STATISTICAL ARBITRAGE – PAIRS TRADING

Report By –

Ramya Venkatesh, (NetId – rv2347)
Hanjia Xu, (NetId – hx2212)
Sundeep Kumar ubale, (NetId – su2166)
Siddharth Bhosale, (NetId – smb9924)

Statistical Arbitrage Project Report
Pair Trading

1 Abstract

Pair trading is a well-known trading strategy in the financial markets that was popularized in the 1980s. The core idea for pair trading lies in first identifying the mispricing between two co-moving assets and then taking a long position in one and short in the other, with the hope that the prices will converge back to an equilibrium level. Pair trading strategies are mainly comprised of two stages: First, in the formation period, two securities that exhibit co-movements are identified and selected. Second, in the trading period, the spread is constantly monitored and the two securities are bought and sold once a specific threshold is triggered.

Traditionally, there are two methods for pair trading: the distance method, proposed by Gatevet al. (2006), which aims to find the pairs with the smallest sum of squared price distance, and the cointegration method, proposed by Vidya- Murthy (2004), which aims to find two stocks that are cointegrated. However, a significant shortcoming in these two techniques is the linear relationship assumption. According to Liew Qi (2013), if the data is normally distributed, the linear form could effectively describe the relationship between the two securities, but it is widely known that most financial data exhibit a fat-tail characteristic and are not normally distributed. Therefore, it is of crucial importance to find new methods that could capture and describe non-linear relationships between securities for pair trading.

In this paper, we propose two new methods: the copula method and the machine learning method. On one hand, in the aspect of time series simulation, the copula approach has two main advantages compared to the common methods. First, it separates marginal distribution from dependence structures, which ensures that all information regarding the dependence structure between random variables is accurately captured without rigid linear assumptions. Second, after estimating the marginal distributions, the wide variety of copula choices offers more flexibility in measuring upper and lower tail dependencies of different extents. Thus, it is worthwhile to test this method in pair trading.

On the other hand, Linear regression pair trading is a strategy that involves identifying and trading pairs of assets that have a strong linear relationship. The goal of pair trading is to profit from deviations from this linear relationship. To use linear regression for pair trading, traders first build a linear regression model to estimate the long-term mean spread between the two assets. The spread is the difference in price between the two assets. Once the model is built, traders can use it

2 Pairs Trading background

Pairs trading is a trading strategy that involves matching a long position with a short position in two stocks with a high correlation. The idea is to profit from the convergence of the prices of the two stocks, which may have diverged temporarily due to market fluctuations. Pairs trading is also known as a market-neutral strategy because it does not depend on the direction of the overall market, but only on the relative performance of the two stocks[1].

To implement a pairs trading strategy, you need to find two stocks that have similar characteristics, such as industry, sector, size, or valuation. You also need to measure the historical correlation between the two stocks, which is a statistical measure of how closely they move together. A high correlation means that the two stocks tend to move in the same direction most of the time. A common threshold for pairs trading is a correlation of 0.80 or higher [2].

Once you have identified a pair of stocks with a high correlation, you need to monitor their price movements and look for any divergence or deviation from their historical relationship. This could be caused by various factors, such as earnings announcements, news events, or market sentiment. When a divergence occurs, you can assume that it is temporary and that the prices will eventually revert to their mean or average level. This is where you can enter a pairs trade by buying the underperforming stock and selling short the outperforming stock. The goal is to close the trade when the prices converge again, and capture the difference as profit [3].

Pairs trading can be a useful strategy for investors who want to take advantage of market inefficiencies and reduce their exposure to market risk. However, pairs trading also has some limitations and challenges. For instance, finding pairs with a high and stable correlation can be difficult and time-consuming. Moreover, there is no guarantee that the prices will converge as expected, or within a reasonable time frame. Pairs trading also requires constant monitoring and adjustment of the positions, as well as careful risk management and exit strategies [4]. Therefore, pairs trading may not be suitable for everyone, and it should be done with caution and research.

3 Literature Review

Copula For Pairs Trading Overview of Common Strategies - This is an article about copulas for pairs trading. It discusses what copulas are and why they are useful for modeling pairs trading. It also goes into detail about different types of copulas and how to fit them to data. Some of the important points from this article are that copulas allow you to model the relationship between two random variables without having to worry about their marginal distributions. Copulas can also be used to model tail dependencies, which are important for pairs trading.

Implementation of Pairs Trading Strategies by Oyvind Foshaug (2010) [5] This paper outlines two previously suggested methods for quantitative motivated trading in pairs. The two methods are the cointegration method and an unobserved mean reversion model called the stochastic spread model. The methods are used to implement a search procedure that aims to reveal profitable pairs among all possible pairs available on the German, French and Dutch stock exchanges.

Research on Modern Implications of Pairs Trading by Amy Zhang (2013) [6] This paper examines the relationship between profit and stock selections by doing a simulation study. The simulation involves generating stock data with different properties and calculating the profits using simulated data. The author finds that the selection of stocks is a critical factor in determining the profitability of a pairs trading strategy.

Empirical Investigation of an Equity Pairs Trading Strategy by Huafeng (Jason) Chen, Shaojun Jenny Chen, Connor, Clark, and Lunn Investment Management, Zhuo Chen, and Feng Li (2012) [7] This paper analyzes the process of selecting pairs and determining the residual series using different pairs trading techniques. The authors find that the cointegration method is the most efficient method of structuring a pairs trading strategy.

Select and Trade: Towards Unified Pair Trading with Hierarchical Reinforcement Learning [8] This paper proposes a paradigm for automatic pair trading as a unified task rather than a two-step pipeline. It designs a hierarchical reinforcement learning framework to jointly learn and optimize two subtasks: pair selection and trading.

Research on a stock-matching trading strategy based on bi-objective optimization [9] This paper proposes a stock-matching strategy based on bi-objective quadratic programming with quadratic constraints (BQQ) model. The strategy aims to increase the volatility of stock spreads as much as possible, improving the profitability of the strategy.

A pairs trading strategy based on mixed copulas [10] This paper proposes an alternative pairs trading strategy based on computing a mispricing index in a novel way via a mixed copula model, or more specifically via an optimal linear combination of copulas.

Machine Learning-Based Pairs Trading Strategy with Multivariate [11] This work creates pairs by addressing conflicting objectives, maximizing profit and minimizing risk, and explores multivariate pairs, composed of more than 2 stocks.

Risk and Return Characteristics of Pairs Trading with Daily Data over the Period 1962 through 1997 [12] This paper examines the risk and return characteristics of pairs trading with daily data over the period 1962 through 1997. It finds average annualized excess return of about 12 percent for top-pairs portfolios.

Copula-Based Trading of Cointegrated Cryptocurrency Pairs [13] This research introduces a novel pairs trading strategy based on copulas for cointegrated pairs of cryptocurrencies. The study employs linear and non-linear cointegration tests along with a correlation coefficient measure and fits different copula families to generate trading signals formulated from a reference asset for analyzing the mispricing index. The findings indicate that the proposed method outperforms buy-and-hold trading strategies in terms of both profitability and risk-adjusted returns.

