Powerful Pairs Trading
Final Project: Group 6
MATH585
25 April 2023

By: Patrick Kuiper, Hao Wu, Sean Moushegian, Brian Lee, Rashaad Ratliff-Brown

Table of Contents

Abstract

The fundamental goal of this project is to create an investment strategy which leverages the theoretical profitability of a portfolio centered on effective risk mitigation.[1] We accomplish this by focusing on investing in distinct opportunities, which are very likely to precipitate high returns, but are temporally rare. We accomplish this by creating a fundamentally simple strategy which meets these goals. First, we conducted detailed research on highly cointegrated pairs, all with large market capitalization, diversified across several sectors with historical stability. We then set entry and exit thresholds via the Ornstein-Uhlenbeck Process which ensures that entry into the associated pairs trade is only executed when a highly unlikely level of divergence is demonstrated. This level of pair divergence, while unlikely, is highly likely to be followed by mean reversion. Our strategy remains unique in that it does not continually search for profitable investments via complex statistical techniques across a huge array of covariates, or make frequent trades. Rather it makes use of our hypothesis that markets will fundamentally behave as expected - and when they do not - there are very brief moments of distinct profitability as securities will likely revert to stability.[2] A simple strategy which leverages this fact will be the most successful.

Introduction

The Saint Petersburg paradox tells us that risks should be mitigated as downside losses require huge reverse moves to overcome due the geometric nature of investments[3] If downside losses are not mitigated this compounding effect of the investment causes significant decreased returns. With this consideration, a successful strategy should be designed incorporating risk mitigation intrinsically.

The primary contribution of this project is structuring a strategy which outperforms market indices, while providing lower variance in returns. This is executed via two simple practices, coupled together: an effective pairs trading strategy via the Ornstein-Uhlenbeck Process, complimented by a detailed search of diversified pairs.[4]

Strategies which are constantly scouring statistical relationships among covariates for opportunity in order to make decisions - by hour, minute, or more often - are likely to come with high variance. Beyond pure variance, these complex strategies are trained with enormous parameter sets, leading to overfitting and poor out of sample performance. Instead, when developing our strategy we considered the holistic

---

[1] See "Amor Fati." Mark Spitznagel, January 2019
[2] See "Why Do People Still Invest in Hedge Funds." Mark Spitznagel, January 2020
[3] See Bernoulli, Daniel. “Exposition of a New Theory on the Measurement of Risk.”
[4] See Rampertshammer, Stefan. “An Ornstein-Uhlenbeck Framework for Pairs Trading.”

statistical assumptions associated with the behavior of asset pricing techniques, along with the long run utility of our portfolio.

We decided to focus towards stability through an investment strategy which first provides consistent returns above market indices. This investment is achieved via a pairs trading strategy, which is profitable whether the market is in a bull or bear cycle, or in other words, this strategy is directionally indifferent with respect to the market. When profitable paris trades are not available, we hold our cash conservatively. This steady hold until profitability requires detailed research into two components:

1. We must identify **profitable and safe pairs** which exhibit steady state behavior in almost any market condition.

2. Pairs must be members of a **diversified basket** and subsequently we must employ an effective method for determining entry and exit thresholds for these pairs**.**

For the remainder of the paper, we will expand on our development of the first and second points outlined above, and then provide analysis of the resulting portfolio.

Background

Pairs Trading

In the market, there are many pairs of stocks that follow similar patterns: when one stock increases (decreases) in value, the other increases (decreases) in step. Intuitively, one might expect there to be many such pairs in common sectors: events and news related to the sector might have a common impact on many stocks, and stocks in a common sector may follow similar patterns.

When we identify a pair of stocks that exhibit the same general patterns, we may attempt to predict the price of one stock based upon the price of the other. Call one stock *P* and the other *Q*. Then, the price of *P, Q* at time *t* can be called P_t, Q_t. We can perform an ordinary-least squares (OLS) regression upon the observations of *log P, log Q* at different timepoints and define a line-of-best-fit with slope β and y-intercept α:

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

The residuals of this regression at time *t* are given by ϵ_t. By the definition of OLS, the residuals will have mean zero. An investor will confidently proceed with pairs trading if the residuals on a given pair are *mean-reverting*: when the residuals drift from zero in absolute value, they tend to return to zero. A pattern is defined between *P* and *Q*, and we can predict the price of *P* at any given time from the price of *Q*. If *P* exceeds this predicted value, we know that the residual is positive -- for now. As investors trade on pairs with mean-reverting residuals, we can expect that the residual will eventually tend to zero, and by extension, the actual price of *P* will eventually tend to its OLS-predicted price. This could happen in one of two ways -- either the actual price of *P* decreases in value, or the predicted price of *P* increases in value (corresponding to an increasing price of *Q*). In the described scenario, an investor following the pairs trading strategy would long stock *Q* (with market weight $\frac{\beta}{1+\beta}$) and simultaneously short stock *P* (with market weight $\frac{1}{1+\beta}$). When the two stocks converge in value, the investor will earn a profit.

The minimum (in absolute value) residual in which we enter a pairs position is called our *entry threshold*, and the residual at which we liquidate an existing pairs position to realize a profit is called our *exit threshold*. The residual at which we liquidate an existing pairs position to avoid downside risk is called a *stop-loss threshold*.

The investor has the advantage of directional indifference: if the price of *P* and *Q* go up by the same dollar amount, we don't care: the gains from our long on one stock will cancel the losses from the other (so long as the capital is split correctly between our positions on *P* and *Q*). The same is true if both stock prices decrease in value. The only bet we make is on the relative spread between *P* and *Q* and this is the only criterion of our profit.

Protective Put/Protective Call Options Strategy

Our initial theory development for this project incorporated a protective put / protective call option strategy. This strategy was later removed from our implementation due to lack of performance; however, we still believe its theoretical qualities warrant discussion.

