[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Pairs Trading Basics: Correlation, Cointegration And Strategy

Mean Reversion & Statistical Arbitrage Excel & R For Trading

Oct 23, 2019

13 min read

By Anupriya Gupta

Pairs trading is supposedly one of the most popular types of trading strategy. In this strategy, usually a pair of stocks are traded in a market-neutral strategy, i.e. it doesn't matter whether the market is trending upwards or downwards, the two open positions for each stock hedge against each other. The key challenges in pairs trading are to:

- Choose a pair which will give you good statistical arbitrage opportunities over time
- Choose the entry/exit points

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Statistics play a crucial role in the first challenge of deciding the pair to trade. The pair is commonly chosen from the same basket of stocks, for instance, Microsoft and Google (technology domain) or ICICI & Axis (Indian Banking) or Nifty Index and MSCI index (market indices). Among each domain, there are thousands of pairs are possible. The best ones are those which are based on mathematical or statistical tests. We will learn about two statistical methods in the next section of pairs trading.

Correlation

Though not common, a few Pairs Trading strategies look at correlation to find a suitable pair to trade.

Correlation is quantified by the correlation coefficient ρ, which ranges from -1 to +1. The correlation coefficient indicates the degree of correlation between the two variables. The value of +1 means there exists a perfect positive correlation between the two variables,

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

The correlation coefficient for the two variables is gi

Correlation(X,Y) = ρ = COV(X,Y) / SD(X).SD(Y)

where, cov (X, Y) is the covariance between X & Y while SD (X) and SD(Y) denotes the standard deviation of the respective variables.

If the correlation is high, say 0.8, traders may choose that pair for pairs trading. This high number represents a strong relationship between the two stocks. So if A goes up, the chances of B going up are also quite high. Based on this assumption a market neutral strategy is played where A is bought and B is sold; bought and sold decisions are made based on their individual patterns.

Just looking at correlation might give you spurious results. For instance, if your pairs trading strategy is based on the spread between the prices of the two stocks, it is possible that the prices of the two stocks keep on increasing without ever mean-reverting.

Spread = log(a) – nlog(b), where ‘a’ and ‘b’ are prices of stocks A

ctively.

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Cointegration.

Cointegration

The most common test for Pairs Trading is the cointegration test. Cointegration is a statistical property of two or more time-series variables which indicates if a linear combination of the variables is stationary.

Let us understand this statement above. The two-time series variables, in this case, are the log of prices of stocks A and B. Linear combination of these variables can be a linear equation defining the spread:

As you know, Spread = log(a) – nlog(b), where ‘a’ and ‘b’ are prices of stocks A and B respectively.

For each stock of A bought, you have sold n stocks of B.

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

With the theory in mind, let us try to answer the ques[illegible]hich you might be thinking of, in the next section of Pairs trading basics.

How to choose stocks for pairs trading?

For any pair of stocks, define the spread as below:

Spread = log(a) – nlog(b), where ‘a’ and ‘b’ are prices of stocks A and B respectively.

Assumption: n, the hedge ratio is constant.

Calculate ‘n’ using regression so that spread is as close to 0 as possible. Hence, we regress the stock prices to calculate the hedge ratio.

Theory: In regression, we get a term called the residuals which represents the distance of observed value from the curve fitting line or estimated value. These residuals tell us how much the actual

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Home Courses Trading Platform Contact Us

Login/Sign Up

Career Growth Tutorials

result. If this value is less than 0.05 or 0.01, we can [illegible] 95% or 99% confidence that the signal is stationary and we can choose this pair.

So far, we have discussed the challenges and statistics involved in selecting a pair of stocks for statistical arbitrage. We understood that by using the cointegration tests, we can say within a certain level of confidence interval that the spread between the two stocks is a stationary signal. In other words, this signal is mean-reverting. The spread is defined as:

Spread = log(a) – nlog(b), where ‘a’ and ‘b’ are prices of stocks A and B respectively. For each stock of A bought, you have sold n stocks of B. n is calculated by regressing prices of stocks A and B.

Having already established that the equation above is mean reverting, we now need to identify the extreme points or threshold levels which when crossed by this signal, we trigger trading orders for pairs trading.

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

N(0, 1) is very useful for creating threshold levels. Fo nple, in pairs trading, we have a distribution of spread between the prices of stocks A and B. We can convert these raw scores of spread into z-scores as explained below. This new distribution will have mean 0 and standard deviation of 1. It is easy to create threshold levels for this distribution such as 1.5 sigma, 2 sigma, 2.5 sigma, and so on.

