDX Components 12/31/2020

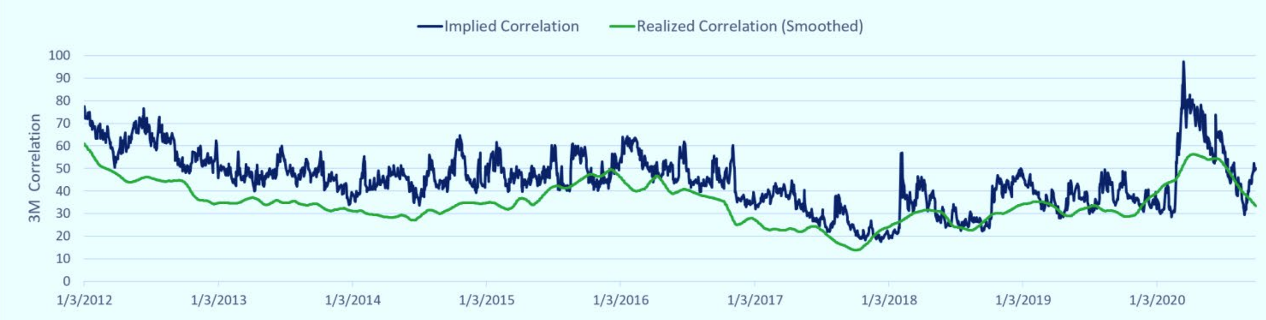
Ticker	Company	Weighting
AAPL	Apple Inc.	12.26%
MSFT	Microsoft Corporation	9.14%
AMZN	Amazon.com, Inc.	8.89%
TSLA	Tesla, Inc.	4.50%
META	Meta Platforms, Inc.	3,57%
GOOG	Alphabet Inc.	3.14%
GOOGL	Alphabet Inc.	2.86%
NVDA	NVIDIA Corporation	2.67%
PYPL	Paypal Holdings, Inc.	2.27%
CMCSA	Comcast Corporation	1.98%

It is typically the case that the top 10 stocks in QQQ represent around half the capitalization of the index. On the table above, the top 10 stocks account for 51.30%. This large representation of just a few stocks in the index makes it suitable for implementing dispersion trades. In addition, QQQ as well as its top constituent stocks are all highly liquid and widely traded. This liquidity in the underlying translates to strong liquidity in the options market as well as market makers take in the lowered costs of delta-hedging.

1.2.2 Dislocation Edge vs Risk Premium

We have now given a viable plan for the execution of our strategy that takes into account liquidity constraints. However, we have yet to answer the critical question of why we should expect our strategy to provide good risk-adjusted returns over long periods of time. Successful back tests on data as defined by returns alone, or even risk-adjusted return parameters, may well have occurred by chance. It is therefore necessary for us to understand the sources of our edge.

In the graph below we compare the CBOE 3M Implied Correlation Index with a smoothed realized correlation measure (60-day rolling window) over the same underlying. Daily return volatilities were annualized using a 252-day period. We can see that the majority of the time, realized correlation is slightly below the CBOE 3M Implied Correlation Index.

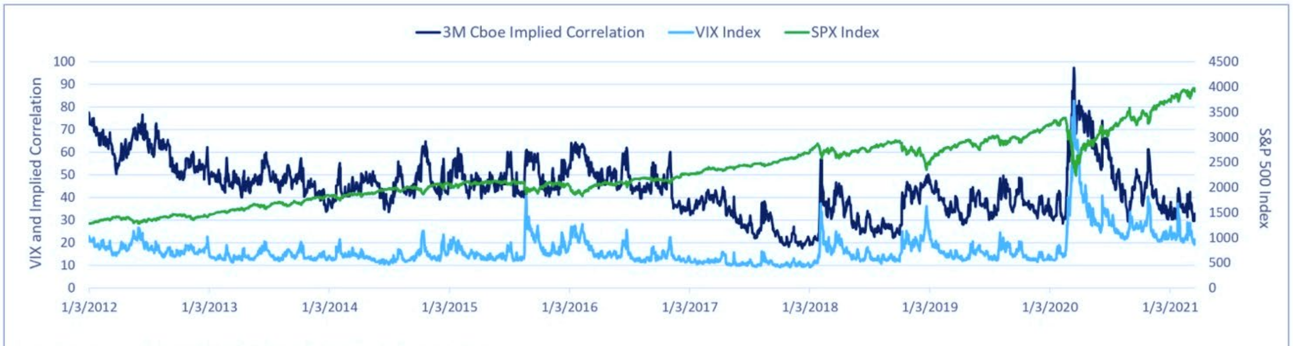


Source: Cboe Hanweck PhD and S&P SPICE Data March 17, 2021

Why is this the case? Recall that the implied correlation is the implied volatility of the index divided by the weighted average volatilities of its components. Academic literature suggests that the difference between implied and realized volatility can be explained by higher expectations of future (i.e. implied) volatilities relative to realized historic volatilities on the index compared to its constituent parts *(CBOE, 2021)*. The financial interpretation of this fact is an “overestimation” of systematic risk compared to realized systematic risk, since the volatility of a well-diversified index is considered exposed only to systematic risk.

Viewed from another perspective, correlation quantifies diversification benefits that market participants expect. Decreasing correlation reduces a portfolio's overall volatility beyond the risk of its components thereby improving the agent’s risk/return tradeoffs. Positive correlation, on the other hand, spikes, and indicate lower diversification benefits and therefore increased systematic risk.

The graph below demonstrates the relationship between the VIX Index and the CBOE 3M Implied Correlation Index. Note that periods of large volatility increases are generally caused primarily by broad market systematic factors, (e.g. March 2020 COVID-19 Crisis) and we see that large declines in the S&P 500 index is matched with a spike in both the Implied Correlation and VIX Indices.



Source: Cboe Hanweck PhD and S&P SPICE Data March 17, 2021

If we consider sudden market movements (typically downwards) and periods of high volatility as "tail events" then the apparent consistent overestimation of systematic risk by implied correlation may simply be reflecting the fact that in the real world, the distribution of returns isn't normal, but is rather on some distribution with a much higher kurtosis. The strong dependency between market returns, volatility, and correlation therefore suggests that the premium from dispersion trading may come from the same fundamental source as the equity risk premium.

Put another way, high correlations imply high equity risk premia (less diversification so higher risk in the index and therefore higher returns required), and vice versa. But stocks drift upwards and crashes downwards (often together), so a position exposed to positive correlation (correlated with equity risk premia) cost carry. Implied correlation trades at a premium to realized correlation because those selling correlation via dispersion trades are capturing a risk premia in a source of carry correlated with conventional equity risk premia. Index options "should" be overpriced because it's an exposure to systematic risk premium.

It is therefore not immediately evident that our strategy has real risk-adjusted edge over the long run, since profitability can be traced both to correlation risk premium as well as market inefficiencies, with only the latter representing alpha. Luckily, a large body of empirical research show that a nontrivial part of trading edge does in fact derive from market inefficiencies.

For example, in *(Marshall, 2009)* the author states, after a broad survey of S&P 500 derivatives data, that "*from a trader's perspective, index option implied volatility tended to be more often rich and component volatilities tended to be more often cheap*." Since shorting correlation involves selling index options and buying component ones, we can conclude that the short correlation trade does in fact capture some relatively value.

In *(Pierpaolo Ferrari, 2019)* the authors found that statistically "*the risk adjusted return of [dispersion trading] used in this empirical analysis has beaten [an appropriately leveraged] buy-and-hold alternative on the S&P 100 index, providing a significant over-performance and a low correlation with the stock market.*" They conclude that their findings also “*provide an evidence of inefficiency in the US options market and the presence of a form of “free lunch” available to traders focusing on options mispricing*.”

But where does this edge actually come from? Who are our counter-parties who deviate from risk-neutrality and therefore allow us to gain edge? In *(Deng, 2008)* and *(Hunter, 2004)* the author points to two main explanations for trading edge:

1) Demand for Index Protection: Institutional investors with large, long positions will buy puts for protection, therefore pushing up index volatility in puts and calls (by put-call parity), hence keeping the index volatility surface higher relative to the average basket volatility surface. This occurs because some large institutional investors such as pension funds have risk-averse mandates. A dispersion trade where we buy Vegas in constituents to hedge our Vega sells in the index can therefore be seen as a liquidity enhancing strategy. The good leg of the trade is in selling volatility, but we buy Vegas to allow ourselves to trade more.

2) Underestimation of Event Risk: Market participants may be underestimating the long-run probabilities of extreme, idiosyncratic events such as bankruptcies, corporate scandals, hostile takeovers, and M&A deals, which favor a long dispersion position since they explode single-stock volatility but have no significant effect on index volatility. This suggests that the long-run price of some equity options may be undervalued (relative to index options, that is). This suggests that some of our trades in index constituents may themselves have positive expectancy.

Having convinced ourselves of the probable edge in dispersion trading, we now turn to an examination of the risks involved in the position.

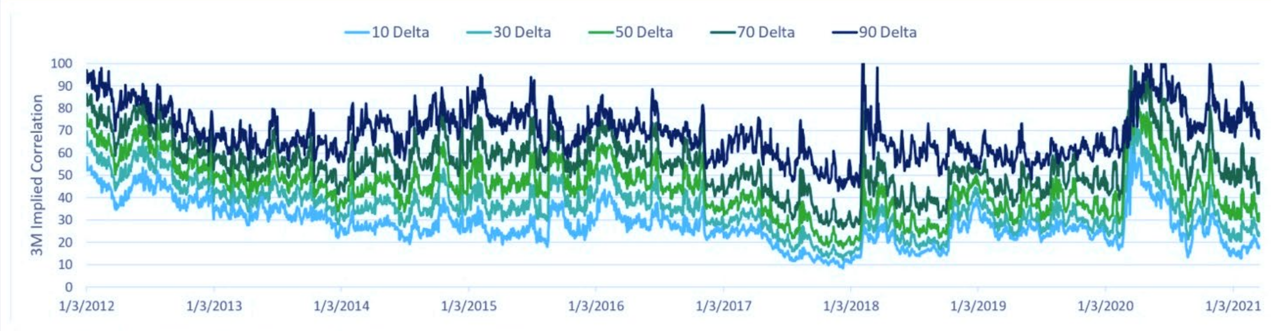
1.2.3 Other Risk Considerations

Point 2 in the section above shows why holding a position with negative exposure to dispersion is dangerous. By selling the constituent vol and buying the index vol, we expose ourselves to a lot of short equity gamma. We may be tempted to think that we have our long gamma on the index side against it and all is well.

The problem is that in a dirty dispersion position where we are only holding a few large equity positions against the index, events such as company bankruptcies and takeover/mergers can cause our book to implode. As such, while we are always happy to buy correlation at 0%, we will only take long dispersion positions in this paper.

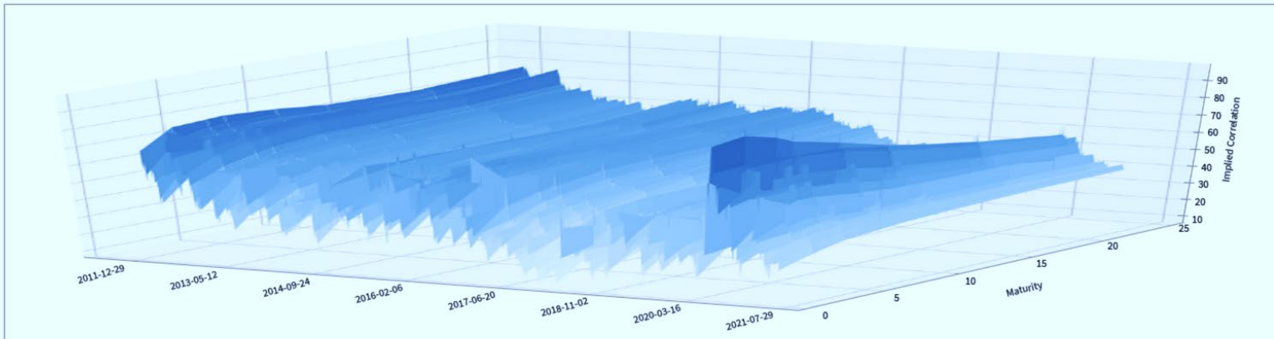
If we are to take on positions that have long dispersion / short correlation, then an important risk management question is how much correlations will increase by in the event of a market crash. We can examine this by looking at implied correlation skews, where the implied correlation is calculated using different options with different delta levels.