If an investor believes that a certain stock will appreciate in value, he or she may take a long position in this stock. Suppose, however, that the investor is risk-averse and is concerned about losing money in the event that the stock price decreases in value. If an investor holds a long position on *N* shares of a security, he or she may purchase an out-of-the-money put option (a put option with a strike price below below the current market price of the underlying security) in such a quantity to purchase *N* shares of the underlying security (typically, this will be *N*/100 contracts). In the event that the underlying security decreases in value, there is a limit to the amount of money that the investor can lose -- once the security decreases in value below the protective put strike price, any further losses are effectively canceled by new profits from the put option contract.

Similarly, an investor may purchase a protective call option -- a call option with a strike price above the current underlying market price -- to protect against upwards movements in a short position on the underlying asset.

This strategy is analogous to insurance -- an agent pays a small fee (in this case, the cost of an OTM put/call option) in exchange for some protection against risk (in this case, downside on the underlying security).

Implied Volatility

As our initial strategy (which was later removed) incorporated options purchases, understanding and analyzing implied volatility - a key component of options pricing - was critical. Among American options, option prices tend to increase as underlying volatility increases -- if the underlying volatility goes up, it is more likely that at some point between the present and maturity, the underlying price will fall above the price of a call (or below the price of a put). One popular model for computing the price of an option is Black-Scholes, which takes (among other factors) current underlying price, strike price, volatility, and expiry as inputs and outputs a fair market price for the contract (note that Black-Scholes assumes that you exercise the option only on the expiry date, as in European options). We can reverse this process: by inputting current market price, expiry, strike price, and underlying price into Black-Scholes, we can retrieve as output a volatility. This is the volatility that investors believe the underlying will undergo, and it is referred to as implied volatility.

Our Final Strategy

Pair Selection Methodology

K-means clustering is a popular unsupervised machine learning algorithm used in various domains. It is particularly useful for algorithmic pairs trading, which involves identifying pairs of assets that have a similar price movement pattern, taking advantage of the price differences between them.

In pairs trading, two assets are chosen based on some predefined criteria, which in our project is their sector and market capitalization. These assets are expected to demonstrate cointegration, or move in a similar pattern. A deviation from this pattern is considered a potential trading opportunity. K-means clustering can help to identify such pairs by grouping assets with similar price movement patterns, specifically if we cluster only on price movements.

To use K-means clustering for pairs trading, historical price data for a set of assets is first collected and grouped according to sector. In this project we employed the most recent 200-days of returns as the historical price data. This data is then preprocessed employing principle component analysis (PCA) to reduce the dimensionality, within sectors of consideration, of each security inorder to perform clustering analysis in a manageable space. We compressed the 200-days of return data into two principal components. We then applied the K-means algorithm to the compressed data to group assets into clusters based on their price movement patterns.[5]

The number of clusters is chosen based on the elbow method or any other appropriate method for determining the optimal number of clusters. In this analysis we employed three clusters. Once the assets were grouped into one of the three clusters, pairs of assets within each cluster were chosen as potential trading pairs. These pairs are expected to have a high correlation in their price movements and hence, cointegration.

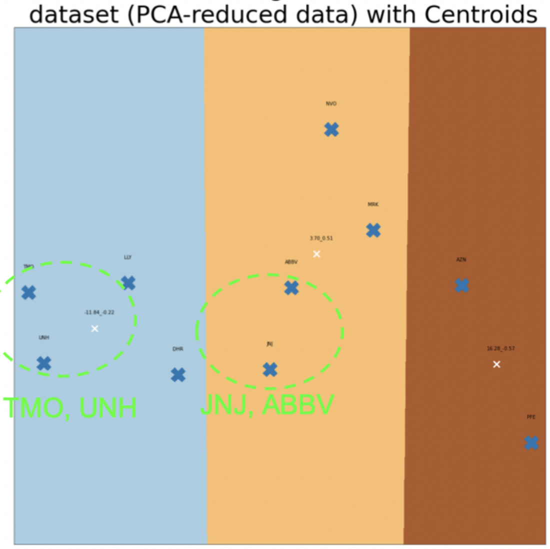
Below we have provided figures of each of the six sectors we employed the K-means clustering technique, and the associated pairs we selected based on three criteria:

1. Pair is in the same cluster (robust performance focused)
2. Pair is in the top three closest euclidean distance (robust performance focused)
3. Pair performs well with in sample performance of pairs trading

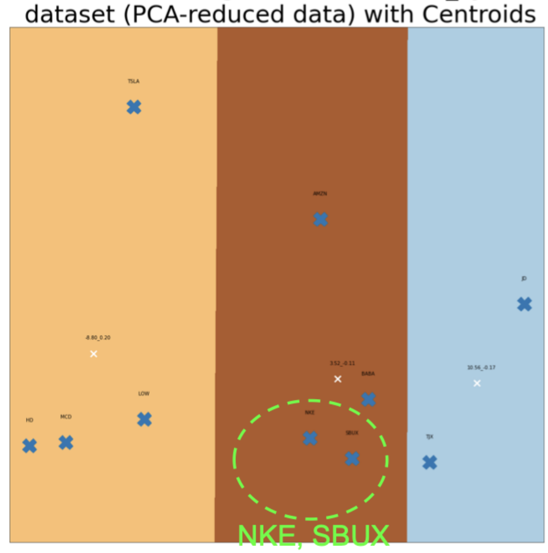
---

[5] See Trachevski, Matthew. (2019). Dynamic Pair Selection with Machine Learning Clustering, Cointegration Testing, a Stop-Loss Policy and Partial Pair Order Buffering.