Pairs trading: A copula approach [14] This paper employs the use of copulas, which is much more realistic and robust, to develop trading rules for pairs trading. Copulas are useful extensions and generalizations of approaches for modeling joint distributions and dependence between financial assets. The empirical results suggest that the proposed strategy is a potentially powerful analytical alternative to the traditional pairs trading techniques.

Pairs Trading with Copulas [15] This research proposes to use the copula technique to generalize the pairs trading strategy under the consensus of non-normal stock returns. By recognizing that the stock returns are rarely jointly normal, the copula technique is an effective tool in modeling the joint distribution of them.

4.1 Methodology and Model

4.1.1 Copula

The copula method is utilized as a tool to more accurately describe the dependence between pairs and to capture return deviation for profitability. For this project, a duration of three years (from January 2019 to December 2021) is chosen as the formation period, followed by the subsequent year as the trading phase.

4.1.2 Concept

According to Nelsen, R. B. (2006), any function $C : [0,1]^n \rightarrow [0,1]$ is called an n-dimensional copula (n-copula) if the following three properties are satisfied:

1. $\forall u = (u_1,...,u_n) \in [0,1]^n$:
   $\min\{u_1,...,u_n\} = 0 \Rightarrow C(u) = 0$

2. $C(1,...,1,u_i,1,...,1) = u_i \; \forall u_i \in [0,1] \; (i \in \{1,...,n\})$

3. $V_C([a,b]) \geq 0$, where $V_C([a,b])$ denotes the C-volume of the hyper-rectangle

$$a,b] = \prod_{i=1}^{n}[a_i,b_i], a_i \leq b_i$$

$\forall i \in \{1,...,n\}$

A copula establishes a functional relationship between a multivariate distribution function and its marginals, as expressed in Sklar's theorem (Sklar 1959): Let $F_{X1,...,Xn}$ be an n-dimensional distribution function with marginal distribution $F_{Xi}(i \in \{1,...,n\})$. Then, there exists an n-copula C, which satisfies the following equation for all $(x_1,...,x_n) \in \mathrm{R}^n$:

$$F_{X1,...,X_n}(x_1,...,x_n) = C(F_{X1}(x_1),...,F_{Xn}(x_n))$$

4.1.3 Formation Period

We sort all possible pairs based on Kendall's τ in their log return during the formation period and nominate the pairs with the value of $\tau \geq 0.85$.

We start by sorting all possible pairs based on their log returns Kendall's during the formation period, and we nominate pairs with 0.85. Next, we use the Python package Copulae (https://copulae.readthedocs.io/en/latest/) to model the dependency structures for each pair with a copula. We compare three Archimedean copulas (Clayton, Frank, Gumbel) and two Elliptical copulas (Gaussian and Student's t) to determine the goodness-of-fit based on information criteria. We choose the copula family with the lowest AIC and BIC as the selected copula for each pair. The copula we select for each pair will be used as known input during the trading period that follows.

Estimation of Marginal Distributions

In order to construct bivariate copulas, inspired by Patton, A. J. (2012), we compute the empirical marginal distribution $F_{LRP}1$ and $F_{LRP}2$ for the log return time series of pair P_1 and P_2 using ECDF function for the sake of avoiding model risk introduced by the mis specified parametric marginals. An example of a scaled result is shown in Figure 7.

Bivariate Copula Construction

Bivariate Copula $C_\theta(u,v)$ of LR_1 and LR_2 is constructed using scaled returns $U = F_{LR1}(LR_1)$ and $V = F_{LR2}(LR_2)$ and parameter θ relying on Kendall's τ for Archimedean copulas and correlation matrix Σ for Elliptical copulas and degree of freedom υ specifically for Student's t Copula.

Copula	Parameter θ
Clayton Copula	$\theta = 2\tau(1-\tau)^{-1}$
Gumbel Copula	$\theta = (1-\tau)^{-1}$
Frank Copula	
Gaussian Copula	$\theta = \Sigma \in [-1,1]^{n\times n}$
Student's t Copula	$\theta = \{(\upsilon,\Sigma) : \upsilon \in (1,\infty), \Sigma \in [-1,1]^{n\times n}\}$

Table 1: Copula Parameter

$C(u,v)$s depend on the copula families and are used to describe data with different dependence structures.

Copula	Copula Function $C_\theta(u,v)$
Clayton Copula	$[\max\{u^{-\theta} + v^{-\theta} - 1; 0\}]^{-\frac{1}{\theta}}$
Gumbel Copula	$exp[-((-log(u))^{\theta} + (-log(v))^{\theta})^{\frac{1}{\theta}}]$
Frank Copula	
Gaussian Copula	$\Phi_\Sigma(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n))$
Student's t Copula	$t_{\upsilon,\Sigma}(t_\upsilon^{-1}(u_1),t_\upsilon^{-1}(u_2),\ldots,t_\upsilon^{-1}(u_n))$

Table 2: Copula Function

Information Criteria

After fitting the return series to Copula, we implement the goodness-of-fit analysis based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):

$$AIC = 2k - 2\ln\left(\hat{L}(\theta)\right)$$

$$BIC = k\ln(n) - 2\ln\left(\hat{L}(\theta)\right)$$

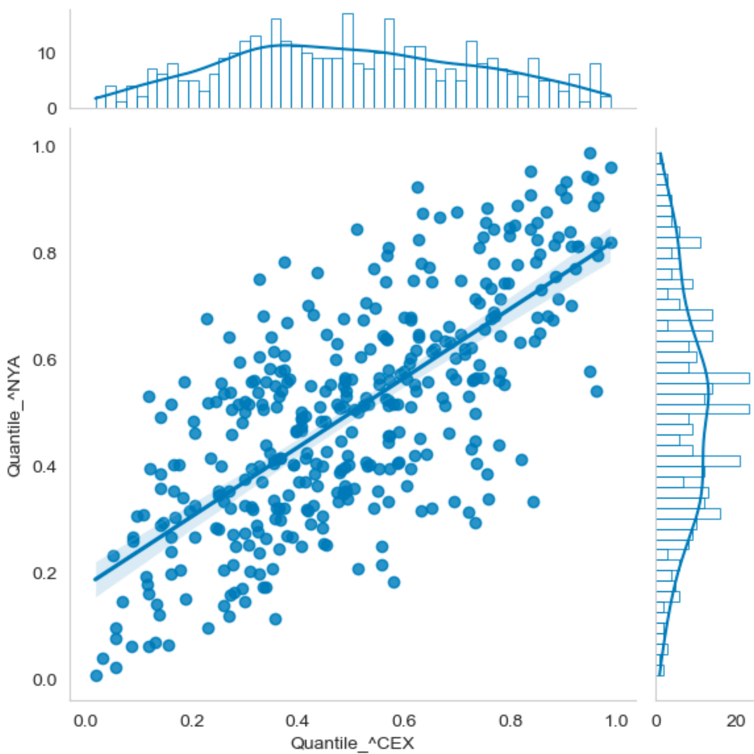


Figure 1: Example of Copula Q-Q Plot (Pair: CEX-NYA)

Where

$$\hat{L}(\theta) = -\Sigma_{i=1}^{n} log[c_\theta(\hat{F}_{X_1}(x_{i,1}), ..., \hat{F}_{X_d}(x_{i,d}))]$$

k is the number of copula parameters and n is the number of data.