The skew relationship signifies market expectations of tail event codependences. For example, the spread between the 90 delta and 50 delta curves measures the curvature of the implied correlation skew curve and therefore the expected movement in implied correlation from a major market shock.



Source: Cboe Hanweck PhD and S&P SPICE Data March 17, 2021

What are potential ways to hedge our risk exposure to correlation? One potential solution lie in the term structure of correlation itself. Like implied volatility surfaces, the correlation surface below shows market expectations of correlation risk over time. Moreover, just like implied volatility surfaces We observe correlation backwardation during the Financial Crisis and COVID-19 Crash, indicating that market participants predicted a medium to long-term reduction in correlation, indicating a similar expected decrease in SPX volatility. Therefore, maybe using longer-dated options in our trades can partially insulate us from the downsides of correlation risk. However, this comes with two costs. The first is that the gains from the trade will also be lesser if longer-dated options are used since we are buying less gamma (though also paying less theta) over time. Moreover, we know that front-month options are typically significantly more liquid than long term options.



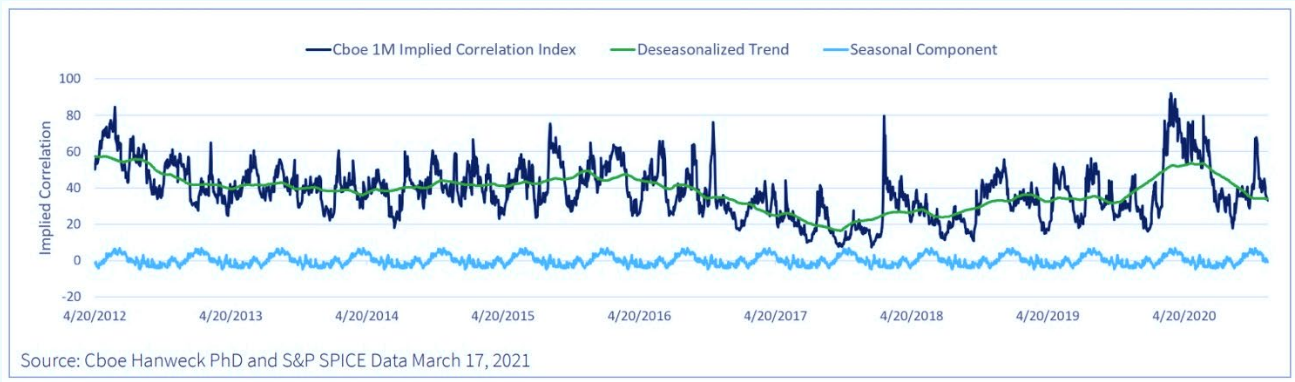
Source: Cboe Hanweck PhD and S&P SPICE Data March 17, 2021

Another possible means by which we can reduce our risk is to hedge it against some index (VIX or CBOE 3M Implied Correlation Index, for example), but at that point there is no guarantee that the cost of our hedge will exceed the edge that we gain in each trade. A more basic technique of reducing sizing appears to be the best means of risk reduction.

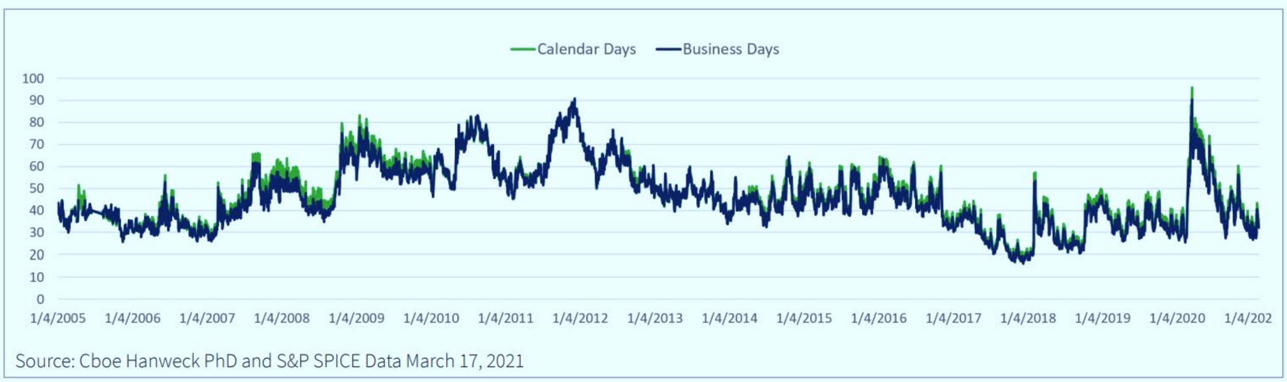
The other two sources of risk that we consider are the seasonal effects of earnings (quarterly) and of the weekend. The graph of de-seasonalized trend and seasonalized component of the CBOE 1M Implied Correlation Index shows that quarterly seasonal pattern is observable.

Considering that most firms release earnings within close intervals, we would expect higher investor uncertainty during earning periods to translate into amplified short maturity option implied volatilities across most component stocks and implied correlations.

However, this effect is not extreme when compared to the non-seasonal components of implied correlation movement.



In addition, our implied correlations are calculated using business day implied volatilities. Moving from business days to calendar days may introduce a well-known ‘Friday Effect’ where market makers remove quotes at the end of a trading week to take the weekend out of the implied volatilities. Therefore, a drop in Friday volatilities followed by an increase during the start of the next trading week may be expected. However, based on our graph below, there is no discernable difference caused by the ‘Friday Effect’.



This suggests that our methodology for calculating implied volatilities is robust. We have now considered a range of risk factors both in our methodology and position. It is now time to move on to implementation as well as the tuning of hyperparameters.

1.3.1 Implementation

Before implementing the strategy, we analyze the data to check for edge. We compared the 50-delta implied correlation and a 1 month smoothed realized correlation from Jan 1 2015 to Jan 1 2020 on QQQ. In *(Bossu, 2006)* it was shown rigorously that under reasonable conditions the average correlation within an index is well-approximated by the index volatility divided by the capitalization weighted volatility of components. Using this metric, the implied correlation was on average 8.74% higher than the realized correlation, this suggests that a dirty dispersion trade can be profitable. Moreover, we are also reasonably sure that the spread between implied and realized corr is mean-reverting because ADF statistic of -4.12 on the residuals of the regression between realize and implied correlation suggest that that we can reject the null hypothesis that there is a unit root in the residual's autoregressive process of any order n. The resolutions we sought in this analysis were daily, as there is no intra-day trading component in our dispersion strategy.

Given the risk and liquidity considerations that we have described, the implantation of our strategy will deviate slightly from the "full" dispersion trade. In particular,

- We have chosen **Invesco QQQ** as the index for the dispersion trade.
- The hedge against the index short vol will be **long vol in 10 stocks** in QQQ.
- These 10 stocks will be the top stocks in QQQ as of Jan 1, 2020.
- **No rebalancing** as stocks drop in and out of the top 10 to minimize price impact.
- **No options rolling** due to liquidity concerns; hold positions till expiration.
- Positions are established monthly and will start off **Vega Neutral**.
- **Daily delta hedging** with underlying stocks to manage risk.

The reader may note at this point that we are executing a typical mean-versal strategy where we find a metric (in this case, the correlation spread) that is mean-reverting and we sell it when its high and buy it when its low (or in this case, do nothing). For our particular strategy, there are a few natural hyper parameters:

- **Realized Correlation Rolling Window**: The length is some positive integer equal to the number of days we consider when calculating realize correlation.
- **Sizing:** How many contracts to trade in each of the components, as a factor of the total asset value of the portfolio.
- **Implied correlation Delta**: This is some positive integer equal to the delta of the call options we use in order to calculate implied correlation.
- **Entry Threshold**: The difference between implied and realized correlation required to enter into a dispersion position. May be expressed as a standard deviation of the historical spread value over some period of time.
- **Exit Threshold**: The difference between implied and realized correlation required to exit a dispersion position.
- **Stop-Loss Threshold**: The difference between implied and realized correlation required to exit a dispersion position. This is a larger number than the entry threshold and signals that the market regime for correlation has shifted.

When conducting back tests for trading strategies, it is crucial to factor in trading costs and liquidity impacts. Failure to do so can lead to inaccurate results that do not reflect the true performance of the strategy in real-world trading conditions. Trading costs, such as commissions, fees, and slippage, can significantly eat into profits and potentially render a profitable strategy unviable. For our **Brokage & Exchange Fee Model,** we use the InteractiveBrokersFeeModel which models the fees of Interactive Brokers. For our **Slippage and Price Impact Model,** we use the VolumeShareSlippageModel which calculates the slippage of each order by multiplying the price impact constant by the square root of the ratio of the order to the total volume. The precise expression is

$$\min\left(\frac{|\text{Order Quantity}|}{\text{Bar Volume}}, \text{Volume Limit}\right)^{1/2} \times \text{Price Impact}$$

We will use the default setting where volumeLimit = 0.025 and priceImpact = 0.1 which was tuned by quant connect to reflect an accurate cost of large transactions. The theoretical justification for the square root law of price impact is strong, but more importantly it is empirically verified (Bloomberg, 2020).

1.3.2 Hyperparameters

A naïve approach is to tune these five hyperparameters by grid search. However, that presents several problems. The first major issue is that we may be overfitting our model to in-sample historical data (Jan 1 2015 to Jan 1 2020) which would make our out of sample (Jan 1 2020 to April 1 2023) performance worse due to overcompensation as discussed in *(David H. Bailey, 2014)*. Furthermore, it was also noted by *(David H. Bailey, 2014)* that with five years of financial data, we should test at most around 40-50 independent hypothesis. Yet with 5 hyperparameters, it implies we should have at most around 2 candidates for each hyperparameter. Therefore, it seems prudent to decide on certain hyperparameters *a priori* based on our understanding of the markets rather than tuning them by in-sample data.

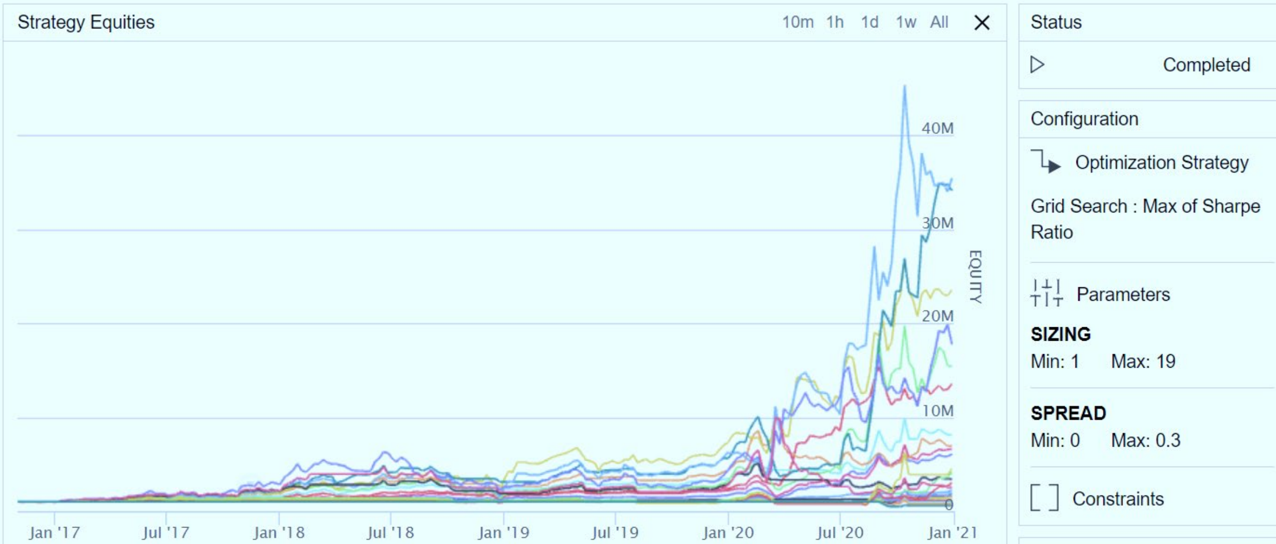
In particular, we can set implied correlation delta = 0.5 since we feel that **ATM options** are the most liquid and therefore best reflect the market expectation of volatility. Moreover, we can set **Exit Threshold = 0%** to avoid holding a short dispersion position. Moreover, because volatility is typically mean reverting, we expect that something like **monthly average** to be reasonably representative. Finally, we hold that **stoploss = 0.82** since that is the 95% percentile of correlation and we feel that correlation being that high implies that usual market conditions have broken down and we are no longer confident in our edge.