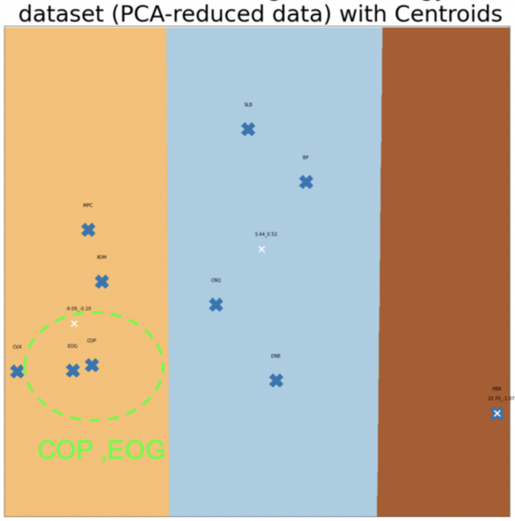
K-means clustering on the healthcare dataset (PCA-reduced data) with Centroids



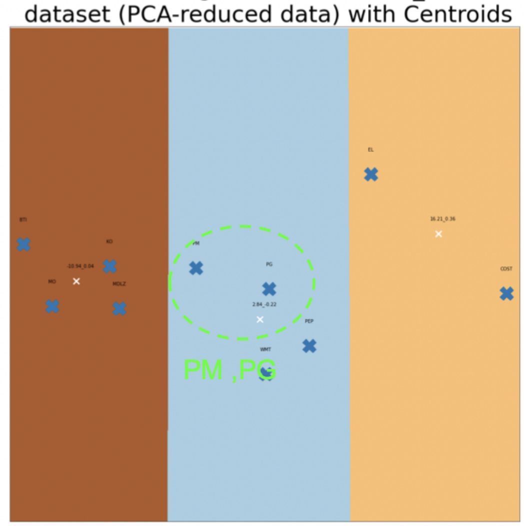
K-means clustering on the consumer_cyclical dataset (PCA-reduced data) with Centroids

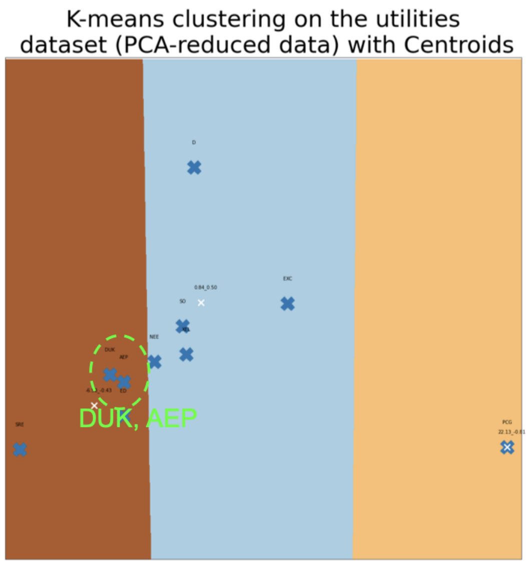


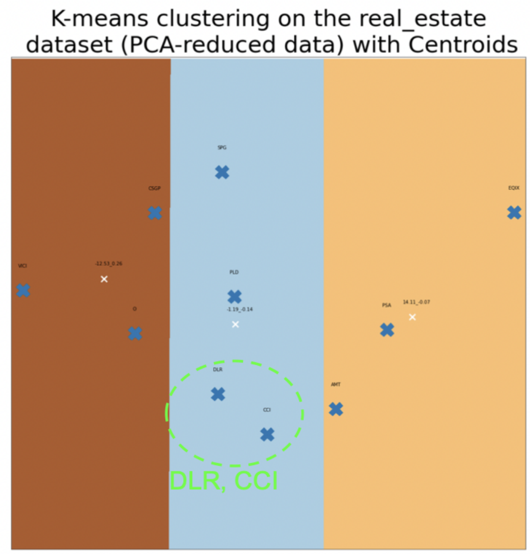
K-means clustering on the energy dataset (PCA-reduced data) with Centroids

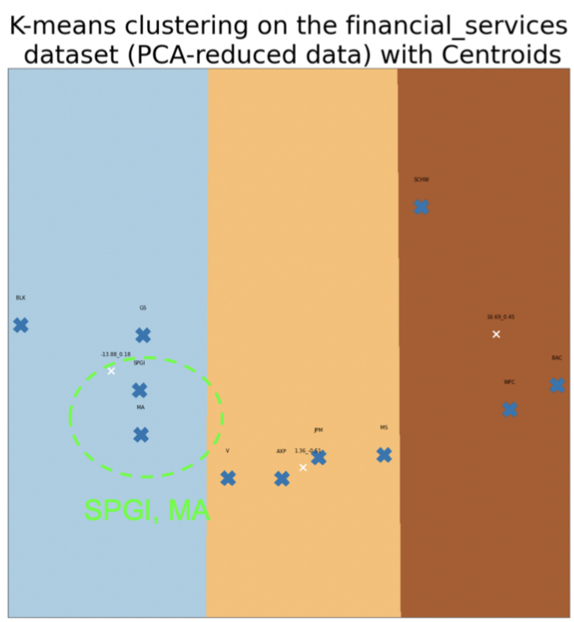


K-means clustering on the consumer_defensive dataset (PCA-reduced data) with Centroids









Additionally, we compared the k-means criteria to the ADF score to benchmark. We find that almost all of the selected pairs are below or close to the typical threshold used to determine ADF cointegration. Please see the table below for results:[6]

[6] See "Statistical Arbitrage by Pair Trading using Clustering and Machine Learning."