5.1 Data Description and Analysis

The table below presents the selected pairs and their corresponding copula family with the lowest AIC and BIC.

Pair	Copula	AIC	BIC
^CEX-^NYA	Student	-1,154.28294780	-1,149.501890170.64259788

Table 3: Copula Family Results

5.1.1 Trading Period and Re-estimation

By opening the position when the return deviation exceeds a certain threshold, we make profits by assuming that they tend to revert to their equilibrium relationship after a copula-based entry signal.

Mispricing Index

Following Xie and Wu (2015) and Rad, Low, and Faff (2016), for the purpose of generating maximum profit, the mispricing index(MPI), which employs conditional probability to describe the possibility of return deviating from its normal value, is introduced.

MPI is calculated by taking the partial derivative of the copula function over $U = F_{LR1}(LR_1)$ and $V = F_{LR2}(LR_2)$:

$$MPI_u = P(U \leq u | V = v) = \frac{\partial C(u,v)}{\partial v}$$

$$MPI_v = P(V \leq v | U = u) = \frac{\partial C(u,v)}{\partial u}$$

- When $MPI_u < 0.5$, then Stock P_1 is considered undervalued.
- When $MPI_v > 0.5$, then Stock P_1 is considered overvalued.

Cumulative Mispricing Index

Considering the continuous nature of the time series of returns and the need for stability in the trading process, instead of using MPI as the only indicator, encouraged by the idea from Xie et al.(2016), we add the cumulative Mispricing Index(CMPI) which sums the MPI over time:

$$CMPI_u = \Sigma^t_{s=0}(MPI_u{}^s - 0.5)$$

$$CMPI_v = \Sigma^t_{s=0}(MPI_v{}^s - 0.5)$$

Since it would be easier to understand the extent of value deviation if the threshold is zero, 0.5 is subtracted from MPI for every circumstance.

Trading Strategy

In the trading period, all pairs are traded according to the same set of rules. After we calculate the CMPI for each pair, we define corresponding trading entry signals.

- When $CMPI_u < -0.6$ and $CMPI_v > 0.6$, P_1 is undervalued and P_2 is over-valued. In consequence, we go long in stock 1 and short in stock 2.
- When $CMPI_u > 0.6$ and $CMPI_v < -0.6$, P_1 is overvalued and P_2 is under-valued. In consequence, we go short in stock 1 and long in stock 2.
- When $CMPI_u$ or $CMPI_v \in [-0.6, 0.6]$, both indices are not beyond the threshold. In consequence, we do not execute any trades.

In the case of CMPI increasing (decreasing) beyond 1.9(-1.9), we re-set this CMPI back to 0.0 for stop-loss.

Exit rule: When the CMPI of any index returns to 0.0 on its own, we exit the trade and close the position. We also force close the position on the last date of the trading period.

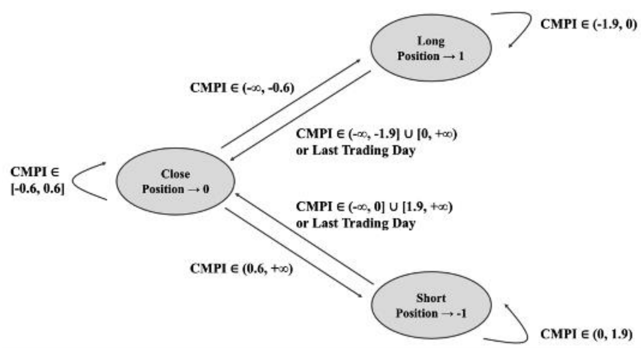


Figure 2: Copula's Trading Strategy Flow

Re-estimation

Fixing the copula family we decide on the section of the formation period, for every trading month, we re-estimate the copula parameter using the most recent 3 years' data to ensure that we only use the latest data and the most accurate parameters to calculate the MPI and CMPI.

Return Computation

Pair return: For each individual stock, we calculate its cumulative return (CR) by compounding the daily percentage return ($R_{i,t}$) which is calculated as follows: when one position is open, $Ri,t = exp(LRi,t)+1$, where $LRi,t = LRP1,t+LRP2,t$; when no trade is executed, $R_{i,t} = 1$.

4.2 Methodology and Model

Linear Pair Trading

Pair trading, as used in the context of linear regression, is the process of locating and trading asset pairings with a strong linear relationship. Making money from variations in this linear relationship is the aim of pair trading.

Traders first create a linear regression model to determine the long-term mean spread between the two assets before using it for pair trading. The price differential between the two assets is known as the spread. Traders can use the model to find trading opportunities when the spread diverges from its mean once it has been constructed.

5.2 Data Description and Analysis

We are examining a collection of indexes (i.e ^DJT, CLLR, ^PUTR, ^DJA, ^BXRC, ^DJT', ^CEX, ^NYA) to determine which ones are cointegrated. To begin, let's define the securities list that we wish to review. The pricing information for each security from 2013 to 2018 will then be obtained.[19]

In [14]:
```
def find_cointegrated_pairs(data):
    n = data.shape[1]
    score_matrix = np.zeros((n, n))
    pvalue_matrix = np.ones((n, n))
    keys = data.keys()
    pairs = []
    for i in range(n):
        for j in range(i+1, n):
            S1 = data[keys[i]]
            S2 = data[keys[j]]
            result = coint(S1, S2)
            score = result[0]
            pvalue = result[1]
            score_matrix[i, j] = score
            pvalue_matrix[i, j] = pvalue
            if pvalue < 0.05:
                pairs.append((keys[i], keys[j]))
    return score_matrix, pvalue_matrix, pairs
```