This leaves two hyperparameters to tune, the entry threshold and the sizing. However, even this does not alleviate all the problems that we will encounter in grid search. Recall previously that a position that is long dispersion (and therefore short correlation) has positive carry since it is exposed to macroeconomic risk and therefore earns a risk premium. As such, if we run the algorithm on a period when macroeconomic risk isn't realized, then we may be told by the back test that we should short correlation no matter what. Put another way, if we were testing a class of investment algorithms that simply buys and holds QQQ, and the only parameter of the algorithm was how much leverage we should take on, then back testing the algorithm on data during a bull market would give us an unreasonably high answer. The problem we face in back testing hyperparameters for our dispersion trading algorithm is not so different. The only difference is that our leverage is created not by borrowing but by entering into highly volatile options positions.

The implication for us is that when we are optimizing our hyperparameters, we should not be optimizing for returns. Instead, we should perhaps optimize for a risk-adjusted return metric such as Sharpe Ratio or PSR, in conjunction with a risk metric such as drawdown. Moreover, these metrics are incomparable. It does not really make sense for us to optimize some function that is a linear combination of the Sharpe Ratio, the Information Ratio, and other metrics, as such an optimization would not be interpretable. As such, it may be better to simply try different hyperparameters and see what they output.

1.3.3 Backtesting & Live Trading

For our **in-sample parameter optimization** we choose period 1/1/2017 – 1/1/2021.



The sizing parameter n is given as the number of contracts of each index component that we trade per $1,000,000 of capital. For example, if $n = 5$ and our position is worth $2,000,000 then we trade 10 contracts on each index component and hedge by selling the necessary number of contracts on QQQ to be Vega neutral. The spread parameter s is given as the amount by which the implied correlation needs to be above the rolling realized correlation for us to enter into a position. For example, if $s = 0.1$ then we enter into a dispersion position only when implied corr > realize corr + 0.1. The sizing parameter has increments of 3 and the spread parameter has increments of 0.01 up to their max values.

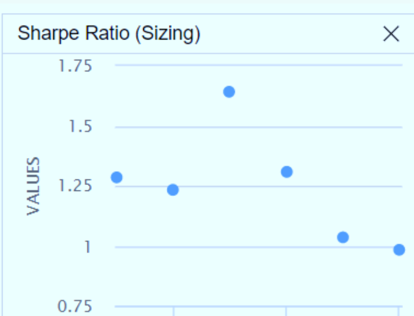
PSR ↓	Sharpe Ratio	Net Profit	Drawdown	Sizing	Spread
83.666%	2.032	3,512.246	67.3%	13	0.1
71.31%	1.585	1,232.114	39.1%	7	0
71.016%	1.285	93.721	7.5%	1	0
63.502%	1.404	710.473	36%	7	0.1
61.542%	1.158	51.102	7.4%	1	0.1
59.807%	1.619	2,233.894	51.1%	13	0
58.955%	1.308	510.977	45.3%	4	0

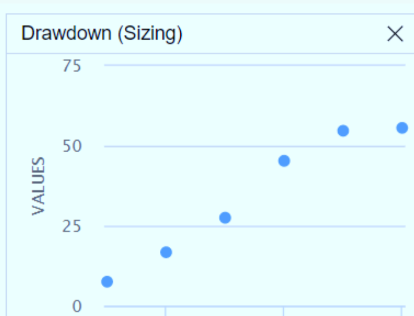
We tested a total of 28 combinations and sorted the results by PSR (Probabilistic Sharpe Ratio). We chose to sort by PSR instead of Sharpe Ratio for two main reasons. First, PSR can be interpreted as the confidence at which we can reject the null hypothesis of the strategy having no alpha in a hypothesis test, which allows for a clear assessment of statistical significance. In addition, Sharpe Ratios are effective when the returns follow a normal distribution, but financial data often exhibit non-normal characteristics, such as skewness and kurtosis. PSR is more robust to these deviations, providing a better assessment of a strategy's risk-adjusted performance *(Marcos Lopez de Prado, 2011)*.

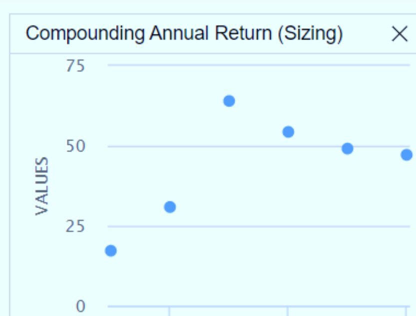
The strategy with the highest PSR we reject due to its large drawdown. The next two strategies have acceptable PSRs and so do the two following that. We made the decision to set **spread = 0** since having spread = 0.1 seems to reduce our net profit significantly while not decreasing the drawdown as much. We decided to further tune the sizing parameter between 0 and 7 since it appears to have a significant effect on profit and drawdown as well as PSR and Sharpe Ratio. Our results are displayed below.

We can see that **sizing = 3** results in both the highest PSR, Sharpe, and CAGR (Inferred from net profit) of the tested values. The drawdown of 27.5% is acceptable especially against the benchmark drawdown of over 50% of SPX during the covid pandemic.

PSR	Sharpe Ratio	Net Profit	Drawdown	Sizing
80.685%	1.642	687.219	27.5%	3
71.016%	1.285	93.721	7.5%	1
61.249%	1.233	206.885	16.7%	2
58.955%	1.308	510.977	45.3%	4
36.153%	1.035	429.303	54.7%	5
32.085%	0.983	401.047	55.6%	6

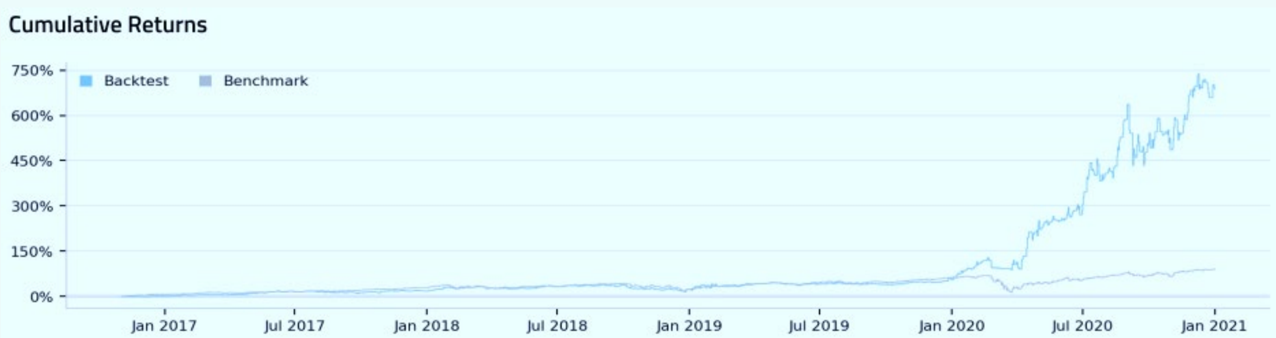






We now look at the back test report in more detail to give interpretation to the strategy.

PSR	80.685%	Sharpe Ratio	1.642
Total Trades	1273	Average Win	1.47%
Average Loss	-0.42%	Compounding Annual Return	64.018%
Drawdown	27.500%	Expectancy	0.731
Net Profit	687.219%	Loss Rate	62%
Win Rate	38%	Profit-Loss Ratio	3.52
Alpha	0.354	Beta	0.863
Annual Standard Deviation	0.284	Annual Variance	0.08
Information Ratio	1.361	Tracking Error	0.247
Treynor Ratio	0.54	Total Fees	$1622.50



Looking at our IS returns graph, we see that although a 64% CAGR is impressive, it only emerges from certain market conditions. In fact, returns seem to follow the benchmark SPX returns for most of the duration of the strategy. We may infer that from this IS backtest that for our OOS backtests, we can expect some to simply following market trends (recall we have exposure to the same fundamental risks) as well as periods of significant outperformance (where the need for liquidity provisions creates genuine dislocations that we can take advantage of.

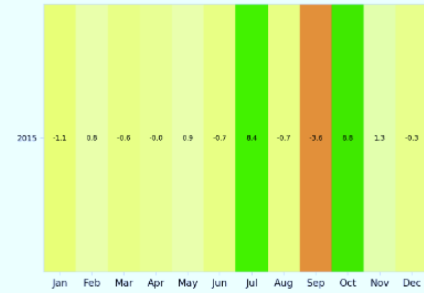
For our OOS backtests, we chose the 3 periods 1/1/2015 – 1/1/2016, 1/1/2016 – 1/1/2017, and 1/1/2021 – 1/1/2022. Of these periods, 2 were suggested to us in-class but we decided to substitute 1/1/2015 – 1/1/2016 for the 1/1/2010 – 1/1/2011 period because many of the stocks that we selected to use to hedge were not traded back in 2010. OOS Results below.

1/1/2015 – 1/1/2016

Key Statistics

Runtime Days	365	Drawdown	6.3%
Turnover	2%	Probabilistic SR	48%
CAGR	12.9%	Sharpe Ratio	1.0
Markets	Option,Equity	Information Ratio	0.7
Trades per Day	0.7	Strategy Capacity (USD)	100M

Monthly Returns

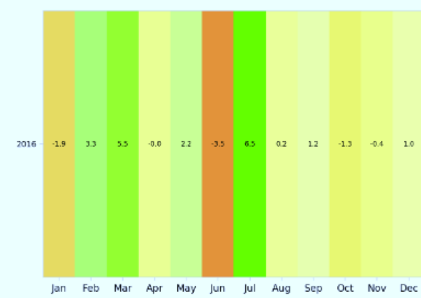


1/1/2016 – 1/1/2017

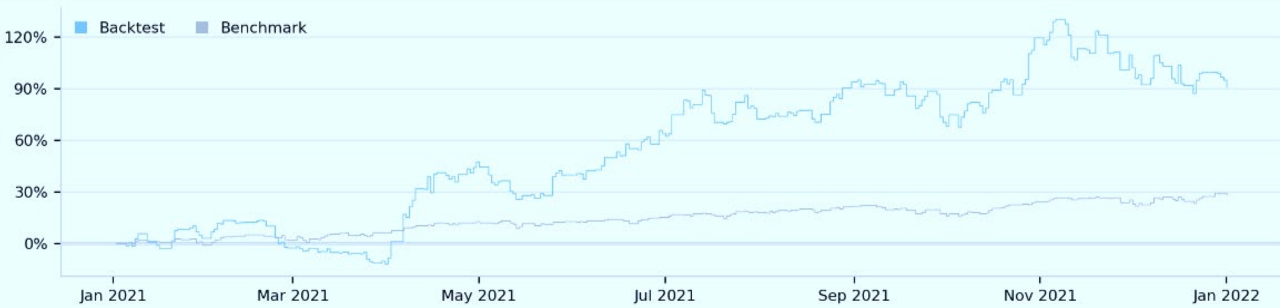
Key Statistics

Runtime Days	365	Drawdown	5.8%
Turnover	3%	Probabilistic SR	53%
CAGR	11.8%	Sharpe Ratio	1.1
Markets	Option,Equity	Information Ratio	0.0
Trades per Day	0.8	Strategy Capacity (USD)	1.4M