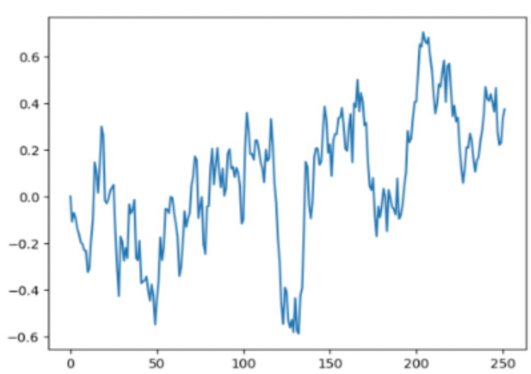
Pair	Industry	ADF Score	Euclidean Distance (2D-PCA)
(JNJ, ABBV)	Healthcare	0.766	1.70
(DUK, AEP)	Utilities	0.048	0.97
(NKE, SBUX)	Consumer Cyclical	0.161	2.15
(SPGI, MA)	Financial	0.293	0.34
(DLR, CCI)	Real Estate	0.059	4.26
(PM, PG)	Consumer Defensive	0.082	5.48
(TMO, UNH)	Healthcare	0.108	1.28
(COP, EOG)	Energy	0.002	1.46

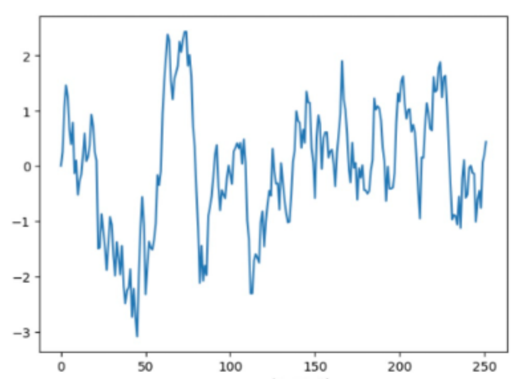
Ornstein-Uhlenbeck Processes (OU) for Entry and Exit Thresholds

Equations in the section are taken from Leung and Li and/or Wikipedia.

The OU process is a stationary gaussian Markov process, which models stochastic dynamics under the assumption of mean reversion. The process can be defined by the stochastic differential equation defined below, where both (θ, μ, σ) are the parameters, and Wt defines a Wiener process. In particular, θ is denoted as the long-term mean of the process, μ is denoted as the drifting power, and σ is the total volatility of the Markov motion.

$$dxt = \ - \mu(\theta - xt)dt + \ \sigma dWt$$
$$Wt = \ Wt - W0 \propto N(0, t)$$





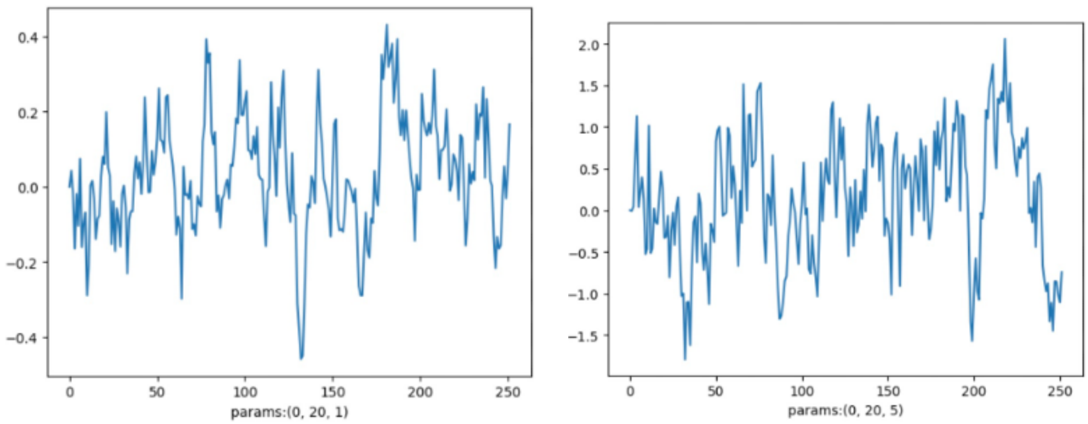


Figure.1 Simulation of OU process with different parameters

Above are four graphs generated by a single simulation where parameters are labeled in the title of the plots. Comparing the simulation where σ = 1 with the simulation where σ = 5, we can observe that the range of xt is wider for σ = 5 than that of σ = 1. Furthermore, comparing the simulation with μ = 10 to simulation with μ = 20, we can observe that the simulation with higher μ tend to converge to the mean with higher frequency. In conclusion, both of the σ and μ are important in the paris trading, since higher drifting power indicates higher probability and higher frequency that the z-score will converge to the mean. Furthermore, the value of σ could also determine when the strategy should enter and exit.[7]

In the normal pairs trading, we typically employ the constants of 2σ and -2σ for thresholds. However, not every pair follows similar patterns. In other words, it is possible that 2σ and -2σ will never be reached during the trading period, and result in unprofitable behavior for the portfolio. However, in our case, in order to find the suitable threshold for different pairs, we use a maximum likelihood estimator to determine the OU parameters (θ, μ, σ) of selected pairs using 100-day rolling windows, and employ Monte-Carlo simulations on the process which is generated by the estimated parameters. We then consider 5% and 95% percentile of the z-score as the exit and entry thresholds.

Step1 (Leung and Li, 2016):

$$f^{OU}\left(z_{i-1}; \theta, \mu, \sigma\right) = \frac{1}{\sqrt{2\pi\sigma^{*}}} \exp exp\left(\frac{-\left(z_i - z_{i-1}e^{-\mu\Delta t} - \theta\left(1-e^{-\mu\Delta t}\right)\right)^2}{2\sigma^{*^2}}\right) \text{(pdf of OU process)}$$

Where z_i is the z-score at t = i

$$\sigma^{*} = \sigma^2 \frac{1-e^{-2\mu\Delta t}}{2\mu}$$

Every single trading day, we obtain past 100 day's z-score , and then maximize the likelihood function:

[7] See Leung, T., & Li, X. (2015). Optimal mean reversion trading with transaction costs and stop-loss exit.

$$l\left(z_{1:100}\right) = -\frac{1}{2}\ln ln\,(2\pi) - \ln ln\left(\sigma^{*}\right) - \frac{1}{2*100}\Sigma\left(z_{i} - z_{i-1}e^{-\mu\Delta t} - \theta\left(1 - e^{-\mu\Delta t}\right)\right)^{2}$$

Step2:

After we have obtain the MLE estimator (θ, μ, σ), we assume the future z-score will follow the OU process statistically with the estimated parameters (θ, μ, σ), and then execute trades based on the expected 5% and 95% percentile of z-score as the trading signal, employing a Monte-Carlo simulation.

