Statistical Arbitrage: Profits through Pairs Trading

Saloni Gupta

Associate Professor, Bharti College, University of Delhi, Delhi
Email Id: salonigupta0609@gmail.com

Abstract. Statistical arbitrage is a popular device among hedge fund managers and assets management professionals. It refers to simultaneous buying and selling two different capital assets to earn super-normal profit. By identifying persistent anomalies that violate the efficient market hypothesis, statistical methods can be used to create a trading strategy to generate profit with high probability. A pair trading is one such trading strategy which is based on statistical arbitrage process. Pairs trading can be simple in concept, but can be one of the most complex types of trading in practice. The starting point of this strategy is that stocks that have historically had the same trading patters will have so in future as well. If there is a deviation from the historical mean this creates a trading opportunity, which can be exploited. Gains are earned when the price relationship is resorted. The basic premise of this strategy is that stock prices follow a mean reverting process. The objective of this paper is to identify arbitrage opportunities and calculating profits earned through these opportunities by using statistical tools.

Many questions need to be answered before one can implement such strategy viz. which pair of stocks should be traded, how much do we buy/sell of each stock, how to catch the signal of an opportunity (i.e opening a position) and when to close the position so that profit could be earned. In this paper we have taken daily closing prices from 1/1/2010 to 1/1/2011 of thirty scrips of BSE-Sensex to form pairs. Pairs are formed on the basis of minimum distances between two stocks. We have decided not to invest anything. That is, purchase the same rupee amount of the long stock as we sell of the short stock so that strategy is self-financing. We open a position when the absolute value of the difference gets larger than two of its historical standardization. To unwind the position, we wait until the first time it crosses zero. To calculate the profit/loss of this strategy, we have used "R-Software". It is observed that profit could be earned through pairs trading if it is applied without losing patience. By identifying persistent anomalies that violate the efficient market hypothesis, statistical methods can be used to create a trading strategy to generate profit with high probability.

Keywords: Arbitrage, Trading, R-software

1 Introduction

To begin, first consider the question "What is Statistical Arbitrage?" There is no single widely accepted definition of statistical arbitrage. Generally it refers to simultaneous buying and selling two different capital assets to earn super-normal profit. The term is also used to connote a specific trading strategy, pairs trading, for example. In short, we can think of statistical arbitrage as a long horizon trading opportunities that generate a profit. It is an extension of the trading

strategies of the existing empirical literature about persistent anomalies. Statistical arbitrage involves trading instruments that are different, but whose prices are mutually inconsistent according to historical price analysis. These price inconsistencies are identified by statistical techniques. A statistical arbitrage trading strategy then devises a set of trades on the inconsistently priced instruments and expects them to return to their historical relationship. Statistical arbitrage is not without risk; it depends heavily on the ability of market prices to return to a historical or predicted prices.

Pairs Trading: Suppose that an asset price process is governed by a mean reverting process given by

$$dX_t = \alpha(X - X_t)dt + \sigma(X_t)dW_t$$

Where $\sigma(X_t)$ is defined as a volatility function

Thus in the long run the asset has an above price of $\overline{X}$ if at time t, the price is above$\overline{X}$, then a trader could short this asset in the expectation that, if the price process continues to be governed by this mean reverting process with level $\overline{X}$ ‹ X_t, he or she will be able to sell and close the position at $\overline{X}$, thus making a profit. Conversely, if the current asset price, X_t satisfied X_t , <$\overline{X}$ then a trader could go long in the asset, expecting to close the position at X, again making a profit. A pair's trader looks for a pair of stocks whose price difference follows a mean reverting process. If one can find such a pair, then when the price of one of the two (say stock X) is high relative to the price of the other (say stock Y), then a pairs trader would short a suitable amount of stock X and go long in a corresponding amount of stock Y, when the prices of the two stocks go back into balance, then the trader close the position. This is a market neutral strategy. It depends upon prices reverting to their long run equilibrium values relative to each other. Although pairs trading sounds too good to be true, several issues need to be addressed:

- Which pair of stocks should be traded?
- How much do we buy/sell of each stock?
- How to catch the signal of an opportunity (i.e. opening a position)?
- When to close the position so that profit could be earned?
- When do we cut our losses when things go wrong?

