International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Cointegration-Based Pairs Trading Strategy in Commodity Markets: A Novel Approach to Statistical Arbitrage

Gautham Mohandas

Charles River Associates, Chicago, United States of America

Abstract: ***Purpose: Pairs trading has been a popular strategy used by most hedge funds in exploiting the anomalies in the market. It is often used for trading overvalued and undervalued stocks, but can this method be extended to identify cointegrated pairs in commodities If so, how can we identify these pairs? This paper predominantly focuses on ways to find the cointegrated pairs in commodities and how to make an effective trade incorporating this strategy. Design/methodology/approach: The paper uses the commodities from the Goldman Sachs Commodity Index from the period 2007 to 2018 as a training dataset and from 2019 to 2022 as a test dataset. A rolling regression window of 1 day and 30 days has been implemented to overcome the issue of look-ahead bias for the trading pairs. Following that, a trading strategy has been created, longing the spread when the z-score is below -2, shorting the spread when the z-score is above 2, exiting the position when the z-score nears zero, and a stop-loss is created when the spread crosses positive or negative 4. Findings: 3 cointegrated pairs have been found from the commodity index which has a cointegrating vector and is statistically significant. A successful trading strategy has been established for those pairs. Originality/value: The paper primarily focuses on the commodities index. First, it is a unique statistical approach to finding the cointegrated pairs within the index. Second, a new trading approach has been proposed based on the z-score as when to long, short, and fix stop-loss of the pairs. Third, the paper also checks the cointegrated pairs by having a rolling regression window which helps differentiate whether the pairs have lost cointegration and eliminates look-ahead bias.***

Keywords: Hedge Funds, Spreads, Regression, Econometric approach

1. Introduction

Pairs trading is a popular statistical trading strategy employed by hedge funds in financial markets, particularly in stocks and futures markets. It is also called statistical arbitrage as traders or investors seek to exploit the relative price movements between two stocks or commodities [1]. The idea behind the trading strategy is to identify any two assets that exhibited a high degree of correlation in price movements that will tend to behave similarly in the future.

The main goal of this strategy is to bet that if the prices of two assets deviate, they will eventually converge back to a long-run equilibrium. This starts with applying methods to identify pairs and then establishing criteria for opening and closing positions. It is taking advantage of temporary mispricing of stocks that often show an economic link. A pairs trading strategy can be considered a mean reversion strategy and involves betting that the price of an asset will revert to the mean. A simple example of mean reversion is buying an asset after its price has fallen dramatically. Pairs trading can also be considered a market-neutral strategy since it matches a long position and a short position in strongly correlated financial securities. In a market-neutral strategy, you hedge against the market risk by simultaneously entering a long position in one stock and a short position in another to mitigate some market risk [2]. Therefore, it should offer profitable trading opportunities in any market environment. Overall, it exploits short-term common asset trends in multivariate time series utilizing different tools such as cointegration.

2. Literature Review

2.1 The Cointegration Method: Framework

Cointegration can be thought of as a relationship between two or more time-related series, or put another way, a stationary equilibrium relationship. Such a long-run relationship is different from mere linear interdependence or correlation. Cointegrated series exhibit a common long-term equilibrium with any short-term deviations from the mean being corrected over time. This correction can usually be represented by an error-correction model (ECM) [3]. While it is often rather difficult to work with non-stationary time series, in this case, it is preferred. When working with non-stationary time series, though, we must be aware of spurious regression problems where completely unrelated time series might appear to be related. This differs from a cointegrated relationship where the time trend has been removed and a genuine relationship between variables exists. By removing the common stochastic (time) trend and modeling the linear combination, we can exploit their relationship which in turn suggests that you can trade the spread profitably based on mean reversion. Stock prices are said to behave like random walks, also called an I(1) series. Cointegration tells you whether some linear combination (I(0)) of the stocks can be created such that the resulting return stream is stationary. This would mean that the pair moves together in lockstep. Such linear combinations are also referred to as cointegrating vectors, meaning that they share a common trend. These cointegrating vectors indicate how the time series are combined. Oftentimes, these are based on some economic link of the underlying securities.