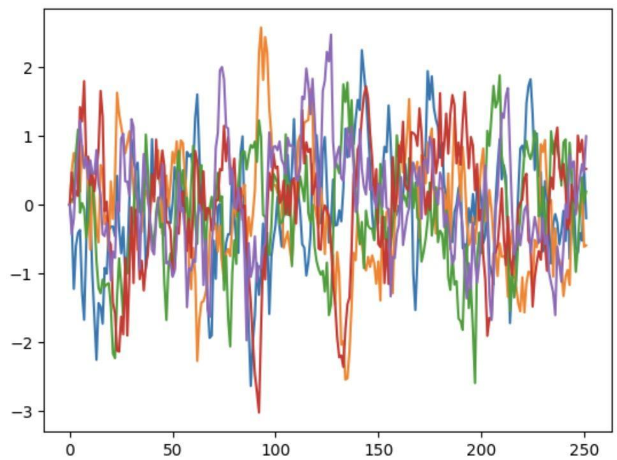


Figure.2 5 simulations with parameter (0,20,5)

Figure 2 shows 5 simulations of the OU process with parameters (0, 20, 5) with rolling windows of 252 trading days. In practice, the entry threshold will be determined by the expected value of 5% percentile and 95% percentile of the z score:

$$entry\ for\ long\ spread: \frac{E\left[\sum_{1}^{5} 5\%ile\ z_{i}\right]}{5}$$

$$entry\ for\ short\ spread: \frac{E\left[\sum_{1}^{5} 95\%ile\ z_{i}\right]}{5}$$

$$exit\ for\ both\ long\ and\ short\ spread: \frac{\frac{E\left[\sum_{1}^{5} 5\%ile\ z_{i}\right]}{5} + \mu_{i}}{4}$$

Limitations:

1. Notice that z is the standardized residual score which is calculated by the regression coefficient beta, however, during the back-testing period, the beta will be negative

during some period between 2017 and 2021. Negative beta will cause our strategy to long or short both stocks in the selected pairs.

2. There may be some periods when the regression beta is very small, and in that case, our pairs-trading returns are largely dependent on the performance of single stock.

Additional Risk Management with Tiingo

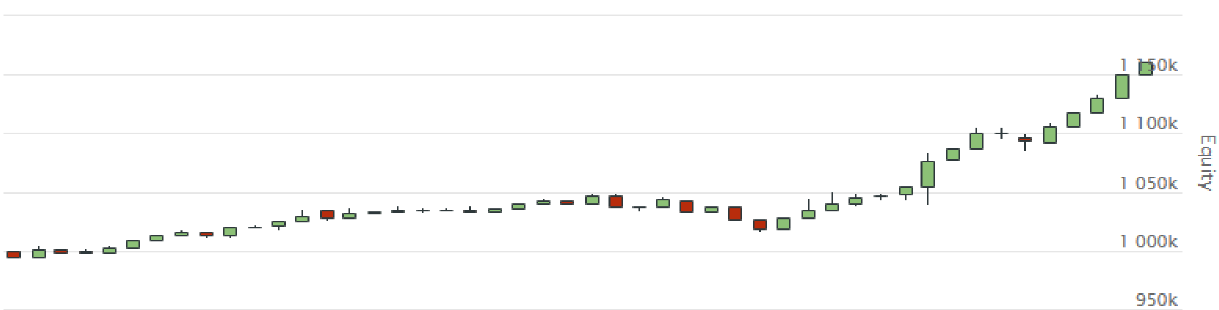
Further risk management is employed incorporating sentiment analysis using the Tiingo library as a risk-management measure. Whenever a pairs trade is placed, we measure the average sentiment score of the past fourteen days for each stock in the pair. Then, while we hold the position, we liquidate both pairs if their sentiment scores dip 3 standard deviations below the average that we calculated.

Performance on In Sample and Out of Sample Periods

Updated Strategy

Below, clearly denoted graphically and in tables for precision, is a breakdown of our strategy's performance in the specified in sample and out of sample testing periods. Backtest URLs are also provided for reference. Of note: Tiingo risk management is not included for these backtest results, and there is some stochasticity to the performance of this strategy. These points are discussed further after the plots.

In Sample (Jan 1 2017 - Jan 1 2021):



Backtest URL	https://www.quantconnect.com/terminal/processCache/?request=embedded_backtest_c4f745ddec625ff7b3a8c2fb9e6abdb6.html
Return	16.04%

Sharpe Ratio	1.137
Drawdown	3.1%

Out of Sample (Jan 1 2023 - Apr 1 2023):

Backtest URL	https://www.quantconnect.com/terminal/processCache/?request=embedded_backtest_878897c62dbca32abc758f41ca3d5f8a.html
Return	1.641%
Sharpe Ratio	2.576
Drawdown	0.7%

LiveTrade (Apr 18 2023 - Apr 24 2023) - Backtested:

Because our livetrade used an old strategy, we decided to backtest the livetrade time period using the newer strategy. It does not execute any trades during this time period.

