Evaluation of Statistical Arbitrage Opportunities in the Indian Derivatives Market

University: Narsee Monjee Institute of Management Studies

# ABSTRACT

In this research paper, we match pairs of equity stocks in the Indian stock market by comparing the movement of the stock prices of various stocks belonging to the same sector; the stocks that show a similar pattern of movement over a specific period of time are taken into consideration for our raw data. Further, we run two regression equations in case of each pair; The first regression equation takes one stock of the pair as independent, other as dependent and vice versa in case of the second regression equation. The regression equation with the least error ratio is selected and an Augmented Dickey-Fuller (ADF) test is done on the residual values of the selected regression equation to check for co-integration between the selected pairs. The Z-Scores of these residual values is considered for selecting the strategic entry and exit points of the relative value trade. The aggregate return we earn via this method in the Indian equity market, beats the risk-free rate, thus providing us with a viable statistical arbitrage opportunity. This is a market neutral arbitrage algorithm which is very similar to the Engle-Granger method, but nonetheless, a simpler version of the same which could potentially give sophisticated investors a net average annualized return of 20-30%

Keywords: *Pairs trading, Statistical Arbitrage, ADF test, Co-integration, Stationarity, Z-Score*

# 

# INTRODUCTION

Even 10 years after the Global Financial Crisis (GFC), we have seen a high amount of turbulence and volatility in the global equity markets, making it difficult even for sophisticated investors to predict the market direction, let alone predict the prices. Hence, investors need a market neutral strategy which can provide them with handsome returns above and beyond the risk-free rate and we think Statistical Arbitrage or Relative Value Trading, on which extensive work has been done since decades, could bridge that gap.

Although, this is not a new quantitative trading strategy and has been exploited for a long period of time, reducing its profitability to some extent, we feel that there could still be pockets of opportunities in the Indian equity markets as the derivatives segment here has only recently gained a lot of traction and is evolving rapidly with new products along with high participation evident from the increasing volumes. This mean reversion strategy, commonly known as pairs trading, that we intend to test and apply, works on the assumption that the price of the underlying asset has a stable trend and the prices fluctuate around this trend. The strategy waits for the actual price to deviate from the predicted values found using the OLS regression equation and then goes long on the under-priced stock while simultaneously short selling the overpriced counterpart, closing the positions as the relationship returns to statistical norms. This mean reversion is the backbone of pairs trading and is the reason for the existence of an arbitrage opportunity. However, it is possible that two stocks which are presently co**-**integrated may not be so in the future; therefore, our dataset needs to be constantly updated with a rolling regression, beta, and Z-scores for better accuracy and risk management. Pairs trading is a strategy that requires certainmarket inefficiencies, which are more prevalent in developing markets like India. In the paper “Nonlinear dependence modelling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100”, Christopher Krauss and Johannes Stübinger demonstrated that refined statistical arbitrage is possible even in markets that seem to be efficient. So, in the future, if our markets become more efficient in exploiting themispricing between co**-**integrated pairs, the strategy could still be viable.

# LITERATURE REVIEW

Pairs trading as a quantitative trading strategy is not only confined to hedge funds which have widely used it to marginally beat the risk free rate but has also become accessible to retail investors in the recent years owing to the exponential improvement in the availability of technology, instant price data and analytical tools. The same has been proven by Sharjil Muktafi Haque and A. K. Enamul Haque in their paper “Pairs trading strategies in Dhaka stock exchange: Implementation and profitability analysis.” The market neutrality of this strategy is extremely important for countries like Bangladesh where the stock markets are extremely volatile due to macroeconomic factors, and even in such a scenario their pairs trading algorithm generated a return of 15.38% using out-sample data.

In the paper “Statistical Arbitrage Pairs Trading with High-frequency Data” by Johannes Stübinger and Jens Bredthauer it was concluded that the best results are achieved by setting the standard deviation 2.5 times (positive or negative) from the mean. They also found that using the parameter setting suggested by literature on daily data was too aggressive as the high transaction cost could not be compensated by higher returns.

In the 2003 paper “A Note on the Effectiveness of Pairs Trading For Individual Investors” by Fabio Pizzutilo, it was shown that a retail investor facing certain real-life restrictions such as initial margins, interest costs, cash guarantees and limitations to short selling significantly impacted the returns generated from pairs trading. However, the returns remained largely positive. The three and five-year beta was found to be zero implying that pairstrading is a market neutral trading strategy.

In “Evaluation of pairs-trading strategy at the Brazilian financial market” by Marcelo Scherer Perlin it was found that this statistical arbitrage technique was a success in Brazilian stock markets when tested using the daily price data. The raw returns given by the bootstrap approach was owed to skill rather than chance.