2.2 Mean Reversion: Error Correction Model

The Error Correction Model describes how the dependent and independent variables behave in the short run consistent with a long-run cointegrating relationship. This model allows long-run components of variables to obey equilibrium constraints while short-run components have a flexible dynamic specification. For the short-run model, the variables need to be in stationary form. The Error Correction Model also incorporates the error corrections term, which are the residuals of the long-run regression but lagged one period. This term estimates how much of the disequilibrium will dissipate in the next forecasting period.

3. Data

The data has been taken from Yahoo Finance [4] from 2007 to 2021 for the Global Sachs Commodity Index (S&P GSCI). The reason behind choosing GSCI is because it is a widely followed commodity price index that was created by Goldman Sachs in 1991 [5]. It was designed to provide a broad and representative measure of commodity price movements. The index reflects the performance of a diversified basket of commodities, including energy, metals, agriculture, and livestock. The GSCI is often used as a benchmark for commodity investments and is used by traders, investors, and financial professionals to track the performance of the commodity market as a whole.The GSCI index is composed of a fixed number of commodities, and the composition is periodically adjusted to reflect changes in the commodity markets. The index includes commodities in different sectors; hence this paper primarily focuses on GSCI and the commodities listed in the index.

4. Methodology

There are different approaches to finding cointegrated pairs. This paper begins by examining individual commodities from the Goldman Sachs Commodity Index. One approach is more intuitive, selecting pairs where we expect cointegrating properties. This would be based on establishing an underlying theory of a shared trend. However, this does not guarantee that the variables cointegrate. The other approach is to test for cointegration using statistical tests. Assuming only one cointegrating vector, the two-step Engle-Granger method tests the residuals of the regression for stationarity. This can be done through any unit root test such as the augmented Dickey-Fuller test. The unit root is a characteristic of a non-stationary time series. The null hypothesis of the Dickey-Fuller test shows that such unit root is present and therefore the time series is non-stationary. Following the Engle-Granger method [6] will yield the coefficient for the linear combination between the pairs. Python has been chosen as the main language for testing the strategy in this paper. We used a statistical data visualization library called Seaborn to test the commodity pairs for cointegration. It displays the p-values of the tests between each pair in a heatmap. After having determined which commodity pairs are cointegrated, we determine their spread based on which we would execute trades.

4.1 In-Sample (07/30/2007-01/01/2019)

All the squares on our heatmap show cointegration, but the darker the green, the stronger the significance of cointegration between a pair of commodities

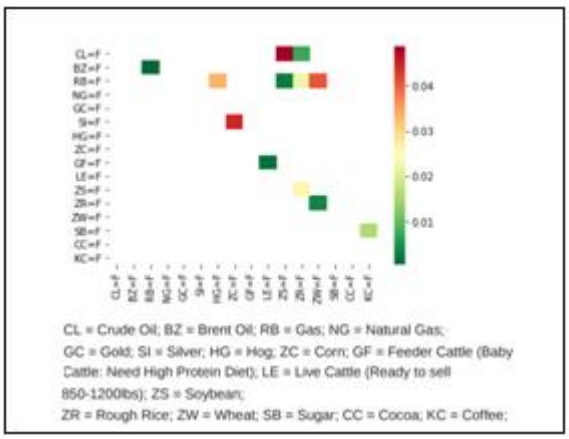

Figure 1: Cointegration Chart

Some strongly cointegrated pairs in Figure 1 do not have any economic link or significance. To move forward, we selected 3 cointegrated pairs that have strong economic links:

1) Baby Cattle (GF) and Live Cattle (LE)
2) Brent Oil (BZ) and Gas (RB)
3) Sugar (SB) and Coffee (KC)

Since we have established cointegration, we can now determine the spread between the assets. To do this, a regression was run on the two commodities, starting with Brent Oil and Gas, and found the constant beta coefficient. We plotted the spread against its mean and tested if the spread was stationary using the Augmented Dickey-Fuller (ADF) Test. From this test, we obtained a p-value for spread stationarity < 0.05. This means that we can reject the null hypothesis that the spread is non-stationary and accept the alternate hypothesis that the spread is stationary. Not only do these two commodities mean revert, but there is also a lot of oscillation. Since they mean revert quite frequently, we have many opportunities in which we can trade.