Monthly Returns



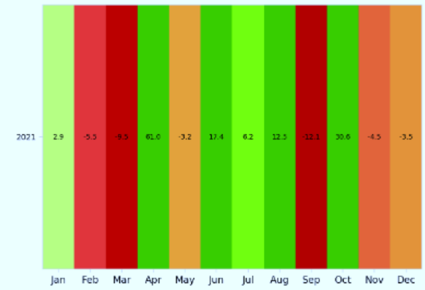
1/1/2021 – 1/1/2022



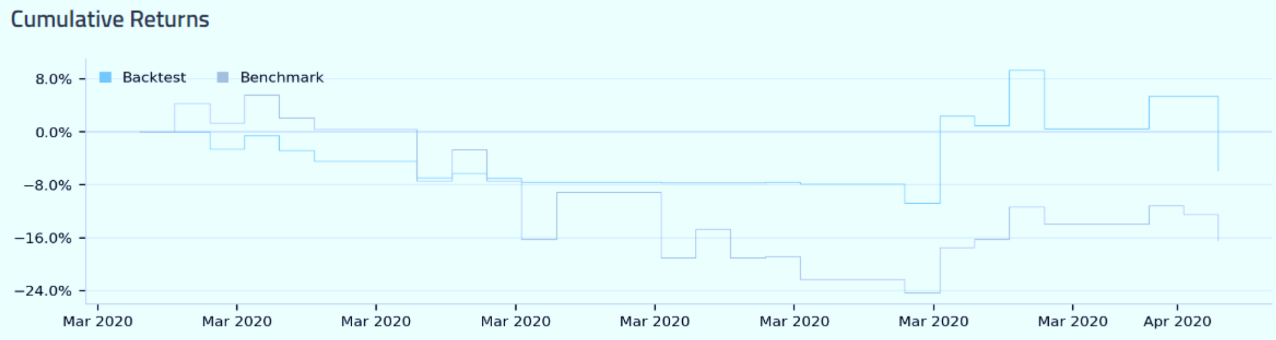
Key Statistics

Runtime Days	365	Drawdown	22.4%
Turnover	9%	Probabilistic SR	68%
CAGR	90.7%	Sharpe Ratio	1.8
Markets	Option,Equity	Information Ratio	1.5
Trades per Day	2.4	Strategy Capacity (USD)	66M

Monthly Returns



March 2020 Stress Test



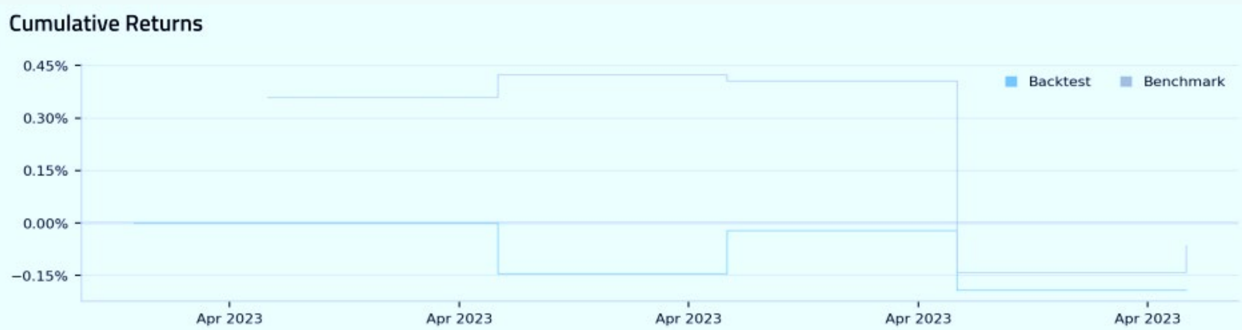
Key Statistics

Runtime Days	31	Drawdown	13.9%
Turnover	1%	Probabilistic SR	31%
CAGR	-49.8%	Sharpe Ratio	-0.4
Markets	Option	Information Ratio	0.7
Trades per Day	1.9	Strategy Capacity (USD)	60K

Our Stress test results in a drawdown less than 15% during a time when the market index has fallen over 25%.

This shows that the algorithm can perform even under intense market conditions.

April 17 – 21 Live Trading



Key Statistics

Runtime Days	5	Drawdown	0.2%
Turnover	0%	Probabilistic SR	25%
CAGR	-13.1%	Sharpe Ratio	-5.2
Markets	Option	Information Ratio	2.9
Trades per Day	11	Strategy Capacity (USD)	650K

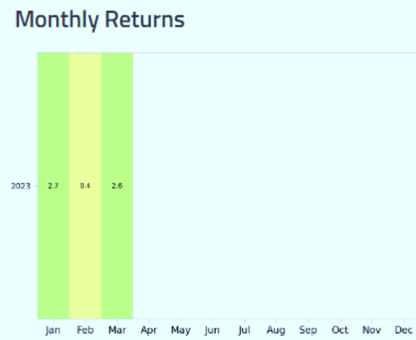
We had time to run live trading for 5 days (one trading week). However, we feel that the sample is not representative due to the small sample size.

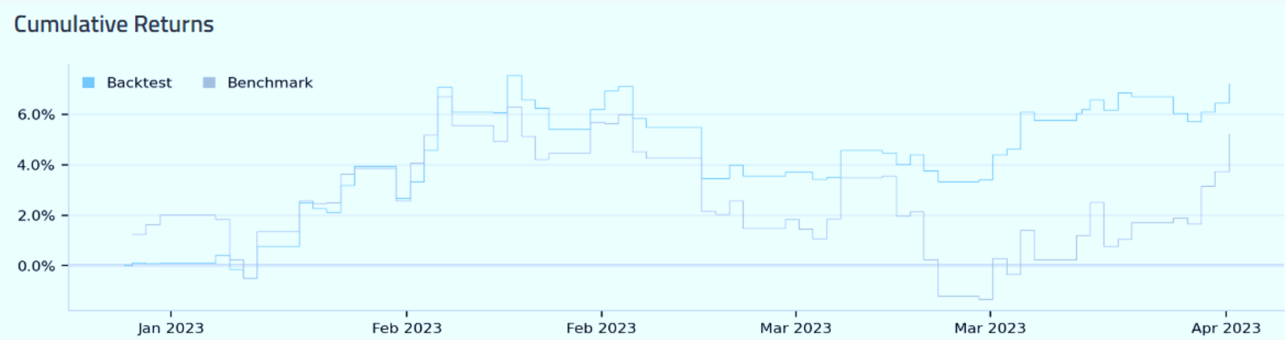
In our analysis of the backtesting results, the IS performance of the trading strategy closely follows the market for substantial periods before exhibiting brief instances of outperformance. Similarly, the OOS results display comparable characteristics, with the strategy aligning with the market for extended durations and then outperforming for short periods. These findings suggest that the trading strategy under consideration may possess a robust foundation for generating returns that are in line with the market, while also offering the potential for periodic outperformance.

January 1 – April 1 OOS (Assigned April 24 2023)

Key Statistics

Runtime Days	79	Drawdown	3.9%
Turnover	3%	Probabilistic SR	75%
CAGR	37.5%	Sharpe Ratio	2.4
Markets	Option,Equity	Information Ratio	1.4
Trades per Day	1.2	Strategy Capacity (USD)	14K





Summary of Results

Test type	Time	CAGR	Sharpe	IR	Drawdown	Daily STD*
IS	1/1/2017 – 1/1/2021	64.0%	1.6	1.4	27%	1.8%
OOS	1/1/2015 – 1/1/2016	12.9%	1.0	0.7	6.3%	0.5%
OOS	1/1/2016 – 1/1/2017	11.8%	1.1	0.0	5.8%	0.5%
OOS	1/1/2021 – 1/1/2022	90.7%	1.8	1.5	22.4%	0.9%
OOS	1/1/2023 – 4/1/2023	37.5%	2.4	1.4	3.9%	0.1%
Stress test	3/1/2020 – 4/1/2020	-49.8%	-0.4	0.7	13.9%	2.8%
Live Trades	4/17/2023 – 4/21/2023	-13.1%	-5.2	2.9	0.2%	0.0%

*Calculated as Annualized STD / $\sqrt{252}$

1.3.4 Limitations and Future Work

Our dispersion trading algorithm demonstrates potential for generating returns. However, it is essential to acknowledge some limitations that could impact overall performance.

One such limitation is the potential inaccuracy of our fee and slippage models. This may lead to underestimation or overestimation of transaction costs, ultimately affecting profitability and risk management. Another limitation is the hardcoded index component hedge, which does not allow for dynamic reallocation. This approach could result in suboptimal risk mitigation and a reduced ability to adapt to changing market conditions. As market dynamics evolve, a more flexible hedging mechanism may lead to improved outcomes. Lastly, the algorithm's undeployed capital remains idle, which depresses returns. By not utilizing this capital in other investment opportunities, the algorithm may be missing out on additional revenue streams that could enhance overall performance.

The implications of these limitations include potential inefficiencies in risk management, reduced adaptability to market fluctuations, and the underutilization of capital, ultimately leading to a possible decline in the algorithm's performance. Addressing these limitations remains a future endeavor that we can undertake as improved modeling, dynamic hedging strategies, and effective capital deployment may significantly enhance the algorithm's overall effectiveness and success.

In addition, we noted previously the presence of periods marked by high correlation with the broader market, as well as periods of significant outperformance. A deeper exploration of the specific market regimes under which the algorithm thrives could yield valuable insights and potentially unlock further enhancements in its performance. Recall that dispersion trading typically generates gains under two distinct scenarios: first, when there are substantial surprises in realized correlation and dispersion, as may be the case during macro events such as the COVID-19 pandemic or global chip shortages; and second, when there is a pronounced dislocation prompted by heightened demand for downside protection in the index.

The question of when dispersion trading is profitable is almost akin to a macroeconomic inquiry. This observation should not come as a surprise, as correlation itself serves as an indicator of the benefits of diversification, which is intrinsically linked to macroeconomic conditions. By examining the relationship between macroeconomic conditions and the algorithm's performance, we may uncover specific market regimes that offer particularly favorable opportunities for dispersion trading strategies.

2 Pairs Trading

2.1.1 Correlation of Stock Returns

Pairs trading is a common algorithmic trading strategy that seeks to profit from highly correlated financial instruments, such as two stocks or two ETFs. The idea is that, for a pair of highly correlated stocks P and Q, their respective returns can be captured in a linear relation. Mathematically, this can be expressed as:

$$\log P_t = \beta \log Q_t + \alpha + \epsilon_t$$

Here, we expect that the residual term ϵ_t to be mean-reverting. Assuming we have a reliable estimate of β and the linear relation holds in the foreseeable future, we can expect the residual term to decrease to zero and accordingly open our position. More specifically, we place $\frac{1}{1+\beta}$ of our capital C on stock P and $\frac{\beta}{1+\beta}$ of capital on stock Q. If the residual is negative, we long the spread (i.e. longing P and shorting Q); if the residual is positive, we short the spread (shorting P and longing Q).

Here we derive the expected P&L of a pairs trading strategy. With a total capital of C allocated to the stock pair P and Q, the expected PnL from time t to T is $\xi \frac{1}{1+\beta} (\epsilon_T - \epsilon_t)C$ where ξ is an indicator variable for whether we are shorting or longing the stock ($\xi = 1$ if we open a long position and $\xi = -1$ if we open a short position). Without loss of generality, we consider the case where we long the spread and calculate the PnL:

$$\begin{aligned}
\text{PnL} &= \frac{C}{1+\beta}\log\left(\frac{P_T}{P_t}\right) - \frac{C\,\beta}{1\,+\,\beta}\log\left(\frac{Q_T}{T_t}\right) \\
&= \frac{C}{1+\beta}[\log(P_T) - \log(P_t)] \;-\; \frac{C\,\beta}{1+\beta}[\log(Q_T) - \log(Q_t)] \\
&= \frac{C}{1+\beta}[\beta\log(Q_T) + \epsilon_T - \beta\log(Q_t) - \epsilon_t] \;+\; \frac{C\,\beta}{1+\beta}[\log(Q_T) - \log(Q_t)] \\
&= \frac{1}{1+\beta}(\epsilon_T - \epsilon_t)\,C
\end{aligned}$$

This is equal to $\xi\,\frac{1}{1+\beta}\,(\epsilon_T - \epsilon_t)C$ because $\xi = 1$ means a long position.