How to calculate z-score?

z = (x – mean) / standard deviation, where x is a raw data point and z is the z-score.

Mean and standard deviation can be rolling statistics for a period of ‘t’ days or minutes or time intervals.

Moving average

We divide the data into subsets of size ‘t’, where,‘t’ specifies a fixed time period for which the average is to be calculated. For example, the moving average of the prices of stock A where ‘t’ is

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

entry points for Pairs Trading.

Defining Entry points

Let us denote the Spread as s. Thus,

Spread = s = log(a) – nlog(b)

Calculate z-score of ‘s’, using rolling mean and standard deviation for a time period of ‘t’ intervals. Save this as z.

Define threshold as anything 1.5-sigma, 2-sigma. This parameter will change as per the backtesting results without risking overfitting to data.

When Z-score crosses upper threshold, go SHORT:

Sell stock A

Buy stock B

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Home Courses Trading Platform Contact Us

Login/Sign Up

Career Growth Tutorials

Defining Exit points

STOP LOSS

Stop loss is defined for scenarios when the expected do not happen. For example, if we chose entry signals at 2-sigma, we are expecting that the spread will revert back to mean from this threshold. However, it is possible that spread continues to blow up. Say it reaches 2.5-sigma and you incurred losses. To prevent further losses, you place Stop Loss at say 3-sigma.

In addition to placing a pre-defined stop-loss criterion such as 3-sigma or extreme variation from the mean, you can check on the co-integration value. If the co-integration is broken during the pair is ON, the strategy warrants cutting the positions since the basic hypothesis is nullified.

TAKE PROFIT

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Home Courses Trading Platform Contact Us

Login/Sign Up

Career Growth Tutorials

risk appetite and backtesting results.

What often works is your experience and a broad range of potent skillsets that allow you to grasp a hold of the complete scenario before jumping to conclusions and help you understand practically. Like we mentioned, your appetite for risk and backtesting results will work for you. Automation and practical applications are the keys here. Anto, who had been trading for 10 years, evolved his skillsets and adapted to the growing markets with the Executive Programme in Algorithmic Trading (EPAT) and is happily trading in this domain.

Let us try to recap what we have understood so far. Pairs Trading is a trading strategy that matches a long position in one stock/asset with an offsetting position in another stock/asset that is statistically related. Pairs Trading can be called a mean reversion strategy where we bet that the prices will revert to their historical trends.

So far, we have gone through the concepts and now let us try to create a simple Pairs Trading strategy in Excel.

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Home Courses Trading Platform Contact Us

Login/Sign Up

Career Growth Tutorials

Why should you download the trading model?

As the trading logic is coded in the cells of the sheet, you can improve the understanding by downloading and analyzing the files at your own convenience. Not just that, you can play around the numbers to obtain better results. You might find suitable parameters that provide higher profits than specified in the article.

Explanation of the model

In this example, we consider the MSCI and Nifty pair as both of them are stock market indexes. We implement mean reversion strategy on this pair. Mean reversion is a property of stationary time series. Since we claim that the pair we have chosen is mean reverting we should test whether it follows stationarity.

Plotting of the logarithmic ratio of Nifty to MSCI makes it appear to be mean reverting with a mean value of 2.088 but we use Dicky [illegible] to test whether it is stationary with a statistical

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Home Courses Trading Platform Contact Us

Login/Sign Up

Career Growth Tutorials

minute closing price only.

- Since this is discrete data, squaring off of the position happens at the end of the candle i.e. at the price available at the end of 5 minutes.
- Only the regular session (T) is traded
- Transaction costs are $0.375 for Nifty and $1.10 for MSCI.
- The margin for each trade is $990 (approximated to $1000).

Input parameters

Please note that all the values for the input parameters mentioned below are configurable.