2. Objective of the Study

It has been well documented that the trading strategies appropriate for an efficient stock market would be significantly different from the strategies appropriate for an efficient market. We have numerous empirical evidences of Indian Stock Market being an inefficient price setter. This calls for a trading strategy based on technical analysis which can outperform the market. Technical analysis is technical trading rules based on graphical patterns of the time series of assets prices. Momentum strategies, contrarian strategies and pairs trading are some techniques of technical analysis. Therefore, the objective of this paper is to identify arbitrage opportunities and calculating profits earned through these opportunities by using statistical tools.

3. Data, Data Sources and Statistical Methodology

It is a market neutral strategy; therefore, we have taken the thirty actively traded stocks of BSE Sensex. The price data under study consists of daily market-close prices from January 1, 2010 to January 1, 2011. Under the study each price series is transformed into its natural logarithm price series. In view of the inherent heteroscedasticity of the data it is considered advisable to transform it into log price. Log transformation is likely to render the price series to be homoscedastic and thereby make the series stationary.

We have followed the methodology suggested by Ngai Hang Chan in his book viz. “Time Series: Application to Finance with R and S-Plus”. First, we examined the historical paths of all the thirty stocks, and chose those pairs that move together most closely. For that we standardized all price series to put them on a common scale and computed certain “distances” between all pairs of the standardized series. Those pairs that have the smallest distances are the ones that have moved most closely together. Altogether, there were 252 trading days. We used the first half that is 126 days, to form the standardized pairs and trade in the second half of the data. Standardized series were formed by subtracting mean and dividing by standard deviation of first half series. The distance between two series is linearly related to their sample correlation coefficient and is minimized by the pair with the largest sample correlation coefficient. Among all possible pairs we considered 10 pairs with the smallest distance.

To deal with second problem as stated above, there are several strategies for choosing how much to invest. One popular strategy is to invest nothing. That is, purchase the same rupee amount of the long stock as you sell of the short stock so that the strategy is self-financing. Now

we start trading with these 10 pairs. For all the ten pairs we calculated mean and standard deviation of their price-ratios for the first half. For each trading day (second half period), we computed price- ratios after subtracting mean and dividing standard deviation of price-ratios of first half.

We open a position when the absolute value of the difference between the standardized series gets larger than some multiple of its historical standard deviation. Mostly traders use a multiple of two but we applied the strategy at multiple of one, two and three as well. The larger the multiple, the more rewarding and riskier it will be. Closing a position is an art. Again, there are various options to choose. We took the strategy to wait until the first time it crosses zero. This is an indication that whatever temporary anomaly separated the two stocks has now been forgotten. If any positions were open the last trading day of the year, they were closed regardless of profit or loss. To apply the above stated strategy we used R- Software and used the following codes:

```
daysf<-1
dayef<-126
dayst<-127
dayet<-252
npairs<-10
cutoff<-2

nstock<-dim(Data)[2]
Data.cor<-cor(log(Data[daysf:dayef,]))
pairm<-c(1:nstock)
for(i in 1:nstock)
pairm[i]<-sort.list(Data.cor[i,])[nstock-1]
pairmm<-t(apply(cbind(c(1:nstock),pairm),1,sort))
pairmu<-pairmm[!duplicated.data.frame(as.data.frame(pairmm)),]

napairs <- dim(pairmu)[1]
toppairs<-pairmu[sort.list(Data.cor[pairmu]),][(napairs-npairs+1):napairs,]
profit<-rep(0.0, npairs)
nopen<-rep(0, npairs)
nclose<-rep(0, npairs)
for(i in 1:npairs){
stock1<-Data[,toppairs[i,1]]
stock2<-Data[,toppairs[i,2]]
ratio.mean<-mean(log(stock1[daysf:dayef]/stock2[daysf:dayef]))
ratio.std<-sqrt(var(log(stock1[daysf:dayef]/stock2[daysf:dayef])))
ratio.trade<-(log(stock1[dayst:dayet]/stock2[dayst:dayet])-ratio.mean)/ratio.std
```

```
open.trade<-0
profit[i]<-0.0
for (j in 1:(dayet-dayst)){
if(open.trade!=0){
if(open.trade*ratio.trade[j]<=0.0){
open.trade<-0
profit[i]<-profit[i]+n1*stock1[j+dayst]+n2*stock2[j+dayst]
nclose[i]<-nclose[i]+1
}
}else{
if(abs(ratio.trade[j])>=cutoff){
open.trade<-sign(ratio.trade[j])
n1<-(-open.trade)/stock1[j+dayst]
n2<-open.trade/stock2[j+dayst]
nopen[i]<-nopen[i]+1
}
}
}
if(open.trade!=0){
profit[i]<-profit[i]+n1*stock1[dayet]+n2*stock2[dayet]
}
}
result<-cbind(names(Data[toppairs[,1]]),names(Data[toppairs[,2]]), nopen, nclose, profit)
sum(profit)
sum(nopen)
sum(nclose)
result
```