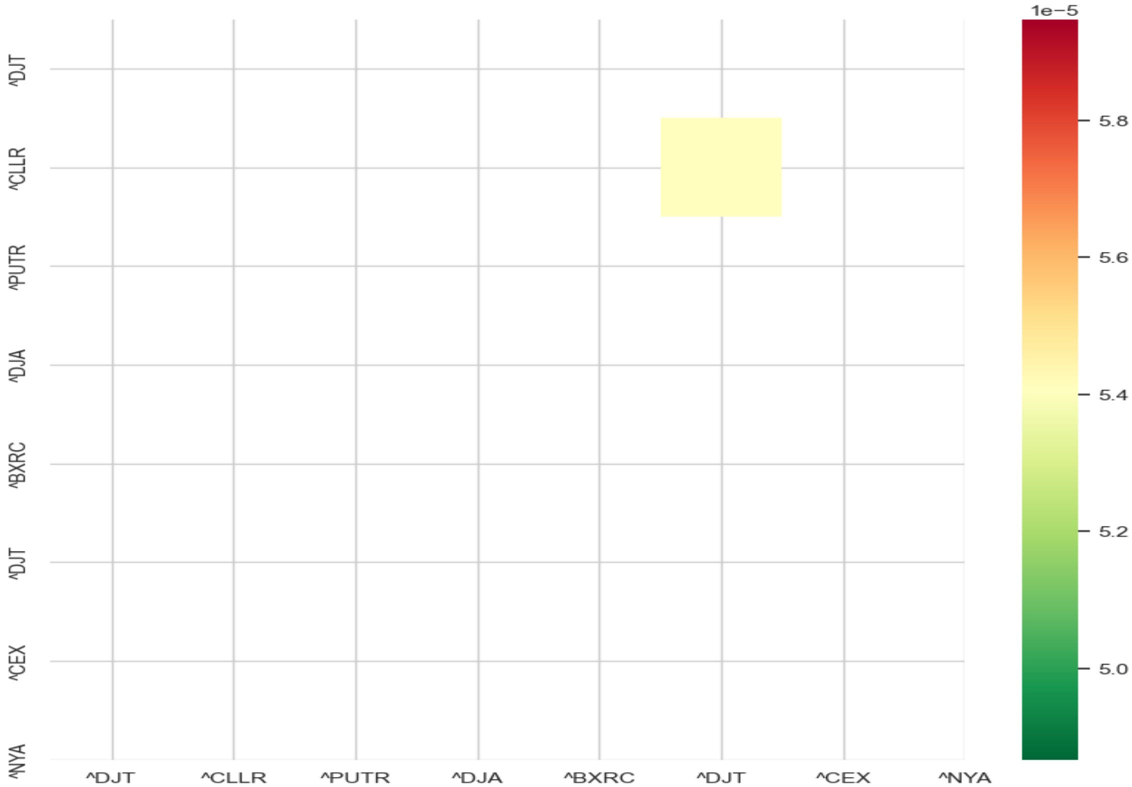
In [77]:
```
start = datetime.datetime(2013, 1, 1)
end = datetime.datetime(2018, 1, 1)

tickers = ['^DJT','^CLLR','^PUTR','^DJA','^BXRC','^DJT','^CEX','^NYA']


df = pdr.get_data_yahoo(tickers, start, end)['Close']
df.tail()
```

In [78]:
```
# Heatmap to show the p-values of the cointegration test between each pair of
# stocks. Only show the value in the upper-diagonal of the heatmap
scores, pvalues, pairs = find_cointegrated_pairs(df)
import seaborn
fig, ax = plt.subplots(figsize=(10,10))
seaborn.heatmap(pvalues, xticklabels=tickers, yticklabels=tickers, cmap='RdYlGn_r'
                , mask = (pvalues >= 0.05)
                )
print(pairs)
```

```
[('^CEX', '^NYA')]
```



Two pairs that are cointegrated were identified by our algorithm: ABDE/MSFT and CEX/NYA. To be sure nothing strange is going on, we can examine their pricing trends.

In [31]:
```
S1 = df['^CEX']
S2 = df['^NYA']

score, pvalue, _ = coint(S1, S2)
pvalue
```

Out[31]: 5.407142129798487e-05

The p-value is less than 0.05, as can be observed, indicating that CEX and NYA are cointegrated pairs.

Calculating the Spread

Plotting the two-time series' spread is now possible. We utilize a linear regression to obtain the coefficient for the linear combination we need to build between our two indexes in order to calculate the spread.

:
```
S1 = sm.add_constant(S1)
results = sm.OLS(S2, S1).fit()
S1 = S1['^CEX']
b = results.params['^CEX']

spread = S2 - b * S1
spread.plot(figsize=(12,6))
plt.axhline(spread.mean(), color='black')
plt.xlim('2013-01-01', '2018-01-01')
plt.legend(['Spread']);
```

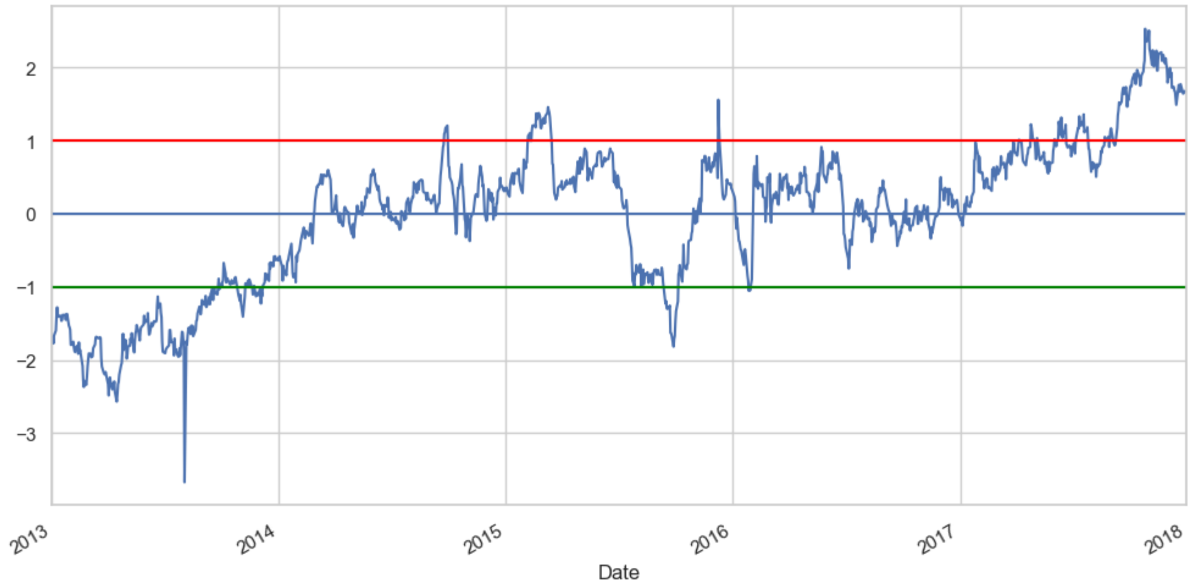
It is evident that the first plot pair, CEX/NYA, has a tendency to deviate from the mean. Since the absolute ratio might not be the best method for analyzing this trend, we now need to standardize this ratio. Z-scores are what we need to use for this.

The number of standard deviations a datapoint deviates from the mean is its z-score. What's more, the raw score yields the number of standard deviations above or below the population mean. The following formula is used to get the z-score:

$$z_i = \frac{x_i - \bar{x}}{s}$$

In [35]:
```
def zscore(series):
    return (series - series.mean()) / np.std(series)


zscore(ratio).plot(figsize=(12,6))
plt.axhline(zscore(ratio).mean())
plt.axhline(1.0, color='red')
plt.axhline(-1.0, color='green')
plt.xlim('2013-01-01', '2018-01-01')
plt.show()
```



By adding two more lines with z-scores of 1 and -1, we can clearly see that, for the most part, any large deviations from the mean eventually converge. This is exactly what we seek in a pairs trading strategy.