- Average of 10 candles (one candle is equal to every 5-minute price) is considered
- A “z” score of +2 is considered for buy and -2 for selling
- A stop loss of $100 and profit limit of $200 is set
- The order size for trading MSCI is 50 (1 lot) and for Nifty is 6 (3 lots)

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Input Parameters		
Average	10	
Buy Threshold	-2	
Sell Threshold	2	
Stop Loss	-100	*USD*
Take Profit	200	*USD*
MSCI Order Size	50	*i.e. 1 lot*
Nifty Order Size	6	*i.e. 3 lots*

S No.	Date & Time	MSCI	NIFTY	I(0)	10 Candle Avg	Std Dev	z-Score	Signal	Buy price	Sell Price	MTM	Status	Profit/Loss	Cumulative
1	17-02-2015 10:00	1092.4	8822.5	2.088928									0	0
2	17-02-2015 10:05	1092.4	8822.5	2.088928									0	0
3	17-02-2015 10:10	1092.4	8822.5	2.088928									0	0
4	17-02-2015 10:15	1092.4	8822.5	2.088928									0	0
5	17-02-2015 10:20	1092.4	8822.5	2.088928									0	0
6	17-02-2015 10:25	1092.4	8822.5	2.088928									0	0
7	17-02-2015 10:30	1092.4	8830	2.089778									0	0
8	17-02-2015 10:35	1092.4	8830	2.089778									0	0
9	17-02-2015 10:40	1092.4	8830	2.089778									0	0
10	17-02-2015 10:45	1092.4	8830	2.089778									0	0
11	17-02-2015 10:50	1092.4	8829	2.089665	2.0892681	0.000438803	0.903791						0	0
12	17-02-2015 10:55	1092.4	8821	2.088758	2.0893417	0.000437219	-1.33475						0	0
13	17-02-2015 11:00	1092.4	8821	2.088758	2.0893247	0.000457905	-1.23732						0	0
14	17-02-2015 11:05	1092.4	8825.5	2.089268	2.0893077	0.000477024	-0.08292						0	0
15	17-02-2015 11:10	1092.4	8825	2.089211	2.0893417	0.000458733	-0.28384						0	0
16	17-02-2015 11:15	1092.4	8824.5	2.089155	2.08937	0.000438665	-0.49058						0	0
17	17-02-2015 11:20	1092.4	8824.5	2.089155	2.0893927	0.000418697	-0.56811						0	0
18	17-02-2015 11:25	1092.4	8823.5	2.089042	2.0893304	0.000400991	-0.72043						0	0
19	17-02-2015 11:30	1092.4	8808.5	2.08734	2.0892568	0.00037655	-5.09017	Buy	8808.5	1092.4		Buy	0	0
20	17-02-2015 11:35	1092.4	8817.5	2.088361	2.089013	0.000673625	-0.96745		8808.5	1092.4	54	Buy	0	0
21	17-02-2015 11:40	1092.4	8822.5	2.088928	2.0888713	0.000643154	0.08841		8808.5	1092.4	84	Buy	0	0
22	17-02-2015 11:45	1092.4	8818.5	2.088475	2.0887977	0.000581419	-0.5555		8808.5	1092.4	60	Buy	0	0

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

The formula –IF(A23>$C$3, AVERAGE(INDEX($E$1 1358, A23-$C$3):E22), "") means that the average should be calculated only if the data sample available is more than 10 (i.e. the value specified in cell C3), otherwise the cell should be blank.

Consider cell F22. Its corresponding cell A22 has a value of 10. Since A22>$C$3 fails, the entry in that cell is blank. The next cell F23 has a value since A23>$C$3 is true. Let's move to the next column.

In column G, the formula, AVERAGE(INDEX($E$13:$E$1358, A23-$C$3):E22) calculates the average value of last 10 (as mentioned in cell C3) candles of column E data. Similar logic holds for column G where the standard deviation is calculated.

The "z" score is calculated in the column H. Formula for calculating "z" score is z= (x-μ)/(σ). Here x is the sample (Column E), μ is the mean value (Column F) and σ is the standard deviation (Column G).

Column I represents the trading signal. As mentioned in the input if "z" score goes below -2 we buy and if it goes above

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

In each new row while the position is continuing, we [illegible] whether the stop loss (as mentioned in cell C6) or take profit (as mentioned in cell C7) is hit. The stop loss is given the value of USD -100, i.e. loss of USD 100 and take profit is given the value of USD 200 in the cells C6 and C7 respectively.

While the position does not hit either stop loss or take profit, we continue with that trade and ignore all signals that are appearing in column I. Once the trade hits either the stop loss or take profit, we again start looking at the signals in column I and open a new trading position as soon as we have a Buy or Sell signal in column I.