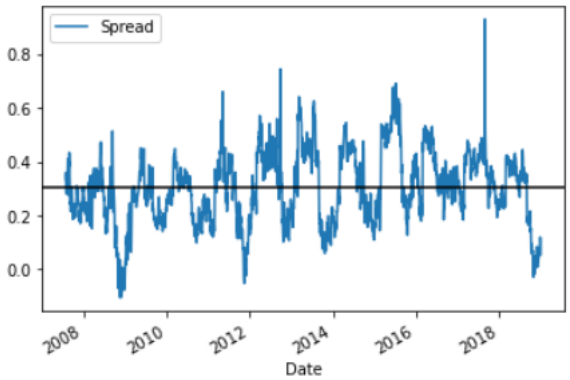

Figure 2: Spread between Brent Oil and Gas

Again a regression was run on Sugar and Coffee, to find the spread between the two commodities. We plotted the spread against its mean and tested if the spread was stationary using the Augmented Dickey-Fuller (ADF) Test. We obtained a p-value for spread stationarity < 0.05, again, indicating that the spread is stationary.

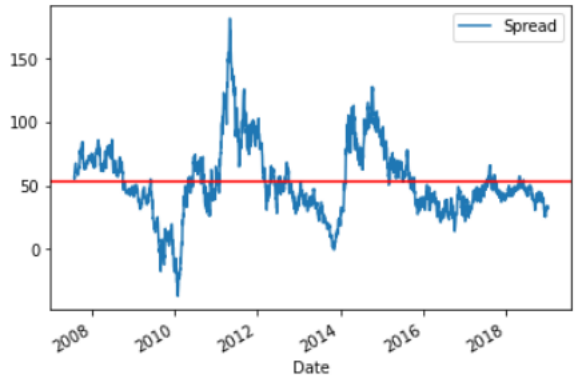

Figure 3: Spread between Sugar and Coffee

Lastly, a regression was run on Baby Cattle and Live Cattle. Following the same steps as the previous pairs, we found the spread, plotted it against its mean, and tested for stationarity. Our results show that the spread is stationary and mean reverts often, giving us many opportunities to trade.

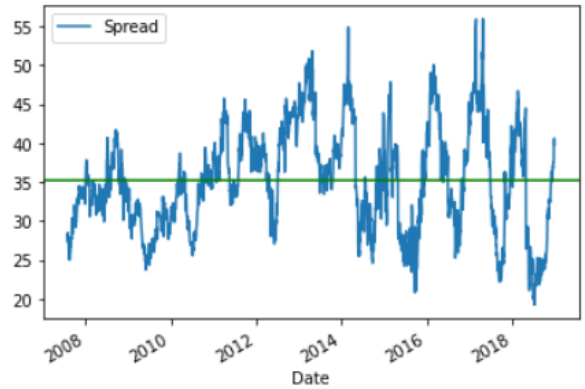

Figure 4: Spread between Baby Cattle and Live Cattle

4.2 Rolling Regression

The problem with the analysis so far is that all these spreads contain look-ahead bias. The constant beta coefficient has been calculated with data in the future from any given point besides the final days. To address this bias, we run a rolling regression for cointegrated pairs. Our new regression creates a rolling beta coefficient that rolls every day starting at day 30 and is reliant upon the previous 30 trading days. We plot the 1-day and 30-day moving average spread to visualize extreme spread events against their own averaged self.

