

Analysis of Statistical Arbitrage using Machine Learning Techniques in Indian Stock Markets

A.U.S.S Pradeep (deepu@iitk.ac.in), Soren Goyal (soren@iitk.ac.in)

Advisor: Dr. Amitabha Mukerjee

Department of Computer Science and Engineering,

IIT Kanpur ,India

March 10,2013

1. Objective:

The aim of the project is to analyze Arbitrage opportunities arising in the Indian stock markets modeled on the set of previous historical data using different learning techniques (like Time delay neural networks and linear regression etc.)

2. Introduction:

Before we describe the problem precisely, some background discussion about statistical arbitrage is necessary. “Statistical arbitrage refers to attempting to profit from pricing inefficiencies identified through mathematical models” [1].

The basic assumption is that prices will move towards a historical average. A portfolio is said to be an arbitrage if it costs nothing to implement, has a positive probability of a positive payoff, and a zero probability of a negative payoff. Loosely speaking, “buy low” and “sell high” trade. [2]

But statistical arbitrage is not risk-free, it has a positive expected payoff and a zero probability of a negative payoff **ONLY** as time approaches infinity, and it’s variance vanishes at time infinity.

3. Motivation:

Arbitrage has the effect of causing prices in different markets to converge. [3] “The speed at which the convergence process occurs usually gives us a measure of the market efficiency”.

Hence a thorough analysis of statistical arbitrage opportunities using the advanced learning techniques is essential in mapping the efficiency of current day Indian market.

4. Work done previously:

Over more than half a century, much empirical research was done on testing the market efficiency, which can be traced to 1930’s by Alfred Cowles, Many studies have found that stock prices are at least partially predictable. The method to test the existence of statistical arbitrage was finally described in the paper “Statistical arbitrage and tests of market efficiency” [4] by, S.Horgan, R.Jarrow, and M. Warachka published in 2002.And an improvement on the paper “An Improved test for Statistical arbitrage”[5] was published in 2011 by the same team which forms the basis for this project.

And further a paper “Machine Learning in Statistical Arbitrage” [1] (2009) tries to implement an approach of Support Vector Regression (SVR) and Principal component Analysis (PCA) to devise a trading strategy to utilize the arbitrage opportunity the iShares FTSE/MACQ traded in the London Stock Exchange Market.

Also a paper “Statistical Arbitrage Stock Trading using Time Delay Neural Networks”(2004)[6] attempts to solve the problem whether TDNN architecture trained on the past history of stock data (NASDAQ dataset from 1975-2000)can accurately predict when to buy a stock.

5. Dataset:

The dataset that we would be using would consist of data from [Bombay Stock Exchange\(BSE\)](#) and [National Stock Exchange of India\(NSE\)](#) given by Yahoo finance, historical prices for the past 10 years. This dataset would be used in both training and testing purposes.

6. References:

[1] “Machine Learning in Statistical Arbitrage” published by Xing Fu, Avinash Patra. (December,2009)

[2] “A Statistical Arbitrage Strategy” a master thesis project by Kun Zhu, Royal Institute of Technology, Stockholm, Sweden. (2005)

[3]Article on Arbitrage on Wikipedia <http://en.wikipedia.org/wiki/Arbitrage>

[4] “Statistical arbitrage and tests of market efficiency” published by S.Horgan, R.Jarrow, and M. Warachka (2002).

[5] “An Improved test for Statistical arbitrage” published by Robert Jarrow, Melvyn Teo, Yiu Kuen Tse, Mitch Warachka (2011).

[6] “Statistical Arbitrage Stock Trading using Time Delay Neural Networks” a Machine learning final year project by Chris Pennock (Fall 2004).