4. Results and Discussion

The clear stated objective of this paper is to show the execution of simple pairs strategy through the use of R-Software. The basic premise of this strategy is that stock prices follow a mean reverting process. This strategy will no longer work, however, if the stock prices go out of order due to fundamental change in the relationship between the two stocks. We formed ten pairs to execute the pairs trading. The results of the above stated simple pairs strategy are shown in table1, Table2, Table3. These results are attained after applying cut-off points at 1, 2 and 3 standard deviation. Different multiples of standard deviation depicts different nature of pairs traders. Trading beyond cut-off point one will have more trading opportunities, but less rewarding and less risky as compare to two and three. Our results show maximum number of opening and closing positions and minimum income (negative profit) with one standard deviation. Most of the researches use two as cut-off point. Here we can see that trading beyond

two has ten opening and two closing position and moderate income (lesser negative profit as compare to one). Trading beyond three also has same number of opening and closing positions but generating positive profit. No arbitrage strategy can assure positive return, the element of risk is always there.

Table1: Positions and Profit/Loss for 10 Traded Pairs

S.No.	Pairs		Open	Close	Profits
1	"st7"	"st12"	2	2	0.01736
2	"st9"	"st11"	2	1	-0.00576
3	"st14"	"st28"	1	0	-0.03869
4	"st25"	"st30"	1	0	-0.00546
5	"st1"	"st13"	1	0	-0.03369
6	"st9.1"	"st15"	3	2	-0.02801
7	"st7.1"	"st8"	1	0	-0.01623
8	"st14.1"	"st29"	1	0	-0.06720
9	"st5"	"st19"	2	2	0.02585
10	"st6"	"st7"	2	1	0.02007
		Total	**16**	**8**	**-0.13177**

Table2: Positions and Profit/Loss for 10 Traded Pairs

S.No.	Pairs		Open	Close	Profits
1	"st7"	"st12"	1	1	0.01395
2	"st9"	"st11"	1	0	-0.02105
3	"st14"	"st28"	1	0	-0.02914
4	"st25"	"st30"	1	0	-0.00546
5	"st1"	"st13"	1	0	0.09560
6	"st9.1"	"st15"	1	0	-0.05262
7	"st7.1"	"st8"	1	0	-0.00839
8	"st14.1"	"st29"	1	0	-0.06228
9	"st5"	"st19"	1	1	0.02460
10	"st6"	"st7"	1	0	0.01238
		Total	**10**	**2**	**-0.03242**

Table3: Positions and Profit/Loss for 10 Traded Pairs

S.No.	Pairs		Open	Close	Profits
1	"st7"	"st12"	1	1	0.02362
2	"st9"	"st11"	1	0	-0.00763
3	"st14"	"st28"	1	0	-0.00946
4	"st25"	"st30"	1	0	-0.00546
5	"st1"	"st13"	1	0	0.09560

6	"st9.1"	"st15"	1	0	-0.04233
7	"st7.1"	"st8"	1	0	-0.00839
8	"st14.1"	"st29"	1	0	-0.04793
9	"st5"	"st19"	1	1	0.02856
10	"st6"	"st7"	1	0	0.01238
		Total	**10**	**2**	**0.03895**

There could be various finer points to this simple strategy but that are beyond the scope of this article. Beta is another tool which could be used to form pairs. Beta is used to determine how many share of each stock to execute for pairs trade. Because beta measures the magnitude of the relationship between two stocks, we can apply beta to the delta of the position to determine the quantity for each stock in the pair. Moreover, we can get better results by using cointegrated pairs. Co integration is a powerful tool which refers not to co-movements in returns, but to co-movement in asset prices. While the individual series may themselves be non-stationary, we say that they are cointegrated if we can find a linear combination that is stationary.