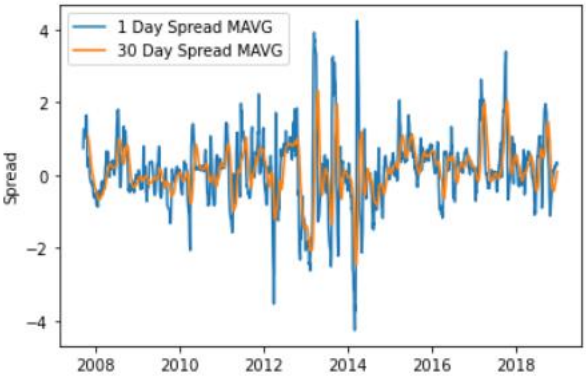

Figure 5: Rolling Spread for Brent Oil and Gas

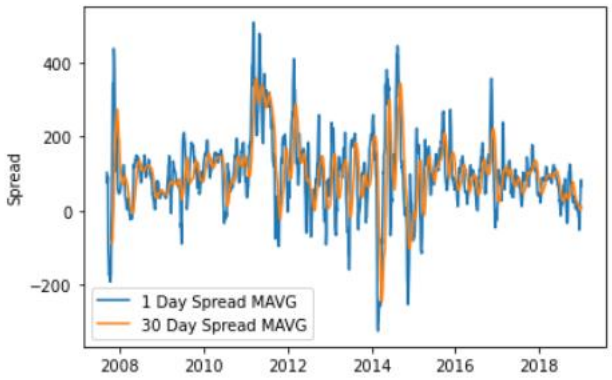

Figure 6: Rolling Spread for Sugar and Coffee

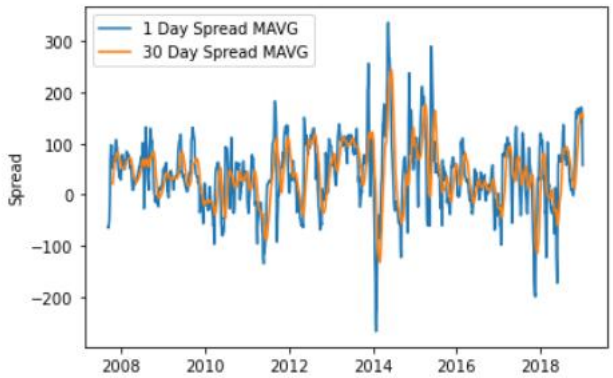

Figure 7: Rolling Spread for Baby and Live Cattle

From Figure 5, 6, 7 we could see that the results show that all pairs remained stationary.

5. Discussions

5.1 Trading Implementation

5.1.1 Baby cattle and Fully grown cattle Futures

Having addressed the look-ahead bias and smoothed out extremities, the following trading strategy has been developed:

- Long the Spread (long asset 1 and short asset 2) whenever the z-score is below -2
- Short the Spread (short asset 1 and long asset 2) whenever the z-score is above 2
- Exit positions when the z-score hits zero
- Stop-Loss of the positions when the spread reaches a z-score of $\pm$ 4

With the steps described above, a trading strategy has been formulated to gain statistical arbitrage. Figure8 shows the implementation of the strategy.

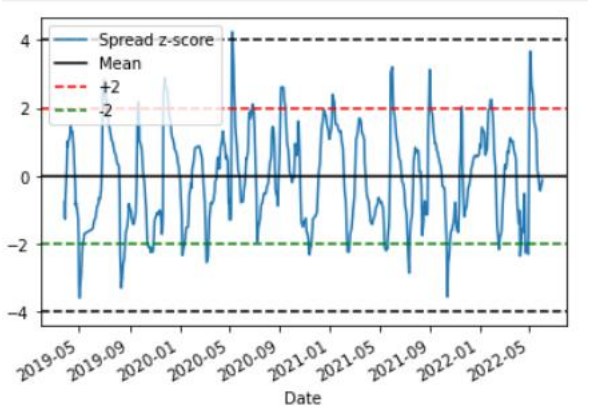

Figure 8: Baby Cattle and Fully Grown Cattle Futures Trading Strategy

Stop-loss is set as a Z-score of±4 as indicated in black dotted lines in Figure 8. It is assumed in this paper that whenever the Z-score crosses of ±4 the pair is said to have lost its cointegration and is advised to close the position to avoid incurring further loss.A Z-score of ±2 is kept as the trade opening position. In Figure 8whenever the Z-score hits ±4, one should long Baby Cattle futures and simultaneously short Live Cattle futures. Since the pair is mean reverting, when the Z-score hits or comes closer to zero, the position should be closed or otherwise sell the Baby Cattle Futures and buy the Live Cattle Futures.

To know the number of units to be short or long, the below formula was used.

Number of units to short or long price of Live Cattle Futures /Price of Baby Cattle Futures

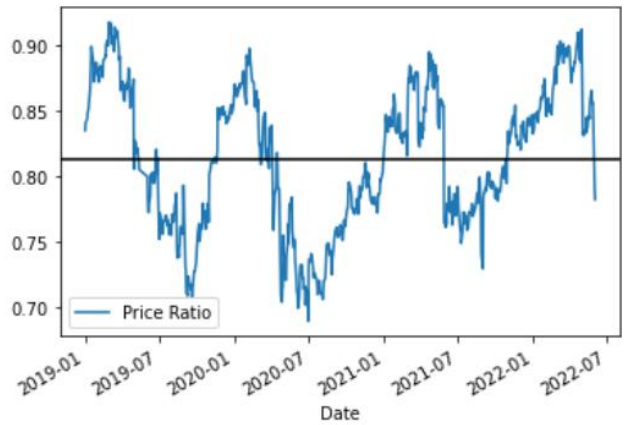

Figure 9: Baby Cattle and Fully Grown Cattle Futures number of shares to trade

Figure 9 is a representative of the ratio of the number of shares to long or short at any point in time. For example, in 07/2020, we could see that the ratio turned out to be approximately 0.68 i.e., for every 1 share of the long position of Baby Cattle Futures, 0.82 shares of Live Cattle Futures need to be short.