Column M represents the trading signals based on the input parameters specified. Column I already has trading signals and M tells us about the status of our trading position i.e. are we long or short or booked the profits or exited at the stop loss. If the trade is not exited, we carry forward the position to the next candle by repeating the value of the status column in the previous candle. If the price movement occurs in such a way that it breaches the given

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Home Courses Trading Platform Contact Us

Login/Sign Up

Career Growth Tutorials

calculates the cumulative profit.

Outputs

The output table has some performance metrics tabulated. Loss from all loss-making trades is $3699 and profit from trades that hit TP is $9280. So the total P/L is $9280-$3699=$5581. Loss trades are the trades that resulted in losing money on the trading positions. Profitable trades are the successful trades ending in gaining cause. Average profit is the ratio of total profit to the total number of trades. Net average profit is calculated after subtracting the transaction costs which amounts to $91.77.

Now it is your turn!

- First, download the model
- Modify the parameters and study the backtesting results
- Run the model for other historical prices
- Modify the formula and strategy to add new parameters and

Our cookie policy s! Play with logic! Explore and study!

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Learn how to implement pairs trading/statistical arbit rategy in FX markets through a project work including live examples. If you want to dig deeper and try to find suitable pairs to apply the strategy, you can go through the blog on K-Means algorithm.

If you want to learn various aspects of Algorithmic trading then check out our Executive Programme in Algorithmic Trading (EPAT). The course covers training modules like Statistics & Econometrics, Financial Computing & Technology, and Algorithmic & Quantitative Trading. EPAT is designed to equip you with the right skill sets to be a successful trader. Enroll now!

Login to Download

Disclaimer: All data and information provided in this article are for informational purposes only. QuantInsti® makes no representations as to accuracy, completeness, currentness, suitability, or validity of any information in this article and will not be liable for any errors, omissions, or delays in this information or any losses, injuries, or damages arising from its display or use. All information is provided on an as-is basis.

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

https://blog.quantinsti.com/pairs-trading-basics/

Sep 28, 2017

Johansen Cointegration Test: Learn How to Implement it in Python

Feb 15, 2021

Mean Reversion Strategies: Introduction, Trading, Strategies and More

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

The "n" parameter.

J

Jan Tarnogrocki • 2 Years Ago

Hi There, In the excel file you are trading 50 contracts MSCI and 6 Nifty. The average net profit per trade is 91.77 USD. But the spread is not taken into consideration. If it is above 1.64 USD per contract the strategy is at break even. In addition slippage on real markets has to be outperformed too.

I do really like this piece of work you are presenting here especially how complex statistic is explained in an easy and understandable way.

But I am afraid the results in the excel file shown are far away from reality.

(I would be happy if I were wrong. But I coded the strategy in Java using tick data for back testing including spread and there is no penny left at the end of the day)

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Home Courses Trading Platform Contact Us

Login/Sign Up

Career Growth Tutorials

opened and closed (not the spread between the cointegrated pairs used for generating trade signals).

The assumption is also posted within the excel file: “Bid Ask Spread is ignored.”

Dirk Damman • 2 Years Ago

Hello, If you change the exit parameters based on I(0) 1or -1. You get the following results. This looks more profitable ?

Output

Loss from SL -893

Profit from TP 7865

Total Profit 6972

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Hi! Thank you for the comment. As we said in the article, this sheet was only for illustration. You can play around the numbers to obtain better results. And you can definitely find parameters that profit higher profits than specified in the article. Hope that answers your query?

QuantInsti • 2 Years Ago

Hi! Thank you for the comment. As we said in the article, this sheet was only for illustration. You can play around the numbers to obtain better results. And you can definitely find parameters that profit higher profits than specified in the article.Hope that answers your query?

Miguel de Lazio • 2 Years Ago

Why is the last value of "I(0)" not used in the 10-period moving average? MA is calculated from previous periods

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

I don't understand why there wo[illegible]e a look-ahead, at least not when I assume that the time column shows the opening time of the bar. The closing prices are available at the end of the bars when the signal is calculated. In the Excel sheet at row 22, closing prices of 10 bars are available. (closing time of this bar is one tick before 10:50, in my program I set this to 10:49:59 and 999 milliseconds which becomes the timestamp of the signal) Then you make a trading decision, e.g. place buy order. The profit in the next bar can be calculated using the closing price and 'closing signal' of the previous bar and closing price of the next bar.