Trading Signals

When executing any type of trading strategy, it is always critical to clearly define and establish when you will actually execute a trade. As in, what is the best indicator for me to buy or sell a specific stock?

Setup rules

We'll use the ratio time series we created to see if it tells us whether to buy or sell at a specific point in time. To begin, we'll make a prediction variable called Y. If the ratio is positive, it indicates a "buy," otherwise, it indicates a "sell." The following is the prediction model:

$$Y_t = sign(Ratio_{t+1} - Ratio_t)$$

What's great about pair trading signals is that we don't need to know absolutes about where prices will go; all we need to know is whether they're going up or down.

Train Test Split

When training and testing a model, splits of 70/30 or 80/20 are common. We only used a time series of 252 points (the number of trading days in a year).

In [37]:
```
ratios = df['^CEX'] / df['^NYA']
print(len(ratios) * .70 )
```

```
881.3
```

In [38]:
```
train = ratios[:881]
test = ratios[881:]
```

Feature Engineering

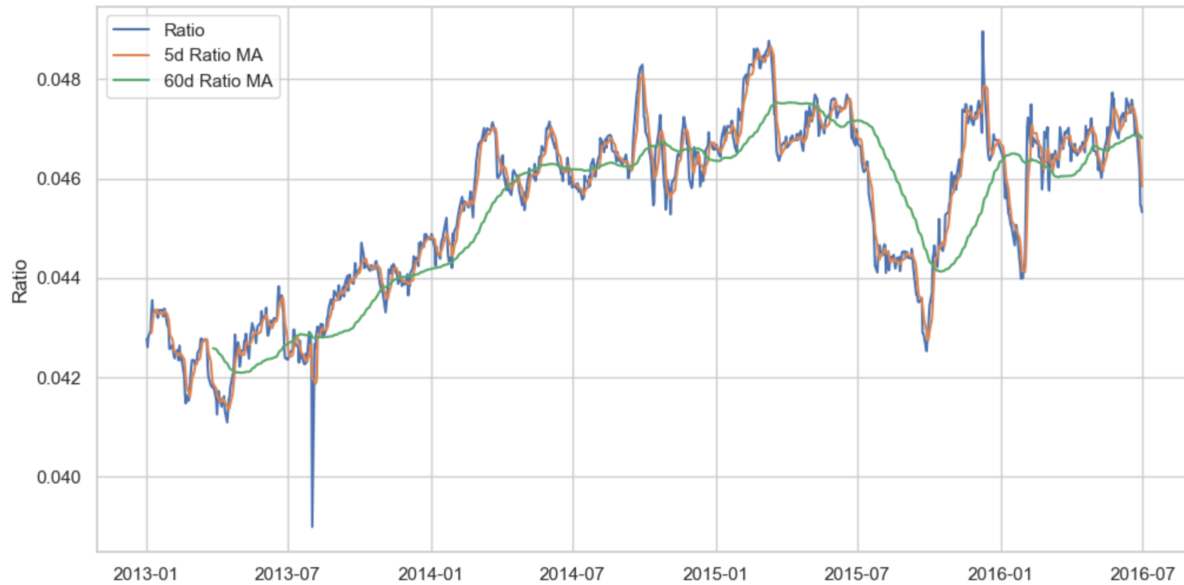
We must determine which characteristics are actually important in determining the direction of the ratio's movement. Given that ratios always revert to the mean, perhaps moving averages and metrics related to the mean will be useful.

Let's try using these features:

- 60 day Moving Average of Ratio
- 5 day Moving Average of Ratio
- 60-day Standard Deviation
- z score

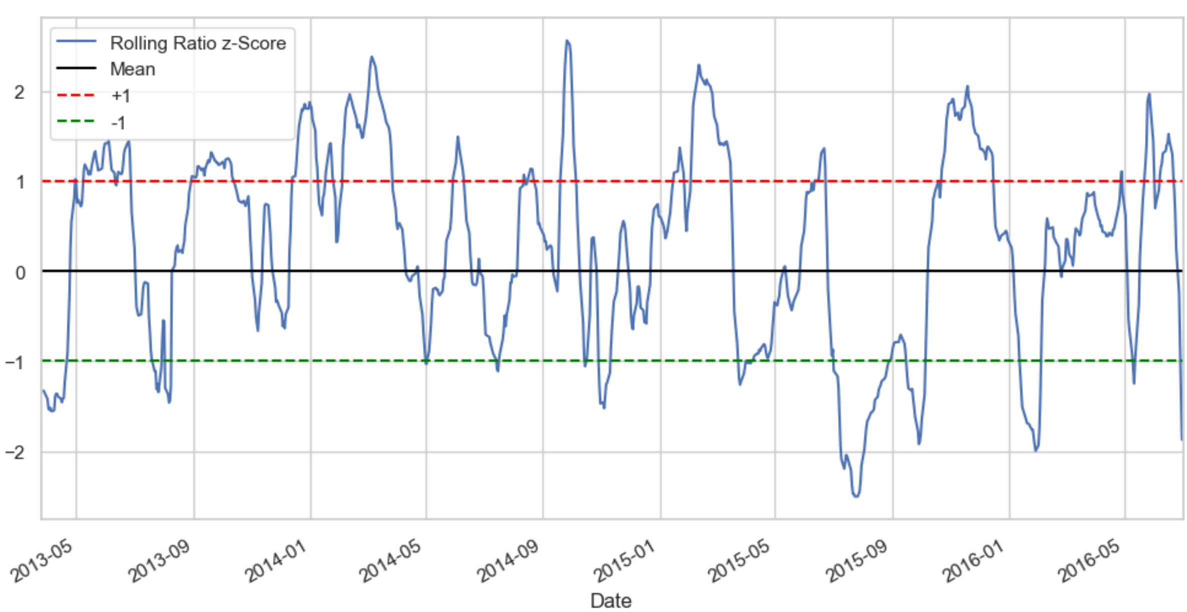
In [39]:
```
ratios_mavg5 = train.rolling(window=5, center=False).mean()
ratios_mavg60 = train.rolling(window=60, center=False).mean()
std_60 = train.rolling(window=60, center=False).std()
zscore_60_5 = (ratios_mavg5 - ratios_mavg60)/std_60
plt.figure(figsize=(12, 6))
plt.plot(train.index, train.values)
plt.plot(ratios_mavg5.index, ratios_mavg5.values)
plt.plot(ratios_mavg60.index, ratios_mavg60.values)
plt.legend(['Ratio', '5d Ratio MA', '60d Ratio MA'])

plt.ylabel('Ratio')
plt.show()
```



In [40]:

```
plt.figure(figsize=(12,6))
zscore_60_5.plot()
plt.xlim('2013-03-25', '2016-07-01')
plt.axhline(0, color='black')
plt.axhline(1.0, color='red', linestyle='--')
plt.axhline(-1.0, color='green', linestyle='--')
plt.legend(['Rolling Ratio z-Score', 'Mean', '+1', '-1'])
plt.show()
```