Backtest URL	https://www.quantconnect.com/terminal/processCache/?request=embedded_backtest_5818572dc3b1cdb665d4180a37ec3a0a.html
Return	0
Sharpe Ratio	0

Drawdown	0

Discontinued Strategy

LiveTrade (Apr 18 2023 - Apr 24 2023) - Deployed Apr 18:

Below are the live-trade is in the Group6-Final-Livetrade project The following statistics were determined on Apr 25 2023 at approximately 10:30am:

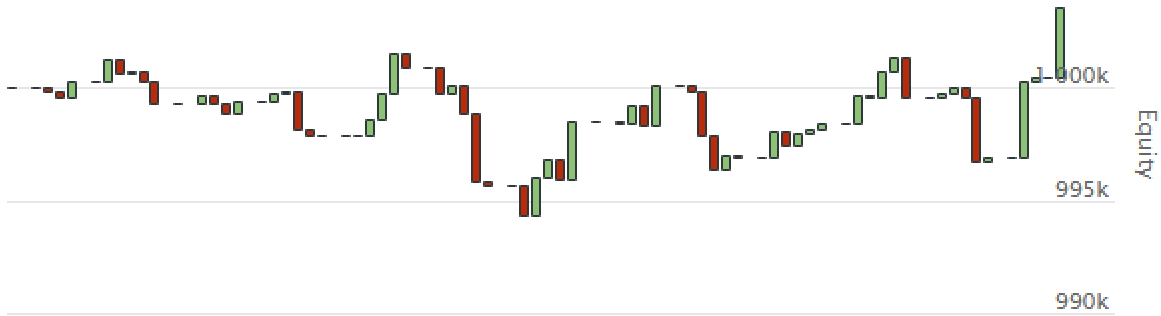
This execution follows a discontinued strategy (using the options insurance). We have backtested over this same period using the newer strategy (results are in the next section):

Return	-4.4%
Sharpe Ratio	Not provided by QC
Drawdown	Not provided by QC

Due to technical runtime limitations, the above results were generated without Tiingo sentiment integration. Even so, we have observed that Tiingo does not cause any pairs to liquidate in these time periods, which makes sense -- our drawdown is very limited, so risk management tools such as Tiingo would not have much reason to execute. The Tiingo component of our strategy proposal should not impact our performance over these periods.

Note: due to the fact that we are estimating quantities related to the OU fitting via Monte Carlo simulation, there is some stochasticity in the results that we attain. Thus, the statistics associated with our backtests will fall within some range if repeated. The results submitted are good, but not atypical, for our algorithm.

Comparison: Traditional Pairs Trading



Backtest URL	https://www.quantconnect.com/terminal/processCache/?request=embedded_backtest_43be5f2ef22782e2d08137d3bc98b2a9.html
Return	0.854%
Sharpe Ratio	1.38
Drawdown	0.7%

Our implementation features better return and Sharpe ratio over this period for this run (again, there exists stochasticity with the OU process). The advantages of OU are less clear for the in-sample period.

Conclusion from Backtests of Updated Strategy

Ultimately, employing our strategy provided a return of 16.04%, with a Sharpe ratio of 1.137, and drawdown of 3.1% for the in-sample testing period. Furthermore, our strategy provided a return of 1.164%, with a Sharpe ratio of 2.67, and drawdown of 0.67% for the out-of-sample testing period. These are very acceptable results and meet the goals of our fundamental strategy with good returns and low risk / drawdown. Additionally, our pairs trading strategy performs better than traditional pairs trading techniques.

We believe the profitable performance of our strategy is due to leveraging the OU modeling process (as demonstrated by our results provided above) and employing the PCA / K-means process for determining profitable pairs.

Discontinued Strategies

Combining Protective Puts/Calls with Pairs Trading

In traditional pairs trading, an investor will implement several thresholds: an entry threshold, which determine when a pairs trade position is entered, an exit threshold, which determines when sufficient profit (considering risk) has been generated and the pairs position is closed for a realized profit, and a stop-loss threshold, which determines at what point we liquidate the position and realize a limited loss to prevent further downside. From our preliminary analysis, we found that many pairs that diverge beyond pre-determined stop-loss thresholds would eventually converge to a profitable position, but the uncertainty and drawdown associated with temporary divergence mandates a stop-loss liquidation to avoid potentially significant losses.

We proposed a strategy to mitigate this problem in our midterm presentation. In the strategy, we define an entry and exit threshold, but do not define any stop-loss threshold. Instead, we purchase an out-of-the-money (OTM) put option on one underlying security of our pair, and an OTM call option on the other underlying security. Under the proposed strategy, the *protective call* and *protective put* option contracts are purchased in a quantity sufficient to buy or sell the entire position of the associated underlying security. The strike price of these OTM options would be purchased a far distance from the current market price of the underlying security, ideally so much so that the OTM options contracts would be inexpensive to acquire.

The ownership of the option contracts enable an investor to eliminate any stop-loss threshold without drawdown fears. In the event that the pair diverges in value, then the options would provide some protection against drawdown and losses. The following figure provides an example:

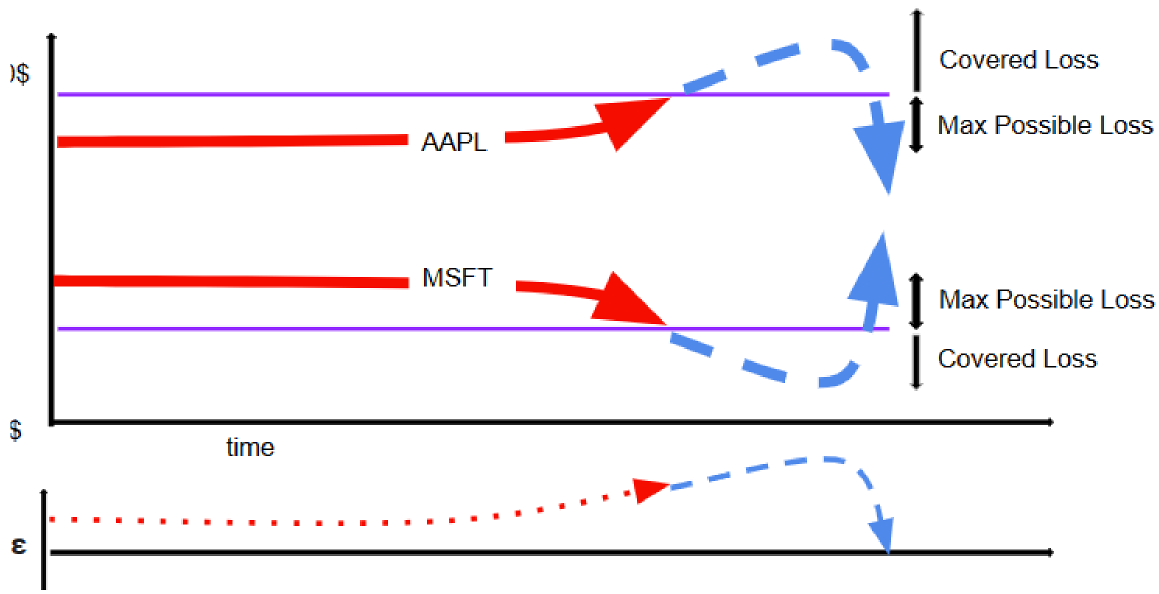


Figure: Options Insurance Protects Wayward Pair
In the above example, an investor bets that the log-prices of AAPL and MSFT converge in price. Instead of implementing a stop-loss order, an investor purchases a protective OTM call option on AAPL and a protective OTM put option on MSFT with strike prices illustrated by the purple lines. If the price of AAPL trends upwards, then losses are limited -- any price movement beyond the protective call strike price will be covered, and any downward MSFT price movement beyond the strike price of the protective put strike price are also covered. Then, an investor can leave this pairs position open without fear of drawdown and potentially profit on eventual convergence (dotted blue line).

In a traditional pairs-trading algorithm, an investor forms a hypothesis only with respect to the spread between the log-prices of the securities in the pair, not on the direction of the market. For example, following the previous example, if the entire technology section improves and the value of a share of MSFT and AAPL both go up by $100, the investor does not care -- the gains from the long position on MSFT cancel the losses from the short position on AAPL. The same is true of a sector downturn. In our implementation, an investor is no longer entirely indifferent to movements in the market as before. If both AAPL and MSFT increase in value by $100, then we will earn a profit on our long position of MSFT and a (possibly limited by the protective call) loss on our short position of AAPL. If the pair diverges in value, it is still possible for a profit to be earned: if both stocks significantly increase (decrease) in value, one profits from the long (short) position and suffers a clipped loss on the short (long) position and it is possible for the profit to exceed the loss. In traditional pairs trading, it is never possible for a trader to profit with a spread that grows in absolute value. The table below summarizes the profit/loss outcomes for different pairs behavior.

Table	
Outcome	Profit relative to traditional Pairs Trading
No Options Exercised- w/ or w/out Spread Shrinking in absolute value	Profits same, less cost of options
One Option Exercised- w/ Spread Shrinking	Profits likely greater
One Option Exercised - w/out Spread Shrinking	Possibly profitable (not possible traditionally)
Two Options Exercised	Losses reduced

Implementation

The success of this strategy heavily depends upon procuring inexpensive option contracts. If the log-prices of the securities of the pair converge in price, we would earn a profit in the traditional pairs-trading approach. In the proposed strategy, the price of the protective call/put options would reduce the total profits enjoyed by the investor. Thus, the proposed strategy is heavily dependent upon procuring inexpensive options. To determine an OTM strike price that will be inexpensive, we conducted the following approach: we retrieved the past M days of returns, randomly uniformly sampled these returns with replacement N times to estimate the security price N days into the future, and repeated this process S times. The S predictions of the future stock price describe a distribution of estimates for a future stock price. Of our S estimates, we would pick the Q quantile estimate and set this as the strike price of our protective call/put option. For example, if M=60, N=30, S=100, and Q=0.05, we would retrieve the past 60 days of MSFT returns, uniformly sample these returns 30 times to estimate the price 30 days into the future, repeat this process 99 times to produce 100 estimates of the 30-days-out price, and select the 5th-percentile such element as the strike price of our protective put. The same process would be repeated to find the 95-percentile future price of AAPL to determine the strike price of a protective call. The expiration date of our protective call/put would be set to N.

Limitations

In experimenting with this approach, we encountered several problems:

- **Limited Option Contract Information:** We are searching for contracts with a medium expiry date and a very OTM strike price. Indeed, these are not the most sought-after option contracts. The QuantConnect platform was not able to consistently produce option contracts that met all of our requirements, including a very OTM strike price.
- **Volatility Smile**: It is a well-documented fact that the implied volatility of contracts increases as the strike price moves further away from the current market price. As our strategy relies upon low option contract prices, we require extremely OTM option contracts. These contracts will have higher-than-typical implied volatilities, and by extension, will typically have higher prices. This curtails the ability to get cheap option contracts -- a necessary component of this strategy.
- **Low Liquidity:** Options that are very OTM are not bought or sold in high quantities. As a result, a market maker will typically charge a higher bid-ask spread, meaning that higher fees will be required to execute an acquisition of very OTM options. This also raises options contract costs -- undercutting profits necessary to sustain this strategy.

Discontinuation

For all of the reasons outlined in the previous section, we determined that purchasing protective calls and puts was a nonoptimal risk management strategy. The sparsity of option contracts information -- and the high prices of those contracts for which we have information -- make this strategy less optimal than a traditional pairs trade with a stop-loss threshold.

Future Improvements

Future Improvements to OU Modeling:

Equations in the section are taken from Leung and Li.

A possible improvement could be achieved by fully implementing the OU modeling methodology outlined in Leung and Li 2016 paper, which provides us with a method which no longer requires calculating the hedge ratio by the regression. In particular, consider the portfolio X, which contains \$1 of share A and \$-β of share B, and denote X_i to be the portfolio value at time i.[8]

[8] See Leung, T., & Li, X. (2015). Optimal mean reversion trading with transaction costs and stop-loss exit.