2.2.1 Universe Selection

The first step in pairs trading is to determine two stocks that are correlated. In many scenarios, this means stocks that are within the same industry or sector. Not wanting to overlook stocks because they are in less-frequent industries, we decided to go with a multiple-sector pairs selection approach. Using QuantConnect's Universe Selection API, specifically coarse and fine filtering, we were able to retrieve an ordered list of the top 500 U.S. stocks by market capitalization. This approach was designed to mimic the S&P500, which includes stocks from many different sectors; however, current QC capabilities don't allow for instant retrieval of a list of S&P500 stocks. Based on Morning Star's Asset Classification library in QC, these stocks were classified into 11 different industry sectors.

Basic Materials	29	**Utilities**	25
Consumer Cyclical	48	**Communication**	40
Financial Services	89	**Energy**	40
Real Estate	19	**Industrials**	54
Consumer Defensive	45	**Technology**	54
Healthcare	56		

From there, stocks were clustered and mean-reverting pairs were identified. These processes are described in the following sections.

2.2.2 Dimensionality Reduction with UMAP

To identify a pair of equities suitable for pairs trading, we need to quantify the correlation between the potential candidates. To examine (Leland McInes, 2018) whether a candidate pair really is mean-reverting, one common approach is to perform some statistical tests on the time series of residual term, such as the Dickey-Fuller Test or Augmented-Dickey-Fuller (ADF) Test. The idea is that the difference between the residual at two consecutive times should be negatively correlated with the residual with statistical significance. However, there are two problems with this method:

1. The ADF test may be too computationally expensive when iterating over all pairs of the universe. Therefore, we need some criteria criteria to narrow down our candidate pool.

2. The ADF test may not be fundamentally justified. If the test is passed by pure chance between two companies that are not fundamentally related (e.g. do not share the same supply chain), then the pairs trading would fail. Ensuring that the pair both comes from the same sector provides an extra filter for to increase the probability of success.
3. The probability that the underlying correlation between the daily returns of the past N days of stocks might be obscured by market noise increases, especially as N increases (i.e. as we look further back into historical data).
4. If we represent the historical returns of stocks as vectors in a Euclidean space and measure the distance between them. Yet, due to the "curse of dimensionality," distance metrics in high-dimensional Euclidean space tend to concentrate around their mean, and therefore the Euclidean distance between every pair of stock becomes roughly the same. In other words, these high-dimensional vectors are scattered more sparsely as the dimension increases.

To tackle these shortcomings, we take inspirations from dimensionality reduction to transform the high-dimensional data points into low-dimensional representations that preserve meaningful features and properties. Dimensionality reduction includes linear and non-linear methods. Because linear dimensionality reduction methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) struggle with dataset that is not linearly separable, we believe nonlinear dimensionality reduction methods are better for our stock data.

Here, we choose UMAP (McInnes et al.), one of the most state-of-the-art nonlinear dimensionality reduction tools. UMAP is mathematically rooted in simplicial homology and has been found to be useful for high-dimensional dataset with a complex and non-linear structure, such as images, audio, and text data. It has the following advantages:

1. It preserves local structures, allowing the points in each cluster in high-dimension to retain their shape and form after the projection.
2. It preserves global structures, ensuring that the distance between clusters are well-preserved. In other words, the clumps do not "mush" together when being projected.

The UMAP algorithm has the following steps:

Step 1: Construct a graph of nearest neighbors, where the weight w_{ij} between vertex i and j reflect probability distribution and is higher if they are relatively closer to each other than to the remaining points.

Step 2: Initialize the low-dimensional embedding using spectral embedding.

Step 3: For each vertex i and each edge (i, j), do the following:

- Apply an attractive force proportional to w_{ij} on the embedding for i-th point towards the embedding for the j-th point.
- Choose a non-neighboring point k of i and apply a repulsive force proportional to $(1 - w_{ik})$ to the embedding for i-th point away from the embedding for the k-th point.

It is worth noting that UMAP allows one to specify the number of neighbors in the graph construction stage and the output dimension. This gives us the flexibility to tune the output dimension that best captures the correlation between two candidate stocks in our algorithm design. We have tested the effectiveness of UMAP, taking into consideration both the runtime and the performance of the strategy, testing UMAP with the output dimension parameters of 2, 3, ..., 10. We have determined that 3 is the optimal parameter. From above, we can see that the minimum number of companies within a certain sector is 19. It is important that UMAP does not project clusters that contain any more than 19 data points, and so the parameters representing the number of neighbors in each cluster was set to be 10.

2.2.3 Metric Filtering

After dimensionality reduction is performed, we choose stocks that are close in their low-dimensional representation as candidates for pairs trading. To save computational time, we iterate through each sector (in contrast to the entire universe) and store the distances within a priority queue for future extraction. Again, we emphasize this strict filtering criteria that only stocks in the same sector can be considered. The reason for this is two-fold: (1) this cuts down our computational time by a factor of 11, and (2) this allows us to filter down the pairs with fundamental analysis.

Every time a new stock pair is needed for trading, we pop a pair (with the lowest distance) from the priority queue, check that it is mean-reverting with an ADF test, and use the obtained p-value to determine whether the pair is ideal for pairs trading, i.e., the residual series is mean-reverting. If it passes the ADF test, we append this to a list containing the currently traded pairs, where it will look for an entry point at every timestep.

Ideally, UMAP, with its lightweight implementation compared to t-SNE, would be called on every timestep of the algorithm. However, this was not feasible within the backtesting nodes of QuantConnect, with an estimated time of ~50.23 seconds to retain it on the data. Therefore, UMAP had be to called in the initial timestep, where it would store all the possible pairs within each cluster and their respective distances. This would be stored in a object attribute, where the priority queue can be called with high-effienency (O(1) time).

There are essentially two ways we can exit out of a trade: (1) we make profit when the exit threshold is triggered, or (2) we close our losses when the pair diverges too much past the risk threshold. If a pair is liquidated, the stored UMAP priority queue is immediately called, and we iterate through the least-distance pairs, ADF-testing them until a new candidate is added.

2.3 Limitations and Future Work

The regular pairs trading strategy has been significantly improved with diversification in 3 pairs with equal weights, dimensionality reduction with UMAP, and sector clustering. However, there still exist some limitations that could impact its performance. One limitation is the stochasticity of UMAP, which could possibly make our algorithm more volatile. UMAP's non-determinism stems from the fact that UMAP uses an approximate nearest neighbor search algorithm to speed up the computation of pairwise distances between high-dimensional points and uses Stochastic Gradient Descent to update the low-dimensional embeddings. To address this issue, one feasible way to improve and stabilize our pairs trading algorithm is to run UMAP multiple times whenever it's invoked and choose the output pairs that are most frequently similar enough in their low-dimensional representation.

We have been able to successfully implement another risk-management tool using sentiment analysis with Tiingo. Our original strategy was to look at the stream of news feed, and use the NLTK package to detect strong sentiment within the pairs that are in our holdings. If this was triggered, then we would lower our risk threshold from 3 to 2.5 standard deviations, which would protect us from losses resulting in sudden movements of the pair. This will essentially lower the upper bound on our losses if we have to liquidate due to extreme divergence. The risk of pairs trading mainly lies in divergence of the two underlying stocks. The loss incurred by divergence could be unbounded, and therefore in our algorithm we set a hard threshold for liquidating to achieve risk control. By default, our algorithm will liquidate if the residual of the linear equation between the returns of the two stocks has a zscore with absolute value higher than 3.

To effectively curb potential loss due to divergence of a pair for which we have an open position, we hope to detect divergence signals as early as possible and act preemptively. To this end, we use sentiment analysis on Tiingo news data to detect whether a pair of stock is likely to diverge soon from their previous linear relation. If signs of stock divergence are detected, we will liquidate before the absolute value of the zscore of residual reaches 3. In our implementation, we use the SentimentIntensityAnalyzer provided by nltk, which assesses input text and generates a positivity score, a negativity score, and a neurality score that are nonnegative and sum up to 1. The sentiment analyzer works as follows: it uses a lexicon containing a list of words and phrases, each with an associated sentiment score that ranges from -1 (most negative) to +1 (most positive). The analyzer will increase the sentiment score for a word that is preceded by an intensifier, such as "very" or "extremely", and decrease the score for a word that is preceded by a negation, such as "not" or "never". We consider a pair of stock likely to have an imminent divergent trend if the absolute value of the product of positivity score for one stock and the negativity score of the other stock is above 0.9, which only happens when the news sentiment regarding the two underlying stocks is drastically different. By setting the threshold high, our algorithm can predict a trend that is no longer mean-reverting, without giving us false alarms frequently.

However, the calling of this news data continuously had severely slowed down our algorithm. As a simple comparison, one run of UMAP takes ~50 seconds, but the computational inefficiency of Tiingo on the QuantConnect API had led to an additional 57 minutes of backtesting time. Unfortunately, the sentiment analysis had to be removed, but we had checked separately within the same time frames of our backtests and found that there were no times when the sentiment indicator was triggered. Therefore, at least for the backtesting periods shown in this paper, the sentiment indicator did not influence our statistics.

Even if we were able to implement efficient sentiment analysis, we would be severely limited in building sophisticated strategies and therefore would run the high risk of underfitting news data. Because backtesting on QuantConnect does not support more sophisticated NLP packages such as Huggingface, which provides APIs to use large language models like BERT, our sentiment analysis may not achieve the most state-of-the-art accuracy and may produce more false indicators than correct one.

3 Capital Allocation

In this section, we discuss our approach to capital allocation across our strategies. It is well known that portfolio diversification is an effective way for an investor to reduce risk. Assuming portfolio assets are not perfectly correlated, distributing one's capital across multiple different assets can unlock portfolios with a given level of risk whose expected returns exceed that of any individual asset. In our capital allocation approach, we treat our strategies as if they are financial assets with specified expected returns and volatilities. Markowitz Portfolio Theory is a classic approach to portfolio allocation which seeks to minimize portfolio volatility for a given expected return. Although Markowitz Portfolio Theory presents an intuitive theoretical strategy, Markowitz Portfolio Theory exhibits numerous practical drawbacks which, without correction, limit its utility in real-world portfolio allocation settings. As such, we implement a capital allocation strategy inspired by traditional Markowitz Portfolio Theory with key adjustments drawn from modern portfolio theory literature.

3.1 Markowitz Portfolio Theory

3.1 Background

Markowitz Portfolio Theory is a classic framework which applies a quantitative approach to the problem of capital allocation (Markowitz, 1952). Such an approach to portfolio optimization relies on three central assumptions as outlined by Zivot in *Introduction to Computation Finance and Financial Econometrics with R* (Zivot, 2021):

1. The returns and covariances of all assets are stationary, ergodic, and jointly normal. That is, it is assumed that the behavior of each asset follows a model of Gaussian White Noise (GWN), and that the behavior of the entire set of assets are fully characterized by mean returns and the covariance matrix (Zivot, 2021).
2. Expected returns, variances, and covariances are known. This assumption is practically impossible (must know forecast future asset performance), however necessary for the sake of Markowitz Portfolio Allocation (Zivot, 2021).
3. Investors prefer portfolios with higher returns and lower volatilities to portfolios with lower returns and lower volatilities. In this sense, an investor's utility function can be fully described by risk and return (Zivot, 2021).