Miguel de Lazio • 2 Years Ago

Hi again, I think I can see what you mean. The VLOOKUP is doing a lookup using the Date &

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

QuantInsti • 2 Years Ago

Yes, you are right. When you change the order of the raw data, keeping all the parameters same, the value of spread changes and thus z-score.Note: In most cases, there would be a minor difference in the spread so it won't make much difference.

Miguel de Lazio • 2 Years Ago

I do not understand how to calculate the relative order size using the 'n' ratio that is found using "log(a) - nlog(b)". If 'n' would be calculated from "a – nb" then the ratio will be directly related to how much the two prices will have to change (relatively speaking) before the spread reverts to the mean. But "log(a) - nlog(b)" gives a completely different 'n' value. So how is it used to manage the risk and order size?

QuantInsti • 2 Years Ago

Thank you for the comment, Miguel and apologies for the delay in the response. When

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Hi! Nice analysis! Can you explain the theoretic assumption behind using rolling mean in the computation of the z-score? Since it modifies the structure of the vector, making it more stationary, why is it a reliable measure to build the strategy on? Maybe my reasoning is wrong, I hope I have explained myself. Thanks, Bye!

QuantInsti • 2 Years Ago

Thank you for your comment. We'd be glad to help you out.

Z-score describes a value's relationship to the mean of a group of values. It is measured in terms of standard deviation from the mean.

Z-score = (x - Mean) / Standard deviation

Mean reversion strategy involves the purchase or sale of financial instruments whose recent performance differs from their historical

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

QuantInsti • 2 Years Ago

Hi! Thank you for the comment. This is one of the assumptions of the strategy. What it tries to say is that if you are taking a position in Nifty futures, you need to keep around $990 or around INR 75,000 with the broker to initiate the trade. This margin varies from broker to broker. We hope this helps with your query.

K

Khalil Elhanafi • Last Year

Thank You for your reply. We need 990$ * 2 for the strategy because in each sell or buy we are taking two position. is that true ?

K

Khalil Elhanafi • Last Year

Thank you for your reply, can we use the same strategy with cryptocurrencies ?

QuantInsti • Last Year

Hi! Yes, you can use the same strategy with cryptocurrencies.

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

We'd be glad to help you, Davic self-paced specialization courses from Quantra can help you learn how to do it. Feel free to connect with our team here: https://bit.ly/31sTxoG

masoud shirin nasab • Last Year

HI, CAN MAKE A EX : FOR FOREX PAIRS : EURUSD AND USDCHF?

QuantInsti • Last Year

Hi! Thanks for your comment. As suggested in the article, there are tests of correlation, and cointegration can be performed to check the feasibility of pair trading between the two instruments. The correlation can be verified using 1. online tools like https://www1.oanda.com/lang/en/forex-trading/analysis/currency-correlation.2. through excel "=CORREL(A2:A7,B2:B7)" 3. through python

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

QuantInsti • Last Year

Hi Vedant, apologies for the late reply. We have fixed the download button. Do check.

도경

?? ? • Last Year

hi what does mtm and status mean?

V

Vibhu • Last Year

MTM, also known as mark-to-market, is the concept of valuing assets and liabilities at the last known price is known as mark-to-market. Consider you are long 1 contract in August corn futures trading at $500. At the end of the day, the price falls to $490. You were long, and the price fell, so you have an unrealised loss of $10. This $10 will be deducted from your account at the end of the day. And the procedure is called daily mark-to-market settlements.

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

S

Simone Bosio • 10 Months Ago

Hi There, great article and superb explenation. The only thing I'm missing is the "Return per month" calculation (cell: K10). Why are you multiplying the average net profit times 2 and dividing the result by 1000? Thank you in advance

QuantInsti • 10 Months Ago

Hi Simone, thank you for your comment. Since we are taking a trade on two instruments, each utilising a margin of $1000, we divide the "Net Average Profit" with $2000, to get the percentage of profit per trade.

However, there is a discrepancy in this. You need to apply the formula as `=K9/(1000*2)`

We will rectify this in the downloads as soon as possible. Hope this helps.

S

Simone Bosio • 10 Months Ago

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

what does it mean by "1 lot" and "3 lots"?

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

About

QuantInsti
Quantra
Blueshift

Events & Announcements

Announcements

Our cookie policy

[FREE WEBINAR] Introduction to Systematic Options Trading Register now!

Our cookie policy