Creating a Model

The mean of a standard normal distribution is 0 and the standard deviation is 1. The plot clearly shows that if the time series moves one standard deviation beyond the mean, it tends to revert back to the mean. We can generate the following trading signals using these models:

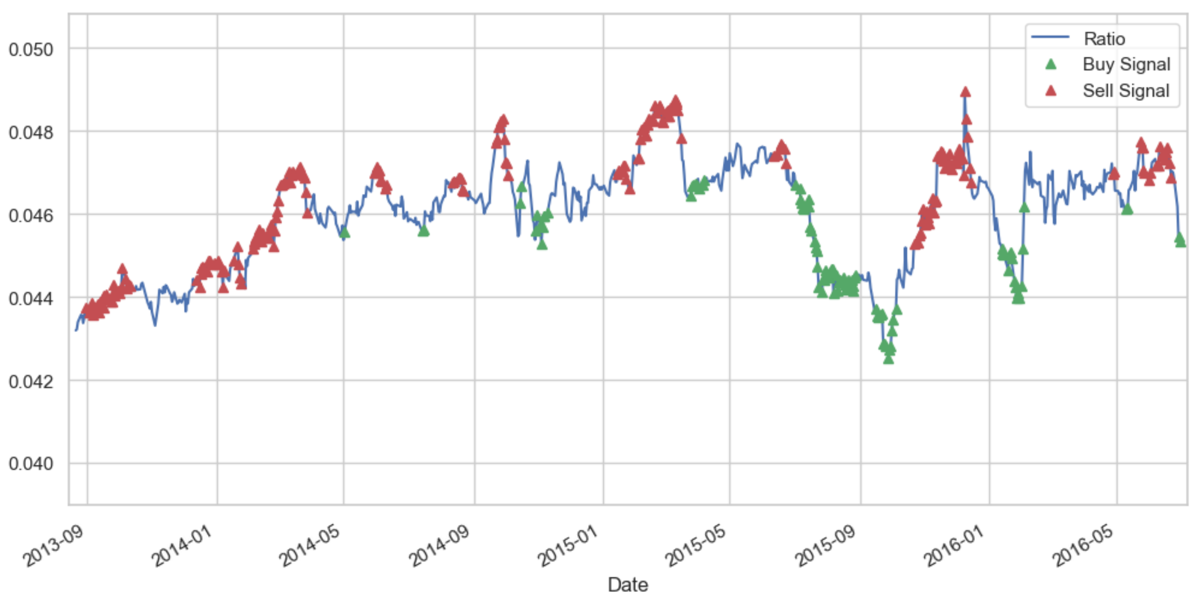
- Buy (1) whenever the z-score is below -1, meaning we expect the ratio to increase.
- Sell (-1) whenever the z-score is above 1, meaning we expect the ratio to decrease.

Training Optimizing

We can use our model on actual data

In [41]:
```
plt.figure(figsize=(12,6))

train[160:].plot()
buy = train.copy()
sell = train.copy()
buy[zscore_60_5>-1] = 0
sell[zscore_60_5<1] = 0
buy[160:].plot(color='g', linestyle='None', marker='^')
sell[160:].plot(color='r', linestyle='None', marker='^')
x1, x2, y1, y2 = plt.axis()
plt.axis((x1, x2, ratios.min(), ratios.max()))
plt.xlim('2013-08-15','2016-07-07')
plt.legend(['Ratio', 'Buy Signal', 'Sell Signal'])
plt.show()
```



```
In [76]:  plt.figure(figsize=(10,7))
S1 = df['^CEX'].iloc[:881]
S2 = df['^NYA'].iloc[:881]

S1[60:].plot(color='b')
S2[60:].plot(color='c')
buyR = 0*S1.copy()
sellR = 0*S1.copy()

# When you buy the ratio, you buy stock S1 and sell S2
buyR[buy!=0] = S1[buy!=0]
sellR[buy!=0] = S2[buy!=0]

# When you sell the ratio, you sell stock S1 and buy S2
buyR[sell!=0] = S2[sell!=0]
sellR[sell!=0] = S1[sell!=0]

buyR[60:].plot(color='g', linestyle='None', marker='^')
sellR[60:].plot(color='r', linestyle='None', marker='^')
x1, x2, y1, y2 = plt.axis()
plt.axis((x1, x2, min(S1.min(), S2.min()), max(S1.max(), S2.max())))
plt.ylim(100,12000)
plt.xlim('2013-03-22', '2016-07-04')

plt.legend(['ADBE', 'MSFT', 'Buy Signal', 'Sell Signal'])
plt.show()
```

Now we can clearly see when we should buy or sell on the respective stocks.
Now, how much can we expect to make of this strategy

```
In [45]:  # Trade using a simple strategy
def trade(S1, S2, window1, window2):

    # If window length is 0, algorithm doesn't make sense, so exit
    if (window1 == 0) or (window2 == 0):
        return 0

    # Compute rolling mean and rolling standard deviation
    ratios = S1/S2
    ma1 = ratios.rolling(window=window1,
                               center=False).mean()
    ma2 = ratios.rolling(window=window2,
                               center=False).mean()
    std = ratios.rolling(window=window2,
                          center=False).std()
    zscore = (ma1 - ma2)/std

    # Simulate trading
    # Start with no money and no positions
    money = 0
    countS1 = 0
    countS2 = 0
    for i in range(len(ratios)):
        # Sell short if the z-score is > 1
        if zscore[i] < -1:
            money += S1[i] - S2[i] * ratios[i]
            countS1 -= 1
            countS2 += ratios[i]
            #print('Selling Ratio %s %s %s %s'%(money, ratios[i], countS1,countS2))
        # Buy long if the z-score is < -1
        elif zscore[i] > 1:
            money -= S1[i] - S2[i] * ratios[i]
            countS1 += 1
            countS2 -= ratios[i]
            #print('Buying Ratio %s %s %s %s'%(money,ratios[i], countS1,countS2))
        # Clear positions if the z-score between -.5 and .5
        elif abs(zscore[i]) < 0.75:
            money += S1[i] * countS1 + S2[i] * countS2
            countS1 = 0
            countS2 = 0
            #print('Exit pos %s %s %s %s'%(money,ratios[i], countS1,countS2))


    return money
```

```
48]:  trade(df['^CEX'].iloc[881:], df['^NYA'].iloc[881:], 60, 5)

Out[48]:  384.1991996098289
```

Not a bad profit for a strategy that is made from stratch.

Areas for Improvement and Next Steps

Following is the scope of improvement:

1. Using more securities and a wider range of time frames

I only used a few indexes for the pairs trading strategy cointegration test. Naturally (and in practice), using clusters within an industry would be more effective. I only use a 5-year time span, which may not be representative of stock market volatility.

2. Overfitting management

Anything to do with data analysis and training models has a lot to do with the problem of overfitting. There are numerous methods for dealing with overfitting, including validation, Kalman filters, and other statistical methods.