5.1.2 Sugar and Coffee Futures

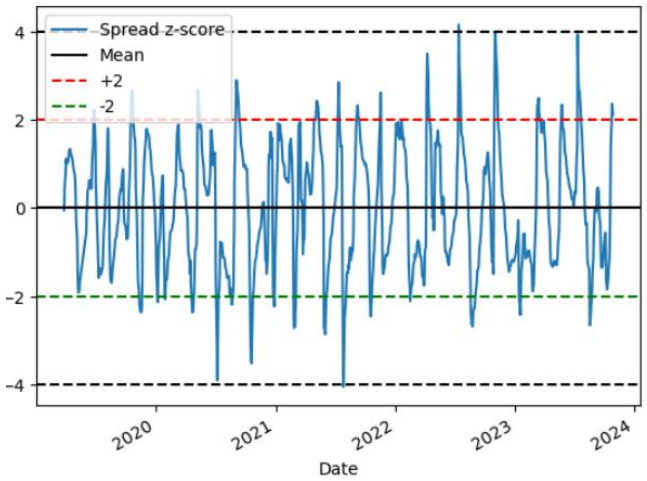

Figure 10: Sugar and Coffee Futures Trading Strategy

In Figure 10, we can see that the Sugar and Coffee are cointegrated based on a 30-day moving average and the spread was within Z-score ±4. There were 2 exceptions where the Z-spread was hitting ±4 and could have possibly resulted in exiting the position with loss as the stop-loss was breached.

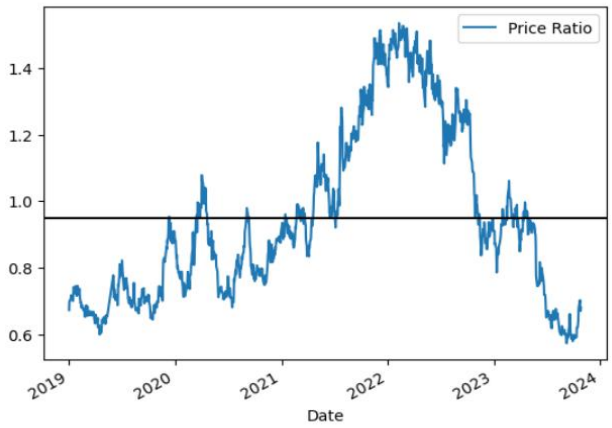

Figure 11: Sugar and Coffee Futures number of shares to trade

Figure11 depicts the ratio of number of units to long and short. Currently, in 2024, the ratio stands at 0.7 meaning, for every 1 long position of Coffee futures, short 0.7 shares of Sugar futures.

5.1.3 Gas and Brent Oil Futures

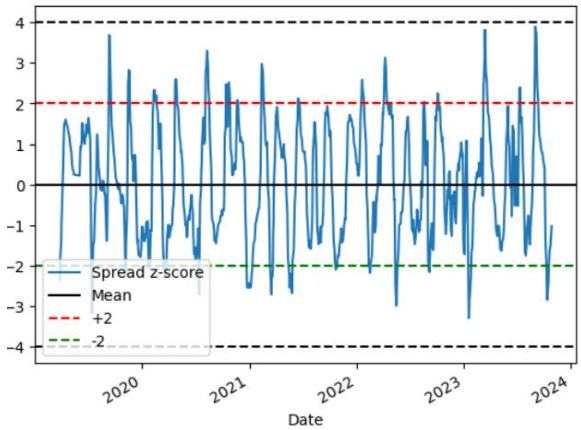

Figure 12: Gas and Brent Oil Futures Trading Strategy

Figure 12shows the Z-spread for Gas and Brent Oil Futures. Here the Z-score was not crossing ± 4 in any of the time windows and the trading strategy was effective.