Given these three assumptions, we may begin to express Markowitz Portfolio Theory mathematically. Suppose we must select a portfolio from a universe of N assets with expected returns $\boldsymbol{r} = [r_1, r_2, \dots, r_N]$ and expected covariance matrix $\boldsymbol{\Sigma} = [\sigma_1, \sigma_2, \dots, \sigma_N]$, where each σ_i is a vector of covariance terms. Suppose we would like to determine a portfolio weight vector $\boldsymbol{x} = [x_1, x_2, \dots, x_N]$. Then, Markowitz Portfolio Theory asserts the following fundamental optimization problem:

$$\min_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{\Sigma} \mathbf{x} \;\; s.t. \boldsymbol{r}^T \boldsymbol{x} \geq r_0$$

In the above equation, r_0 is the target level of return the investor desires. Thus, the above equation expresses the investors desire to minimize portfolio volatility ($\sqrt{\boldsymbol{x}^T \boldsymbol{\Sigma} \mathbf{x}}$) for a given level of portfolio returns, $\boldsymbol{r}^T \boldsymbol{x} \geq r_0$. One can add additional constraints to the above optimization problem according to their individual needs. For instance, we implement the constraint that all portfolio weights sum to 1 ($\mathbf{1}^T \boldsymbol{x} = 1$), where $\mathbf{1}$ is a vector of all 1's, as well as the constraint that all portfolio weights are positive (no shorting), ($x_i \geq 0 \; \forall \; i$). That is, our capital must be fully invested at any given moment, and we are not allowed short positions on any given strategy.

The above optimization problem is convex and achieves a unique solution so long as the covariance matrix is invertible. The optimal portfolio weights, $\boldsymbol{x}^*$, have a closed form solution which can be solved mathematically through Langrange multipliers (Zivot, 2021). Each optimal solution is considered an efficient portfolio allocation. That is, it describes a portfolio that achieves the minimum possible expected volatility for a given level of risk. Investors that adhere to assumption 3 above would, in theory, invest exclusively in efficient portfolios and thus limit their capital allocation decision to a tradeoff between potential risk-return combinations (Zivot, 2021).

3.1.2 Efficient Frontier Example

The set of all efficient portfolios is collectively known as the Efficient Frontier. The below figure graphs the returns and volatilities of 1000 simulated fully invested long-only portfolios of three stocks, JPM, GS, and WMT with random portfolio weights. The three assets are denoted by the black dots. As seen, the range of possible portfolios allows an investor to achieve a vast array of expected mean and variance combinations by selecting different portfolio weights. Notably, the investor can achieve portfolios with lower volatilities than any single asset through diversification. The graph also includes the efficient frontier of the three assets. As can be seen, the efficient frontier marks the set of portfolios with maximum possible returns for a given level of volatility. Since this portfolio is long-only with total weight set to one, the portfolio with maximal expected return is just one fully invested in JPM (since JPM has the highest expected return).

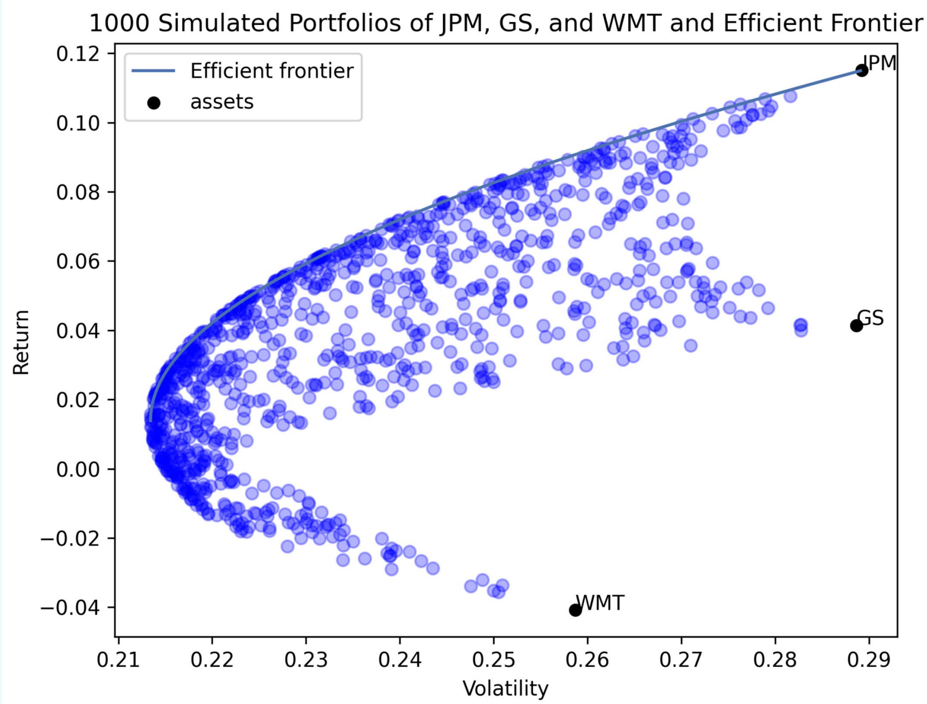


Figure 1: Plot of 1000 simulated portfolios and the efficient frontier for three stocks specified on 252 days of data. Graph partially generated with Pyportfolioopt library (Martin, 2021).

The shape of the above graph is known as the Markowitz bullet (it may be shown that the graph of any Efficient Frontier of non-perfectly correlated risky assets is the upper half of a hyperbola). In the limiting case where there exist perfectly negatively correlated assets, the Efficient Frontier may actually collapse to a straight line that intersects the y-axis (Zivot, 2021). That is, it would be possible to achieve a portfolio with zero expected volatility through the appropriate combination of the perfectly-negatively correlated assets (Zivot, 2021). The point on the efficient frontier with the lowest volatility (the furthest point to the left on the above graph) represents the minimum volatility portfolio. One would be able to extend the Efficient Frontier shown above by relaxing the constraints on portfolio weights. For instance, if shorting were allowed, the investor would be able to reach expected returns beyond the expected return of JPM, albeit at the cost of higher volatility.

In theory, one may include a “risk-free” asset in their portfolio (such as a US Treasury Bond) which achieves some level of return known as the risk-free rate with zero volatility (Zivot, 2021). In doing so, the efficient frontier would be reduced to a straight line from the risk-free asset to some point on the risky Efficient Frontier known as the tangency point. This line is known as the capital allocation line. Any investor following assumption 3 would be expected to divide their portfolio entirely between the risk-free asset and tangency portfolio, thereby choosing a point somewhere on this line according to their risk tolerance. The tangency point exhibits the property of having the highest Sharpe ratio of any possible portfolio, which is defined mathematically as follows:

$$\text{Sharpe Ratio} = \frac{r_p - r_f}{\sigma_p}$$

In the above equation, r_p is the expected portfolio return, r_f is the risk-free rate, and σ_p is the expected portfolio volatility.

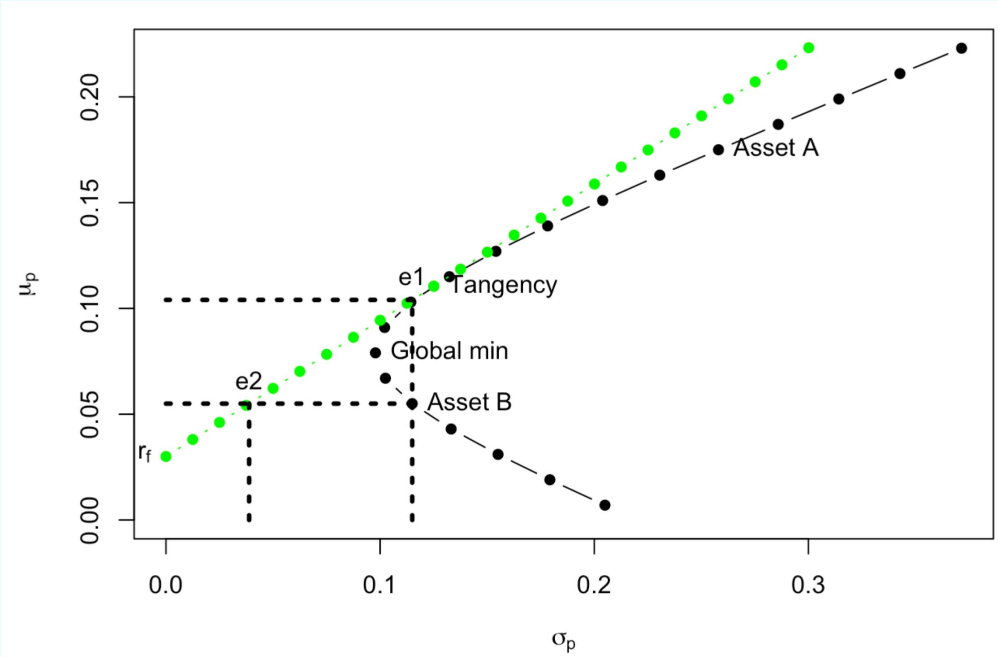


Figure 2: Graph of the efficient frontier and capital allocation line for a portfolio of two risky assets and a riskless asset. Image Source: Introduction to Computation Finance and Financial Econometrics with R (Zivot, 2021)

The above figure draws out the Efficient Frontier and capital allocation line for portfolios consisting of some combination of assets A and B. The capital allocation line intersects the Efficient Frontier at the tangency point, and all portfolios on the capital allocation line consist of linear combinations of the risk-free asset and tangency portfolio (image sourced from *Introduction to Computation Finance and Financial Econometrics with* R (Zivot, 2021).

3.1.3 Markowitz Portfolio Optimization Shortcomings

Markowitz Portfolio Theory presents an intuitive approach to capital allocation with mathematical underpinnings. However, Markowitz Portfolio Allocation often fails to achieve strong results in real-world settings. The shortcomings of Markowitz Portfolio Theory have been explored thoroughly in literature. In 2006, DeMiguel et. al. compared a naïve 1/N capital allocation strategy (evenly dividing capital across N assets) to a Markowitz mean-variance allocation strategy and a set of close derivatives of this model. The paper finds that, in out of sample data, none of the tested strategies consistently outperform the 1/N portfolio on several key metrics including Sharpe ratio (DeMiguel, Garlappi, & Uppal, 2006). *The Markowitz Optimization Enigma: Is 'Optimized' Optimal*, Michaud notes the "enigma" of limited adoption of Markowitz Portfolio Optimization in industry despite its widespread prevalence in academic circles (Michaud, 1989). Michaud offers numerous limitations of the Markowitz Portfolio Optimization which could potentially explain its limited popularity in practice.

Michaud lists error maximization as one of the driving reasons behind this apparent "enigma". Michaud notes how a Markowitz Mean-Variance (MV) framework is reliant on accurate estimates of future returns and covariances. If an investor uses sample returns and covariances as their estimates for future returns and covariances, any deviation from historical results could lead to suboptimal portfolio allocations. In this sense, a traditional MV approach would tend to underweight assets that underperform in-sample, yet overweight assets that overperform in-sample. Any estimation error in returns and covariances could have severe consequences for portfolio performance. In a 2021 paper, Mynbayeva, Lamb and Zhao show that MV optimization results in unequal portfolio weights even when all return means and variances are identical. Even if past returns and covariances are statistically identical to future returns and covariances, random fluctuation in past data could lead to suboptimal portfolio allocations. The paper authors also find that MV optimizers perform worse as the number of assets goes from 20 to 50 (Mynbayeva, Lamb, & Zhao, 2022).

Michaud also outlines how MV approach may lead to unstable solutions. Markowitz portfolio optimization depends upon inverting the covariance matrix of assets. If the covariance matrix is ill-conditioned (such as if multiple assets have nearly identical returns), it may lead to extremely unstable optimal portfolio weights. Best and Grauer demonstrate and closed-form expression which places an upper bound on the change in portfolio weights to changes in change in mean returns. They show that the change in portfolio weights is dependent on ratio of maximum to minimum eigenvalue of covariance matrix. The authors also suggest that this ratio may become higher as the number of considered stocked increases (Best & Grauer, 1991). Further, Michaud asserts that MV optimization may be misleading as it produces a single optimal set of portfolio weights when there may be other portfolios that produce nearly identical expected returns and volatilities with drastically different portfolio weights (Michaud, 1989).