We can also choose between stocks, options, or stock plus options. Each has its own advantages and disadvantages. Stocks are relatively easy to execute in actively traded stocks, but have virtually unlimited risk if you are wrong. That is, if you expect that a spread between two stocks will revert to a mean, but if it does not, you can lose a lot of money on both the long and short positions of pairs trade. Options are good vehicle for pairs trading, and can simply be used as stock substitutes: long calls for long stock, long puts for short stock. Options have limited risk, but can be tougher to execute quickly. Within option, various spreads like vertical, horizontal and diagonal could be executed. Pairs trading require patience in acquiring the knowledge, finding and analysing suitable opportunities, and executing the trade. Pairs trading can have a wide range of risk/reward possibilities.

References

Alexander, S. S., 1961, "Price Movements in Speculative Markets: Trends or Random Walks", *Industrial Management Review* II, May, 7-26.

Bombay Stock Exchange Website; *www.bseindia.com*

Chan, Ngai Hang, 2010, *Time Series: Application to Finance with R and S- Plus*, Wiley Publication

Chaudhuri, S. K., 1991b, "Seasonality in Share Returns: Preliminary Evidences on Day-of-the-Week Effect", *Chartered Accountant*, November, Vol. 40, No. 4, 407.

Chowdhury, S.S.H., Sadique, M.S. and Akhter, M.S., 2002, "The Size Effect on the Speed of Stock Price Adjustment to Information in the Indian Stock Market", *Asian Economic Review*, 111-123.

Conrad, J. and Kaul, G., 1988, "Time-Variance in Expected Returns", *Journal of Business*, 61(4), 409-425.

Engle, R. F., and Granger, C. W. J, 1987, "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica* 55,251-276

Fama, Eugene F., 1965, "The Behaviour of Stock-Market Prices", *The Journal of Business*, January, Vol. XXXVIII, 34-105.

Fama, Eugene F., 1970, "Efficient Capital Markets: A Review of Theory and Empirical Work", *The Journal of Finance*, May, Vol. 25, 383-417.

Fogler, H. Russell, 1978, *Analyzing the Stock Market: Statistical Evidence and Methodology*, Grid Inc., Columbus, Ohio.

Gatev, E., Goetzmann, W., and Rouwehorst K. G.,2006, "Pairs Trading: Performance of a relative value arbitrage rule, *Review of Financial Studies* 19, 797-827

George, T. J. and Hwang, C. Y., 2004, "The 52- Week High and Momentum Investing, *Journal of Finance* 59, 2145-2176

Hogan, S., Jarrow, R., Teo, M., and Warachka, M., 2004, "Testing Statistical Arbitrage with Application to Momentum and Value Strategies, *Journal of Financial Economics* 73, 525-565

Jegadeesh, Narasimhan and Titman, Sheridan, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", *The Journal of Finance*, March, Vol. XLVIII, No. 1, 65-91.

Kendall, M. G., 1953, "The Analysis of Economics Time Series", *Journal of the Royal Statistical Society* (Series A), XCVI, 11-25.

Lo, A. and Mackinlay, C., 1988, "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test", *Review of Financial Studies* 1, 41-66.

Poterba, James M. and Summers, Lawrence H., 1988, "Mean Reversion in Stock Prices", *The Journal of Financial Economics*, 22, 27-59.

Seth, A. K., and Gupta, Saloni, 2005, "Understanding Volatility at BSE: A Quality Control Approach (Implications for Investors), *Decision*, Vol. 32, No.1, Jan-June.

Summers, Lawrence H., 1986, "Does the Stock Market Rationally Reflect Fundamental Values?", *The Journal of Finance*, July, Vol.41, No.3, 591-601