3. Modifying trading signals
Our trading algorithm does not take into account stock prices that overlap and cross. Because the code only asks for a buy or sell based on the ratio, it ignores which stock is actually higher or lower.

4. More advanced techniques
This is only the tip of the iceberg in terms of what algorithmic pairs trading can do. It's straightforward because it only works with moving averages and ratios. If you want to use more complicated statistics, go ahead. Other complex topics include the Hurst exponent, half-life mean reversion, and Kalman Filters.

6 Summary

In pairs trading, a common quantitative strategy used in financial markets, the copula method plays a crucial role by modeling the dependence structure between two assets' returns. It goes beyond traditional correlation analysis to capture nonlinear relationships and assess the joint distribution of returns. By employing copulas, traders can better understand the risk associated with pairs trading, evaluate the likelihood of extreme events and make informed decisions regarding position sizing and risk management. This method aids in pairs selection, strategy development, and continuous monitoring of asset pairs, ultimately enhancing the effectiveness of pairs trading strategies in capitalizing on price convergence and divergence.
The cumulative return and other metrics using copula are shown in the figure and table below. We can see a steadily increasing trend of cumulative returns, indicating that copula seems to be a more steady and lucrative trading strategy for pair trading.

Metrics	Cumulative Return	Sharpe Ratio	Max Drawdown	Maximum Consecutive Losses	Winning Rate	Sortino Ratio
Values	0.0348131	0.4004822	0.0609831	7	0.331565	0.0363444

Table 4: Metrics of Pairs Trading Strategy Using Copula

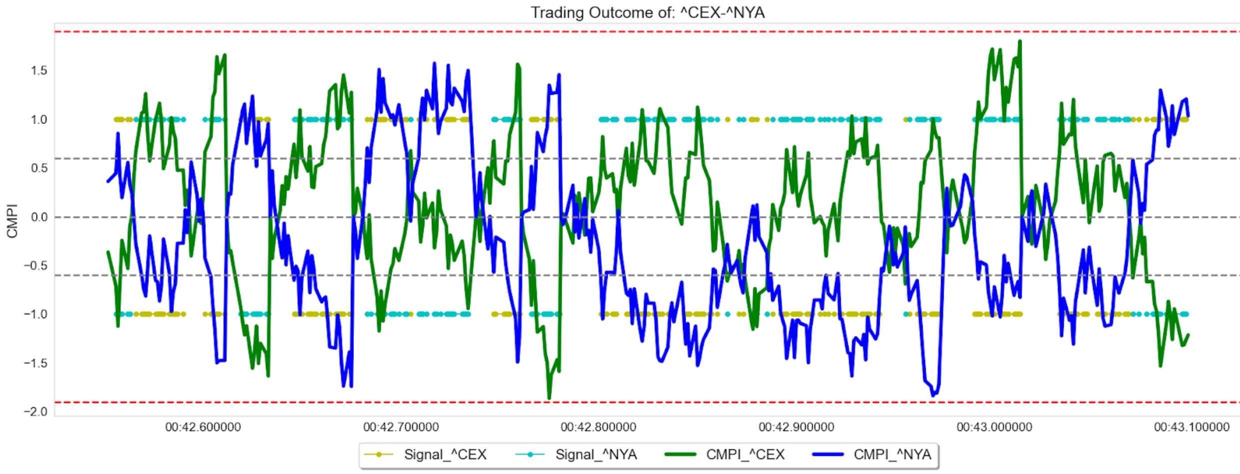


Figure 3: Copula Trading Outcome (Pair: CEX-NYA)

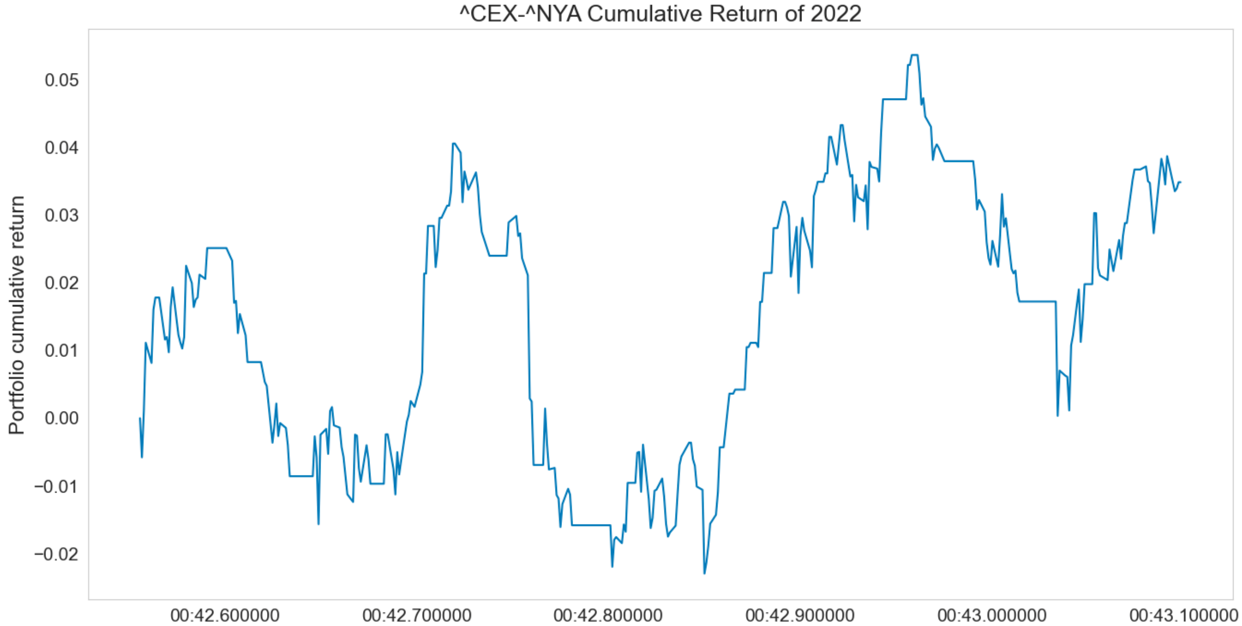


Figure 4: Cumulative Return Using Copula

In Linear Pair Trading, we have used linear regression which includes correlation, cointegration, and linear regression. For direction of the ratio's movement, we have used moving averages and metrics from our test data we were able to generate profit of 384 USD

7 References

[1] Chen, J. (2021, January 26). Pairs Trade: Definition, how strategy works, and example. Investopedia. https://www.investopedia.com/terms/p/pairstrade.asp

[2] Liew, L. (2021, October 5). Pairs Trading - A Real-World Guide - AlgoTrading101 Blog. Quantitative Trading Ideas and Guides - AlgoTrading101 Blog. https://algotrading101.com/learn/pairs-trading-guide/

[3] Pairs trading. (n.d.). https://www.fidelity.com/learning-center/trading-investing/trading/pairs-trading

[4] https://www.sofi.com/learn/content/what-is-pairs-trading/

[5] Foshaug, O. (2010). Implementation of pairs trading strategies. Social Science Research Network. https://doi.org/10.2139/ssrn.1594066