Finally, Michaud notes how MV optimization may fail to consider other potential factors often key to investment decisions. For instance, MV optimization generally does not consider asset liquidity or transaction costs. If an MV optimizer produces wildly different optimal portfolios for updated input data, it may lead to high transaction costs as the investor may need to place substantial orders to re-optimize their portfolio. Further, MV optimizers may produce extreme portfolio weights on certain assets (Karl Härdle, Klochkov, Petukhina, & Zhivotovskiy, 2022). Such a portfolio may be infeasible when an investor must allocate capital across a wide number of assets. In our case, we want to distribute capital across all our strategies given the fundamental assumption that they are each individually likely to be profitable, and that diversification would help reduce risk.

3.2: Modern Adjustments to MV Optimization

There exists significant research into ways to mitigate the drawbacks of traditional Markowitz portfolio optimization. In a 2022 paper, Ortiz et. al. state that there exist multiple categories of proposed modifications to traditional Markowitz optimization, icluding Bayesian approaches, moment restriction, and portfolio restriction (Ortiz, Contreras, & Mellado, 2022). Our approach to portfolio optimization makes use of two distinct strategies, portfolio weight regularization and covariance matrix shrinkage.

In a regression context, regularization seeks to reduce the variance of a given model at the expense of increased bias. This concept may be extended to Markowitz Portfolio Theory as MV optimization essentially reduces to a constrained regression problem in which one seeks to find portfolio weights which minimize volatility at some level of expected return. One from of regularization, l_p, regularization, adds a term equivalent to the magnitude of the model parameters raised to the power of p. Two commonly used special cases of l_p regularization are the Lasso regression (p = 1) and the Ridge regression (p = 2). In 2008, Brodie et. al. examines the use of a tunable l_1 penalty term added to the MV optimization function. The researchers found that adding an l_1 penalty produced several positive effects on the optimization problem. For one, the researchers found that the penalty stabilized the problem, reducing the sensitivity of the optimal portfolio weights to changes in input data. Adding the penalty also naturally disincentivized short positions, as shorting generates portfolio weights whose absolute values become greater. While less applicable to our strategy, the researchers also found that the l_1 strategy encouraged sparse portfolios (Brodie, Daubechies, De Mol, Giannone, & Loris, 2009). The Ridge regression is also commonly used in the context of inverse problems (Carrasco & Noumon, 2012). We ultimately choose to implement l2 regularization as it encourages a broader distribution of capital than l_1 regularization (as the l_2-term coefficient approaches infinity, the optimal weights will converge to a 1/N equal weight portfolio).

In 2003, Ledoit and Wolf describe a process called shrinkage which aims to reduce estimation error in the covariance matrix of asset returns (Ledoit & Wolf, 2003). Ledoit and Wolf note that estimating the covariance matrix of a set of assets tends to be a very difficult task, with the sample covariance matrix formed from past data often being subject to significant estimation error. Instead, Ledoit and Wolf propose the following matrix, $\hat{\Sigma}$, to replace the sample covariance matrix:

$$\hat{\Sigma} = \delta F + (1 - \delta)S$$

In the above expression, S is the sample covariance matrix and F is known as the shrinkage target. δ is called the shrinkage constant and determines the extent to which the sample covariance matrix is "shrunk" to F. The authors state how the shrinkage target should be some highly structured matrix that nevertheless reflects qualities of the value being estimated. The authors propose the use of the constant correlation model in which the diagonal of F is equivalent to the diagonal of S, except all correlations are set to the average correlation across all assets. That is, all off-diagonal terms in F are set to the same value. This choice of F serves to limit the effect of extreme values in S by drawing them closer to the mean value. Ultimately, Ledoit and Wolf find that applying shrinkage may help improve the information ratio of a portfolio set to track a benchmark index on real stock data (Ledoit & Wolf, 2003).

3.3 Capital Allocation Strategy/Implementation

Our approach to capital allocation involves the use of l_2 regularization and shrinkage as described by Ledoit and Wolf. We make use of the Pyportfolioopt Python library which offers a convenient way to implement Markowitz portfolio optimization with customizable constraints and objective functions, among other features (Martin, 2021). A notable drawback of the Pyportfolioopt package, however, is its difficulty optimizing Sharpe ratio given the presence of additional objective function terms (such as the L2 loss). Since optimizing Sharpe ratio is not a convex optimization problem (without modification), the package implements a change of variables which may cause other objective functions to not work properly. As such, we used standard MV optimization with a specified risk tolerance instead of seeking to optimize Sharpe ratio. Further, the Pyportfolioopt package seems to only allow an efficient frontier which maximizes returns subject to a custom objective function (such as volatility plus the L2 regularization term). For our purpose, it may have been useful to visualize an efficient frontier which maximized returns *minus* the regularization term, for instance, subject to a given risk tolerance. This way, certain returns would be unachievable compared to the optimization problem without the L2 regularization term. With Pyportfoliopt's current implementation, adding an L2 regularization term instead makes certain levels of volatility unachievable (e.g., on the far left of the curve).

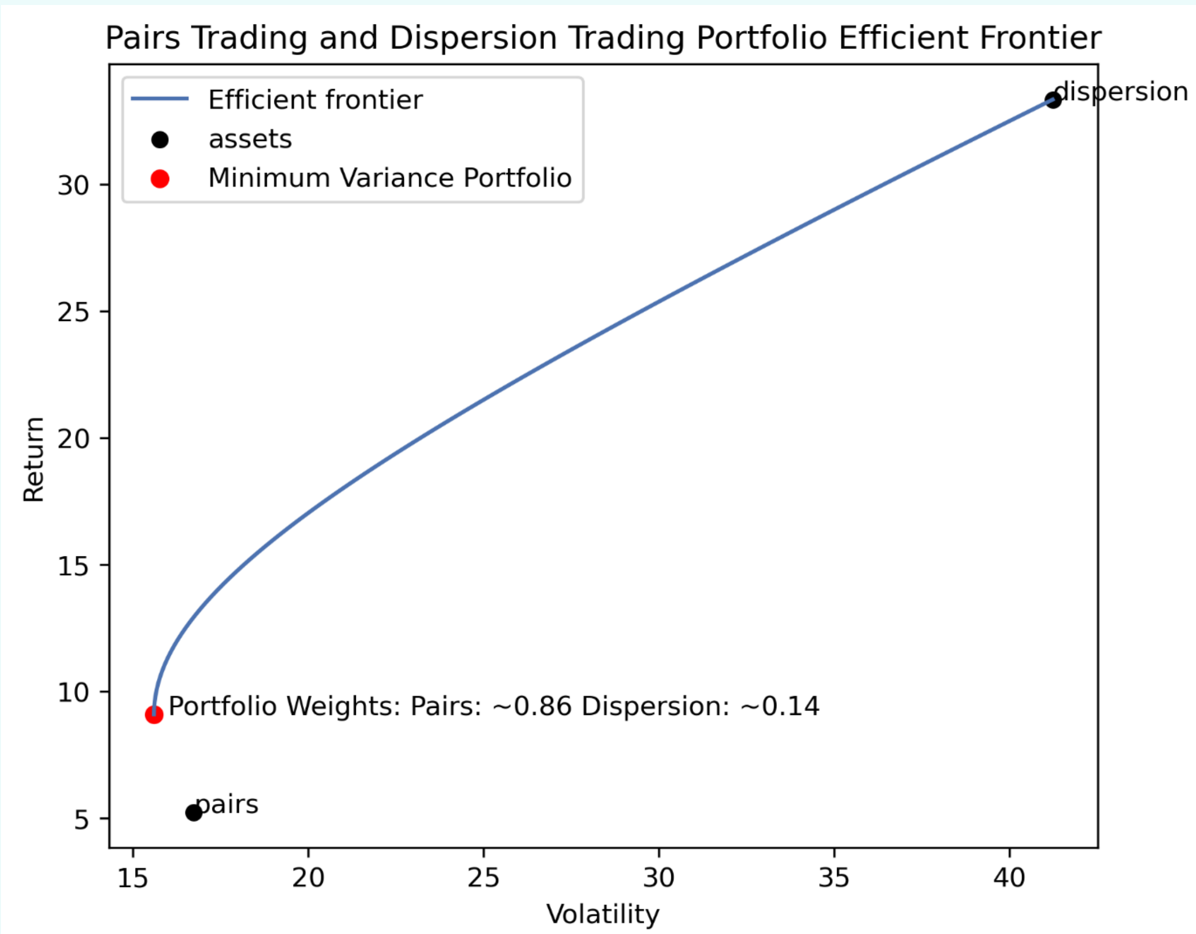


Figure 3: Graph of Efficient Frontier for portfolio allocation between pairs trading strategy and dispersion strategy on four yers of historical data for one particular back test. Red dot denotes minimum variance portfolio, and the associated weights chosen for live trading. Above graph partially generated with Pyportfolioopt package (Martin, 2021)

From the above graph, one can see our chosen portfolio expected return and volatility on four years of returns data for the pairs and dispersion trading algorithms. We chose to use the minimum variance portfolio, or the portfolio at the farthest-left point of the efficient frontier in the above graph, for live trading. The above graph also implements L2 regularization and Ledoit-Wolf covariance matrix shrinkage. This portfolio would ideally minimize the risk of our strategy in live trading, although it may reduce expected returns given past data. Although the pairs algorithm is stochastic in nature, we assume it would behave relatively similarly if the algorithm were repeated (and thus our capital allocation would apply to future data).

4 Results

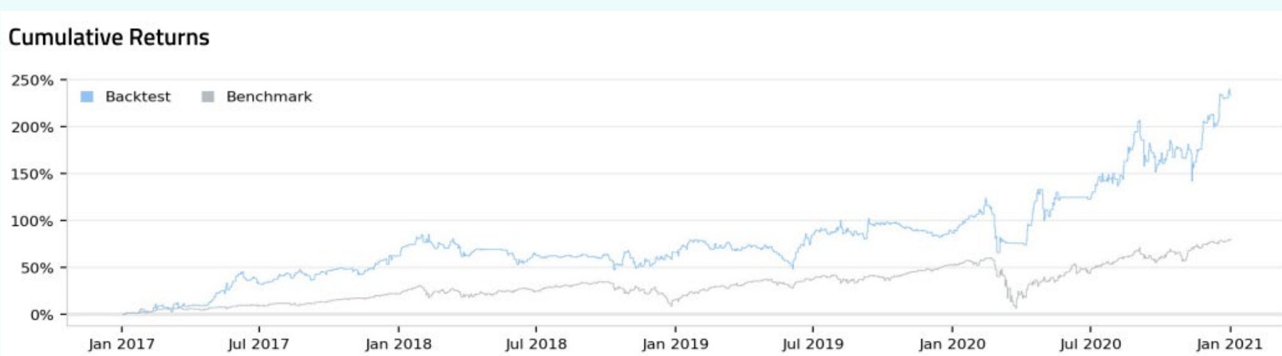
A summary of the results for the one IS period, four OOS periods, one stress period, and one live testing period is shown below. The fee and slippage models were implemented identically to the ones mentioned in the dispersion trading chapter.