“High-frequency pairs trading on small stock exchange” by Andreas Mikkelsen and Frode Kjær showed that with a shorter formation period, trading the profits go up for high-frequency stocks. Also, the distance approach has shown lower risk; while on the other hand, the co-integration method is less trading intensive.

**METHODOLOGY**

1. For statistical arbitrage to work we need to find pairs that are co-integrated and not correlated; as correlation need not imply that the difference in prices of the paired stocks will mean revert. However, if the stocks are co-integrated it implies that the difference in stock prices of the pair shall mean revert as their regression residuals shall be stationary in nature, thus giving us an opportunity for a statistical arbitrage by trading that spread from an extreme value as it moves towards its mean.
2. To find stocks that might be co-integrated; we have considered various pairs of stocks belonging to the same sector with similar business models, being affected by the same macroeconomic fundamentals to shortlist for a co-integration test.
3. We then prepare two regression equations for each pair with the prices of one stock being the independent variable and the other a dependent variable; vice versa for the second regression equation. Next, we find the error ratio for both of the regression equations; the error ratio here is defined as a ratio of the standard error of the intercept and the standard error of the residual. We select the regression equation with the lowest error ratio.
4. An Augmented Dickey-Fuller test (ADF) is carried out on the residuals of the selected regression equation (using NumXL) to check for stationarity; this essentially means that the difference between the predicted price via regression and the actual price of the stock is mean reverting and that there is no trend.
5. The testing procedure for the ADF test is applied to the following model:

|  |
| --- |
| Augmented Dickey-Fuller Test in Excel (ADF) formula |

1. In the above model,

$ \Delta  $ is the first different operator

$  \alpha $ is a constant

$  \beta_1 $ is the coefficient on a time trend

$  \beta_2 $ is the coefficient on a squared time trend

1. We use this model to test for a unit root which is the same as testing that$ \gamma = 0 $
2. In all, the Augmented Dickey-Fuller Test in Excel (using NumXL)the hypothesis is as follows:

|  |  |
| --- | --- |
| Null hypothesis of Augmented Dickey Fuller (ADF) Test |  |

|  |  |
| --- | --- |
| Alternate hypothesis of Augmented Dickey Fuller (ADF) Test |  |

1. Where:

$ H_{o} $ is the null hypothesis (i.e. $ y_t $ has a unit-root)

$ H_{1} $ is the alternate hypothesis (i.e. $ {y_t} $ does not have a unit-root)

1. The test statistics ($ \tau $) value is calculated as follows:

|  |  |
| --- | --- |
| Tau calculation formula for Augmented Dickey Fuller (ADF) Test |  |

1. Where:

$ \hat{\gamma} $ is the estimated coefficient

$ \sigma_{\hat\gamma} $ is the standard error in the coefficient estimate

1. The calculated statistic value is compared to the relevant critical value by referring to the Dickey-FullerTable since it is not possible to use a standard t-distribution here.
2. If the calculated ADF test statistic lies below the critical value defined as per the ADF test table i.e. in the rejection region, then we reject the null hypothesis and accept the alternate hypothesis that the data is stationary. Alternatively, we can also look at the P-value and conclude that if the P-value is below the defined level of significance we shall reject the null hypothesis and accept the alternative hypothesis implying that the data is stationary.
3. If the Augmented Dickey-Fuller test for the residuals shows that the residual values are stationary, we conclude that the pair is co-integrated.
4. Finally, we find the Z-scores of the residual values by dividing each residual value with its standard error or the standard deviation of the residual values. This is done for ease of calculations.
5. We define the parameters of entry and exit of the trades in the form of Z-scores for example: +2 SD and -2SD. Also, we set a target Z-score where we can book our profits and also set a stop loss; which might vary from one pair to another.
6. We take a long position on shares that are underpriced and take a short position on shares that are overpriced as per our model output. The position sizing should ideally be decided using the beta values, however, we shall use one futures lot of each stock in the pair for return calculation and back-testing purposes; since it requires a lower amount of capital to be blocked.

# DATA ANALYSIS

1. As defined in the research methodology, the pairs selected for the stationarity test were as follows; the output of the ADF test on their regression residual values have also been mentioned in the table below:



*Note: The regressions have been run on these pairs for a period of five hundred trading days or more.*

1. Out of the above-mentioned pairs, we find eighteen pairs whose regression residuals are concluded to be stationary as per the Augmented Dickey-Fuller test results and hence, considered them to be co-integrated. Out of these eighteen pairs, we shortlist the pairs which are at an extreme; defined here as having a P-value of less than 0.5%.
2. We find eight such pairs with P-values below 0.5% from which we trickle down to just five pairs which have historically provided higher trading opportunities; for the purposes of return calculations and backtesting.