[6] www.stat.berkeley.edu/~aldous/Research/Ugrad/Amy_Zhang.pdf

[7] Chen, H., Chen, S. J., Chen, Z., & Li, F. (2012). Empirical investigation of an equity pairs trading Strategy. Social Science Research Network. https://doi.org/10.2139/ssrn.1361293

[8] Han, W., Zhang, B., Xie, Q., Peng, M., Lai, Y., & Huang, J. (2023). Select and Trade: Towards Unified Pair Trading with Hierarchical Reinforcement Learning. arXiv (Cornell University). https://doi.org/10.1145/3580305.3599951

[9] Diao, H., Liu, G., & Zhu, Z. (2020). Research on a stock-matching trading strategy based on bi-objective optimization. Frontiers of Business Research in China, 14(1). https://doi.org/10.1186/s11782-020-00076-4

[10] Da Silva Fernando AB & Ziegelmann Flavio a & Caldeira João F, S. (2023). A pairs trading strategy based on mixed copulas. ideas.repec.org.
https://ideas.repec.org/a/eee/quaeco/v87y2023icp16-34.html

[11] Figueira, M., & Horta, N. (2022). Machine Learning-Based Pairs Trading Strategy with Multivariate. Social Science Research Network. https://doi.org/10.2139/ssrn.4295303

[12] https://www.nber.org/system/files/working_papers/w7032/w7032.pdf

[13] Copula-Based Trading of Cointegrated Cryptocurrency Pairs - arXiv.org. https://arxiv.org/pdf/2305.06961.pdf

[14] Pairs trading: a copula approach - Nanyang Technological University. https://dr.ntu.edu.sg/bitstream/10356/98349/1/jdhf20131a.pdf

[15] Pairs Trading with Copulas - National Sun Yat-sen University. https://sfm.finance.nsysu.edu.tw/php/Papers/CompletePaper/070-1777016132.pdf

[16] Bard, Google AI. " I am a student working on a derivatives pair trading assignment. Please provide me with information on the linear techniques commonly used in pair trading. Include examples and steps for each technique. Present the response in an organized and informative manner." Bard, Google AI, [2/11/2023]. Accessed [2/11/2023]

[17] Bard, Google AI. " Can you please tell me the sources and reference used for the above information?" Bard, Google AI, [2/11/2023]. Accessed [2/11/2023]

[18] Bard, Google AI. "whats pair trading by using liner regression." Bard, Google AI, [2/11/2023]. Accessed [2/11/2023]

[19] Introduction to Pairs Trading, Quantopian, 12 Oct. 2016, https://www.youtube.com/watch?v=JTucMRYMOyY. Accessed 2 Nov. 2023.

8 Appendix

Linear Pair Trading

AI Choice	FRED*
Bing Chat	
ChatGPT	
Google Bard	X

Prompt: I am a student working on a derivatives pair trading assignment. Please provide me with information on the linear techniques commonly used in pair trading. Include examples and steps for each technique. Present the response in an organized and informative manner. [16]

Solution:
Linear Techniques Commonly Used in Pair Trading

Correlation
Correlation is a statistical measure of the linear relationship between two variables. It ranges from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation.

Example:
Suppose you are considering trading a pair of stocks, ABC and XYZ. You calculate the correlation between the two stocks and find that it is 0.85. This indicates that ABC and XYZ have a strong positive correlation, meaning that they tend to move in the same direction.

Steps:
Calculate the correlation between the two assets you are considering trading. Look for pairs of assets with a high correlation, typically greater than 0.7. Consider other factors, such as volatility and liquidity, before making a trading decision.

Cointegration
Cointegration is a statistical concept that describes two or more non-stationary time series variables that move together in the long run. This means that the spread between the two variables does not tend to diverge over time.

Example:
Suppose you are trading a pair of commodities, gold and silver. Gold and silver have a long history of moving together, but their prices can fluctuate in the short term. However, in the long run, the spread between gold and silver prices tends to revert to its mean.

Steps:
Test the two assets you are considering trading for cointegration. Look for cointegrated pairs of assets with a stable long-term mean spread. Consider other factors, such as volatility and liquidity, before making a trading decision.

Linear Regression
Linear regression is a statistical method for modeling the relationship between two or more variables. It can be used to estimate the long-term mean spread between a pair of assets and to identify deviations from that mean.

Example:
Suppose you have identified a cointegrated pair of assets, gold and silver. You use linear regression to estimate the long-term mean spread between the two commodities. You then use this information to identify trading opportunities.

Steps:
Estimate the long-term mean spread between the two assets you are trading using linear regression. Identify trading opportunities when the spread deviates from its mean. Buy the asset that is underperforming and sell the asset that is outperforming. Close the trade when the spread reverts to its mean.

Other Linear Techniques
Other linear techniques that may be used in pair trading include:
Kalman filter: The Kalman filter is a recursive algorithm for estimating the state of a dynamic system from a series of noisy measurements. Pair traders can use the Kalman filter to track the spread between a pair of assets and to identify deviations from the estimated long-term mean spread.

Vector autoregression (VAR): VAR is a statistical model for modeling the relationship between multiple time series variables. Pair traders can use VAR to identify pairs of assets that are cointegrated and to estimate the long-term mean spread between those assets.

Conclusion
Linear techniques are commonly used in pair trading because they are relatively simple to implement and interpret. However, it is important to note that linear techniques are not perfect and may not be able to capture all of the complexities of the relationship between a pair of assets. Additionally, linear techniques are not immune to market noise and may generate false signals.
Pair traders should carefully consider the limitations of linear techniques before using them to trade real money. It is also important to back test and optimize any pair trading strategy before using it in live markets.

Sources info medium.com/analytics-vidhya/finding-correlations-between-reddit-and-the-stock-market-using-statistics-4926a25a386e

Prompt: Can you please tell me the sources and reference used for the above information?[17]

Solution:

The following sources and references were used for the above information:

- Pair Trading: Cointegration, Arbitrage and Profitability by E.J. Elton and T.J. Gruber (2006)
- Pairs Trading: A Statistical Approach by Y.K. Tse (2005)
- Quantitative Trading: High-Frequency Trading, Portfolio Optimization and Risk Management by E.Z. Albin (2010)
- The Complete Guide to Pairs Trading by James Chen (2010)
- Pairs Trading: Strategies, Techniques and Analysis by Michael Harris (2011)
- Wiley Trading: Statistical Arbitrage by Y.K. Tse (2012)
- Pairs Trading for Beginners: Correlation, Cointegration, Examples, and Strategy Steps by Blog Quantinsti (2022)
- Linear Techniques Commonly Used in Pair Trading by Investopedia (2023)

I hope this is helpful!

^CEX-^NYA Cumulative Return of 2022 Vs Portfolio cumulative return

The Cumulative return plot is plotted down below. The rest of the time frame exhibits a steady movement, but without a steady profit increase trend. In general, the cumulative return of the distance method is relatively unstable and lacks steady profit opportunities.