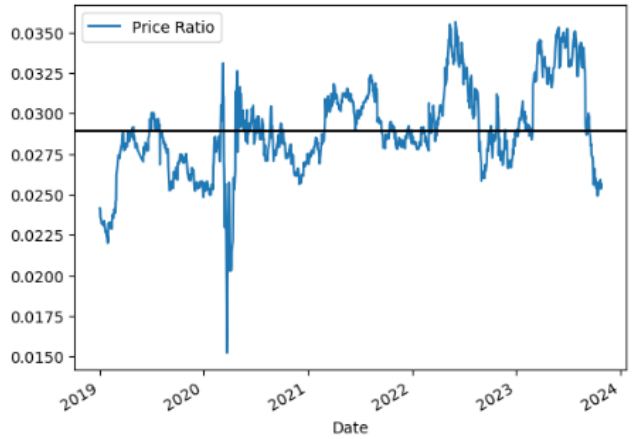

Figure 13: Gas and Brent OilFutures number of shares to trade

Figure 13 represents the share ratio as to how many shares to long or short per Gas future.

5.2 Advanced Methods

Copula in Latin, Copula means 'link' and the key idea here is modeling the relation between the two random variables to avoid adding idiosyncratic marginal distribution. It is another more complex tool that can be used for pairs trading [7]. This method enables the analysis of multiple random variables' dependency structure. A few advantages of using this advanced model are that it separates the marginal distributions from studying their “relation”. Further, it enables you to adjust the dependency structure to your liking and to work with multiple stocks in a group, instead of just a pair. All conventional trading strategies are primarily based on the assumption of linear association and use correlation coefficients or cointegration as a dependency measure. The use of cointegration for pairs trading assumes a symmetric distribution of spread around a mean of 0 and is bound by this linearity restriction. Copula, however, provides a strong framework for modeling dependence structure without such rigid assumptions. It divides individual marginal behavior estimation and dependence structure into two different ways. From an economic view, it gives analysts a chance to use different marginal distributions to account for the diversity in the risks of financial assets. Therefore, copula can be applied regardless of the form of the marginal distributions while providing much higher flexibility for practical application. Among the different types of Copula, the student-t is most often used for pairs trading due to the lower tail dependence [8]. Overall, Copula offers increased flexibility when working with multivariate time series.

6. Conclusion

To sum up, pairs trading is a strategy that is widely employed, usually by hedge funds, aiming to make a profit in any market environment. It has been around for quite some time with different and emerging methods for finding pairs and setting up trades. Our research revealed that cointegrated pairs can lose this property over sample periods. Thus, one should never solely focus on one pair but trade multiple pairs for further diversification. In general, pairs trading is an effective portfolio diversifier due to its market-neutral properties. To make this model more robust, we could test it across more pairs and try to retrieve high-interval data to exploit higher quantities of spreads. To find additional pairs we could use proven algorithmic methods to find additional pairs. Lastly, our model does not account for transaction costs or slippage which further reduces the strategy's profitability. Overall, our basic pairs model gives a broad overview of how statistical arbitrage strategies exploit temporary mispricing for commodity futures.

References

[1] Elliott, Robert J., John Van Der Hoek*, and William P. Malcolm. "Pairs trading." Quantitative Finance 5.3 (2005): 271-276.
[2] Clarence C.Y. Kwan, A note on market-neutral portfolio selection, Journal of Banking & Finance, Volume 23, Issue 5,1999, Pages 773-800, ISSN 0378-4266,https://doi.org/10.1016/S0378-4266(98)00114-9.
[3] Schmidt, Arlen David. "Pairs trading: a cointegration approach." (2009).
[4] https://finance.yahoo.com/quote/GD=F?p=GD=F&.tsrc=fin-srch
[5] https://www.spglobal.com/spdji/en/indices/commodities/sp-gsci/#news-research
[6] Engle, Robert F., and C. W. J. Granger. “Co-Integration and Error Correction: Representation, Estimation, and Testing.” Econometrica 55, no. 2 (1987): 251–76. https://doi.org/10.2307/1913236.
[7] Rad, Hossein, and Low, Rand Kwong Yew and Faff, Robert W., The Profitability of Pairs Trading Strategies: Distance, Cointegration, and Copula Methods (June 3, 2015).
[8] Miao, George J.. “High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach.” International journal of economics and finance 6 (2014): 96.