Step1: estimated $(\beta, \theta, \mu, \sigma)$

The hedge ratio β is estimated by $argmax_{\beta} l(\theta, \mu, \sigma | x_{1:N})$. Take the data of GLD and SLV from 2023-02-01 to 2023-04-01 as an example. Assume we invest \$1 in GLD and \$-β in SLV. The left graph shows the relationship between the log-likelihood and 100* β. In that case, a portfolio with \$1 in GLD and \$-0.58 in SLV will provide us with the highest likelihood.

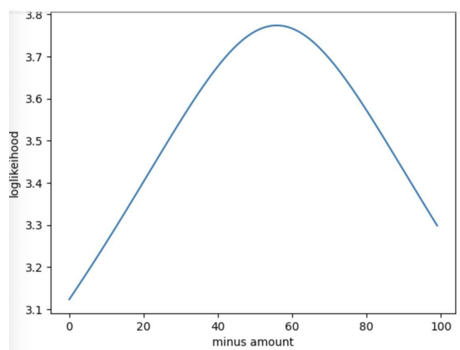


Figure3 log likelihood change with respect to \$-β amount

Step2: using differential equations to find $b(entry\ threshold)$ and $d(exit\ threshold)$

We may further refer to Li and Leung's double optimal stopping problem (Li & Leung, 2016), which would allow for the calculation of the optimal entry and exit threshold. In particular, with the estimated $(\beta, \theta, \mu, \sigma)$, we can first find b in order to make equation (3) holds. With that exit threshold b, we can then able to get the value of equation (4), and by plugging equation (4) into equation (5), we can obtain the entry level d.

$$F(x, r) = \int_{0}^{\infty} u^{\frac{r}{\mu} - 1} e^{\frac{-u^2}{2} + (x - \theta) u \sqrt{\frac{2\mu}{\sigma^2}}} \qquad (1)$$

$$G(x, r) = \int_{0}^{\infty} u^{\frac{r}{\mu} - 1} e^{\frac{-u^2}{2} - (x - \theta) u \sqrt{\frac{2\mu}{\sigma^2}}} \qquad (2)$$

$$F(b) = (b - c) * F'(b). \qquad (3)$$

$$V(d) = (b - c)\frac{F(x)}{F(b)} \quad if \quad d < b$$
$$or \quad d - c \quad if\ d > b \quad (4)$$

$$G(d)\left(V'(d) - 1\right) - G'(d)(V(d) - d - c \quad (5)$$

Below is the figure which shows the optimal entry and exit threshold of pairs that contains GLD and SLV. Based on the given time period, the optimal entry point is around .37, and the optimal exit threshold is around 0.475.

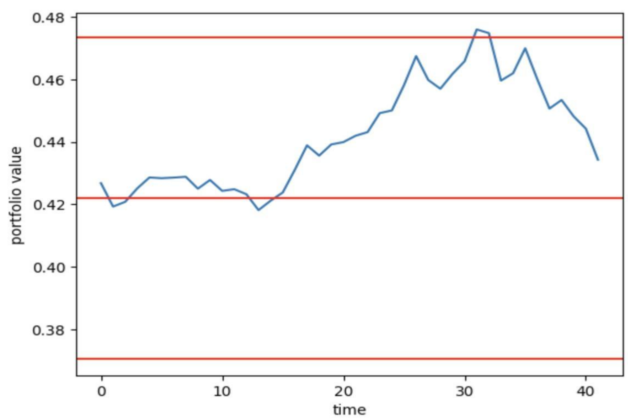


Figure.4 optimal entry and exit threshold for GLD and SLV from 2023-02-01 to 2023-04-01

Contributions

Patrick Kuiper: basic pairs model, initial protective options model and write up / presentation composition.

Sean Moushegain: protective options implementation and write up / presentation; background writing; alternative strategy ideation and exploration.

Brian Lee: risk management with Tiingo, incorporation separate parts of strategy to work together, write up / presentation

Hao Wu: OU methodology, future improvement of threshold selection and hedge ratio estimation

Rashaad Ratliff-Brown: Assisted with K-means algo for pairs trading and laid groundwork for clustering section write up

References

Bernoulli, Daniel. “Exposition of a New Theory on the Measurement of Risk.” *Econometrica*, vol. 22, no. 1, 1954, pp. 23–36. *JSTOR*, https://doi.org/10.2307/1909829. Accessed 24 Apr. 2023.

"Amor Fati." Mark Spitznagel*,* January 2019, https://www.universa.net/UniversaResearch_SafeHaven_AmorFati.pdf.

"Why Do People Still Invest in Hedge Funds." Mark Spitznagel*,* January 2020, https://www.universa.net/Universa_Spitznagel_SafeHaven_HedgeFunds.pdf.

Rampertshammer, Stefan. “An Ornstein-Uhlenbeck Framework for Pairs Trading.” (2007).

Rocco, Marco, Extreme Value Theory for Finance: A Survey (February 3, 2012). Bank of Italy Occasional Paper No. 99, Available at SSRN: https://ssrn.com/abstract=1998740 or http://dx.doi.org/10.2139/ssrn.1998740

Narcisa Kadlcakova, Lubos Komarek, Zlatuse Komarkova & Michal Hlavacek (2016) Identification of Asset Price Misalignments on Financial Markets With Extreme Value Theory, Emerging Markets Finance and Trade, 52:11,2595-2609, DOI: 10.1080/1540496X.2015.1087792

Leung, T., & Li, X. (2015). Optimal mean reversion trading with transaction costs and stop-loss exit. *International Journal of Theoretical and Applied Finance*, *18*(03), 1550020.

Trachevski, Matthew. (2019). Dynamic Pair Selection with Machine Learning Clustering, Cointegration Testing, a Stop-Loss Policy and Partial Pair Order Buffering. 10.13140/RG.2.2.35787.34081.

“Statistical Arbitrage by Pair Trading using Clustering and Machine Learning.” (2019).