Summary of Results

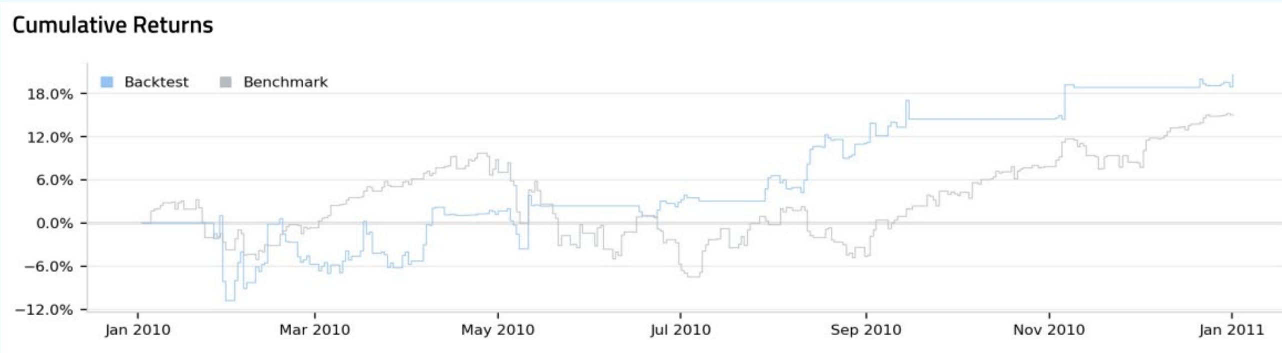
Test type	Time	Returns	CAGR	Sharpe	IR	Draw-down	Daily STD*
IS	1/1/17 – 1/1/21	233.71%	35.1%	1.2	0.6	25.8%	0.014
OOS	1/1/10 – 1/1/11	20.63%	20.6%	0.8	0.2	11.7%	0.011
OOS	1/1/16 – 1/1/17	17.02%	17.0%	1.3	0.3	10.6%	0.006
OOS	1/1/22 – 11/1/22	75.69%	96.3%	1.3	1.6	35.0%	0.047
OOS	1/1/23 – 4/1/23	16.01%	82.7%	2.8	1.4	3.3%	0.010
Stress test	3/1/20 – 4/1/20	2.16%	27.7%	0.7	1.3	7.7%	0.023
Live trade	4/17/23 – 4/21/23	0%	0%	N/A	0	0.0%	0.000

**Calculated as Annualized STD /* $\sqrt{252}$

For the IS period, the optimal portfolio was composed of a ratio of 0.19:0.81 between the dispersion and pairs trading strategies, which provided some balance between the high mean /high variance returns of the dispersion during this period, and the low mean/low variance returns of the pairs strategy. While the cumulative returns of 233.71% for the in-sample period significantly outperforms the S&P500 benchmark, the CAGR of 35.1% is significantly smaller than the previous CAGR in the dispersion trading. However, the profit potential in relatively static market conditions early in the period had improved, increasing its robustness within bear markets. Furthermore, the drawdown of 27.5% in the pure dispersion trading had slightly improved to 25.8%.

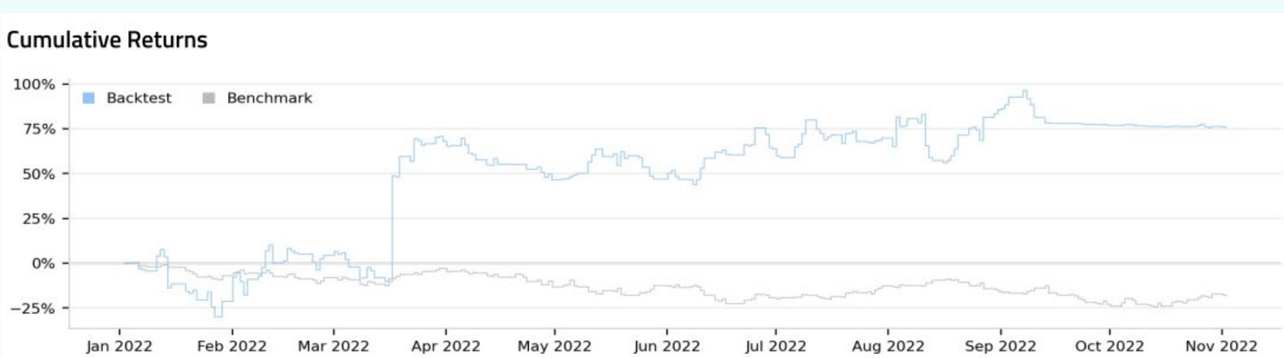


As mentioned before, several of the stocks that we selected to use for hedging were not traded back in 2010. Meta had IPOed back in May 2012, Tesla in June 2010, and Google had its stock split (into GOOGL and GOOG) in March 2014. Therefore, for the OOS period beginning in 2010, these stocks were substituted with the top largest components of QQQ, including Oracle, Cisco, and Qualcomm. Due to the infeasibility to keep this invariant across all periods, we expected the 2010~2011 returns to deviate from our expectations. As expected,

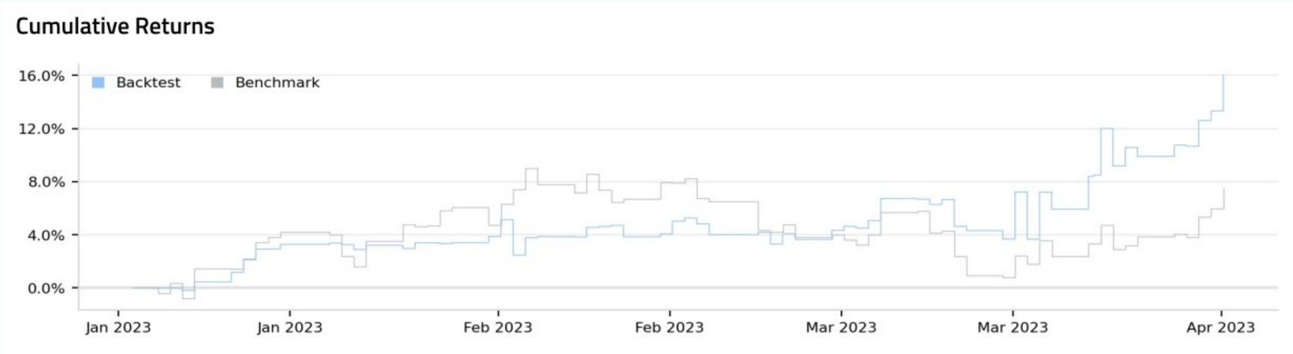


The performance of the total strategy in 2016 had performed relatively stable, with a max drawdown of 10.6% and maintaining an alpha of ~8% throughout the first three quarters.

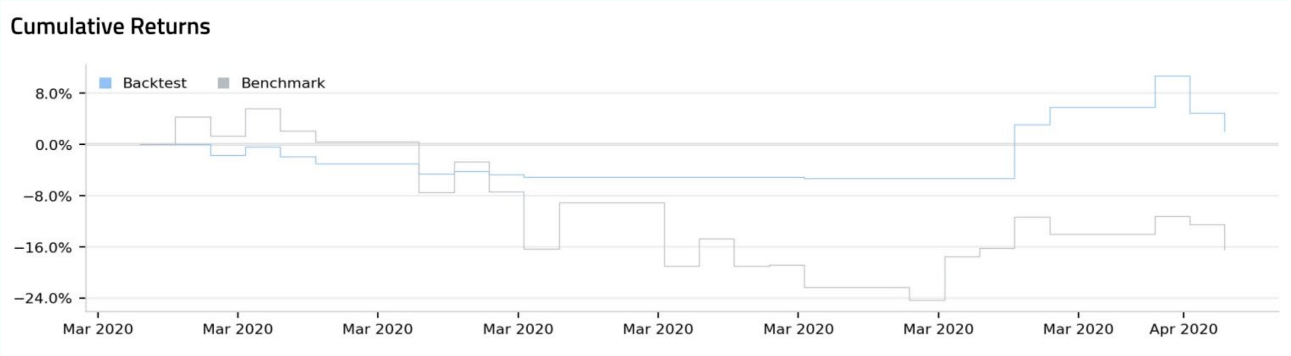
Similarly, to the 2017~2018 period within the IS period, our strategy had performed much better than the benchmark index within the 2022 bear market. Our optimal allocation from the start of this period had been to put the vast majority of our capital (93%) into the pairs strategy, while keeping 7% in dispersion. We believe that during the tumultuous times when the Fed had been rising interest rates, the peaking of the Russia-Ukraine war, and a 20-year low global economic growth rate of 2.7% the pairs trading was able to take advantage of statistical anomalies to enter into profitable positions ("World Economics Outlook", 2022).



The 2023 OS period shows that our algorithm performs well in extremely recent conditions, with returns of 16% within the first 3 months of 2023 and a Sharpe ratio of 2.8, indicating both strong returns and low variance.



Our stress test within the time period containing the explosion of the coronavirus shows that our algorithm performs well even under intense market condition. We can maintained strong stability with a 2.12% gain and a maximum of a 7.7% drawdown, compared to the 25% fall in the benchmark index.



The live testing period showed no trades, since none were executed during this time period, though the logs show that the pairs (BSBR/RJF, BBD/TRV, ENPH/IBM) have been stored, but the entry thresholds were not triggered.

References

Best, M. J., & Grauer, R. R. (1991, April). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *The Review of Financial Studies*, 315-342. Retrieved April 23, 2023, from https://academic.oup.com/rfs/article/4/2/315/1571031

Bloomberg. (2020, October 1). *Moving the market: A look at prices and impacts from trading*. Retrieved from https://www.bloomberg.com/professional/blog/moving-the-market-a-look-at-prices-and-impacts-from-trading/

Bossu, S. (2006). A new approach for modelling and pricing correlation swaps in equity derivatives,. *Global Derivatives Trading & Risk Management*.

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009, July). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences, 106*, 12267-12272. doi:10.1073/pnas.0904287106

Carrasco, M., & Noumon, N. (2012, June). Optimal Portfolio Selection using Regularization. Retrieved April 22, 2023, from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=424be1079b0c168d0462e3db0f7f471a1f688a8b

CBOE. (2021). *Implied Correlation*. Cboe Exchange, Inc.

David H. Bailey, J. M. (2014). Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance. *Notice of the AMS*.

DeMiguel, V., Garlappi, L., & Uppal, R. (2006, June). 1/N. Retrieved from https://dx.doi.org/10.2139/ssrn.911512

Deng, Q. (2008). *Volatility Dispersion Trading*. SSRN.

Hunter. (2004). *Dispersion*. Capital Structure Demolition LLC.

Karl Härdle, W., Klochkov, Y., Petukhina, A., & Zhivotovskiy, N. (2022). Robustifying Markowitz. Retrieved from https://arxiv.org/abs/2212.13996

Ledoit, O., & Wolf, M. (2003, November). Honey, I Shrunk the Sample Covariance Matrix. Retrieved from http://www.ledoit.net/honey.pdf

Leland McInes, J. H. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction . *ArXiv e-prints*.

Marcos Lopez de Prado, D. H. (2011). The Sharpe Ratio Efficient Frontier. *Journal of Risk*.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91. Retrieved 4 22, 2023, from http://www.jstor.org/stable/2975974

Marshall, C. M. (2009). *Dispersion trading: Empirical evidence from U.S. options markets*. Global Finance Journal.

Martin, R. A. (2021). PyPortfolioOpt: portfolio optimization in Python. *Journal of Open Source Software, 6*, 3066. doi:10.21105/joss.03066

McInnes, L. H. (2018). *Umap: Uniform manifold approximation and projection for dimension reduction*. arXiv preprint arXiv:1802.03426.

Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is 'Optimized' Optimal? *Financial Analysts Journal*, 31-42. Retrieved April 22, 2023, from http://www.jstor.org/stable/4479185

Mynbayeva, E., Lamb, J. D., & Zhao, Y. (2022). Why estimation alone causes Markowitz portfolio selection to fail and what we might do about it. *European Journal of Operational Research*, 694-707. Retrieved from https://www.sciencedirect.com/science/article/pii/S0377221721009863

Ortiz, R., Contreras, M., & Mellado, C. (2022). Improving the volatility of the optimal weights of the Markowitz model. *Economic Research-Ekonomska Istraživanja, 35*(1), 2836-2858. Retrieved April 23, 2023, from https://www.tandfonline.com/doi/full/10.1080/1331677X.2021.1981963

Pierpaolo Ferrari, G. P. (2019). *Dispersion trading: an empirical analysis on the S&P 100 options*. Investment Management and Financial Innovations.

World Economic Outlook: Countering the Cost-of-Living Crisis. (2022). International Monetary Fund.

Zivot, E. (2021, June 11). Introduction to Computational Finance and Financial Econometrics with R. Retrieved April 22, 2023, from https://bookdown.org/compfinezbook/introcompfinr/