1. After calculating the Z-scores of the stationary residuals, we define the entry and exit parameters (long and short) for the selected pairs in terms of Z-scores ranging from positive or negative 1.5 to 2 standard deviations.
2. Since we are trading in the futures market, we only pay a certain percentage of the contract value which may differ from one stock to another and from one point in time to another depending on the stock’s volatility, liquidity and market conditions. For the purposes of the research, we shall assume no leverage and calculate the unlevered profits that can be generated from such a quantitative trading strategy.
3. The following are the results for each of the pairs traded in the period between 12th December 2017 to 5th October 2018, with long and short signals defined in terms of Z-scores along with the stop-loss and target profit levels also being defined in terms of standard deviations of the residual values; which may differ from one pair to another.











1. It can be observed from the above back tested profit and loss statements of the traded pairs, that inspite of getting less frequent trading opportunities, sometimes as low as two trades in an entire year, we are able to generate handsome profits while also reducing the potentially high transactions costs that come with high-frequency trading.

# The cumulative or aggregate return table below shows that this relatively small portfolio was able to generate an absolute return of 12.48% (unlevered), with our capital being blocked for less than a year.



1. It is important to note here that above return calculations do not take into consideration the transaction costs, brokerage fees and capital gains tax.

# CONCLUSION

The objective of our research paper was to examine the existence and feasibility of relative pricing strategies like the statistical arbitrage using co-integrated pairs of equity stocks in the Indian derivatives market and find out if such strategies have the robustness to beat the risk free rate (State Bank of India fixed deposit rate of 6.4% p.a.) and provide a significant alpha. Based on empirical results we found that there were quite a handful of pairs that exhibited a relation of co-integration and could thus be exploited for generating above-average returns by using the divergence window where the pairs of stocks are mispriced. In times of high volatility such as the one experienced since the beginning of 2018, where we have seen the Volatility Index (VIX) spike to multiyear highs, a market neutral strategy could be a good value addition to a trader’s portfolio. From our research we can conclude that the statistical arbitrage model tested on an out of sample data of an entire year, did not just marginally beat the risk free rate i.e 6.4% p.a, but gave us an absolute unlevered return of 12.48% (positions were open for 30 days on average), average annualized return of 55% with an average hit ratio of 85%. Ideally, the test can be carried out for a longer horizon to test the viability much more rigorously and use the tools at our disposal such as machine learning algorithms, neural networks to find out the most co-integrated pairs, best entry and exit points, thus resulting in higher alpha and better risk management

.

# 

# REFERENCES

* Qazi, L. T., Ur Rahman, A., & Gul, S. (2015). Which pairs of stocks should we trade? selection of pairs for statistical arbitrage and pairs trading in Karachi stock exchange. *Pakistan Development Review, 54*(3), 215-244.
* Krauss, C., &Stubinger, J. (n.d.). Nonlinear dependence modelling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100, 39.
* Sharjil, M. H., & Haque, A. K. E. (2014). Pairs trading strategy in Dhaka stock exchange: Implementation and profitability analysis. *Asian Economic and Financial Review, 4*(8), 1091-1105.
* Stübinger, J., &Bredthauer, J. (2017). Statistical arbitrage pairs trading with high-frequency data. *International Journal of Economics and Financial Issues, 7*(4)
* Pizzutilo, F. (2013). A note on the effectiveness of pairs trading for individual investors. *International Journal of Economics and Financial Issues, 3*(3), 763-n/a.
* Perlin, M. S. (2009). Evaluation of pairs-trading strategy at the Brazilian financial market. *Journal of Derivatives & Hedge Funds, 15*(2), 122-136.
* Mikkelsen, A., &Kj, F. (2018). High-frequency pairs trading on a small stock exchange. *International Journal of Economics and Financial Issues, 8*(4), 78-88.
* Yahoo Finance - Business finance, stock market, quotes, news. (n.d.). Retrieved October 18, 2018, from <https://in.finance.yahoo.com/>
* Jarque, Carlos M.; Anil K. Bera (1980). "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". Economics Letters 6 (3): 255-259.
* Ljung, G. M. and Box, G. E. P., "On a measure of lack of fit in time series models." Biometrika 65 (1978): 297-303
* Enders, W., "Applied econometric time series", John Wiley & Sons, 1995, p. 86-87
* Shapiro, S. S. and Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)", Biometrika, 52, 3 and 4, pages 591-611
* SBI FD Interest Rate, SBI Fixed Deposit Rates Oct 2018. (n.d.). Retrieved October 20, 2018, from<https://www.myloancare.in/fixed-deposit/fd-interest-rates/sbi>
* Engle, R., & Granger, C. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica,55*(2), 251-276. doi:10.2307/1913236