An Evaluation and Implementation of Pairs Trading in Cointegrated Indian IT Stocks

Anirudh Jayaraman[1]

[1]M.Sc Student at Indira Gandhi Institute of Development Research (IGIDR)
Goregaon (East), Mumbai 400 065

anirudhjigidr.ac.in

Abstract. *This term-paper aims to test the efficacy of pairs-trading, a strategy to obtain profits from a market neutral approach on a co-integrated pair of stocks. While there are various proprietary approaches to implementing pairs trading making literature on this subject rare, the focus of this paper is to showcase the applicability of this strategy using freely available data and open source software.*

1. Introduction

Pairs trading is a very well known market-neutral Wall Street strategy of using statistical arbitrage to make profits. This method was pioneered by Gerald Bamberger and the quant group at Morgan Stanley in the 80s. It brought forth so to speak, the birth of statistical arbitrage.

At the heart of pairs trading lies the selection a pair of securities that are co-integrated so that due to mean reversion of the co-moving pair, any divergence beyond a threshold can be profited from in the long run by taking opposite positions on the 2 securities. Intuitively, one can imagine this strategy to work for more than 2 securities as well, since the principle of co-integration can still be used to conduct arbitrage.

This paper therefore constructs a portfolio of 2 securities such that the spread is an I(0) process. To make the most of pairs-trading, the efficacy of such a strategy hinges heavily on modeling this I(0) series. However, as a first-cut implementation of this technique and for the sake of code-simplicity, modeling the I(0) process has been avoided in order to go after the low-hanging fruit instead. The caveat being, since literature on pairs trading is of proprietary nature and strategies are mostly applicable to specific cases only, there isn't much theoretical verification.

Pairs trading demands good position sizing, market timing, and decision making. In spite of the absence of much of a downside risk, there are scarce opportunities, and for profiting, the trader must be one of the first to capitalize on such an opportunity.

2. Literature Review

The most cited and referenced work on pairs trading literature comes from **Gatev, Goetzmann and Rouwenhorst (1999)**, who uses a simple standard deviation approach

to show profitability of pairs trading - formally laying the foundation for academic literature on the subject.

Vidyamurthy (2004) offers a detailed implementation strategy using the co-integration framework although without empirical results.

Elliott, van der Hoek and Malcolm (2005) apply a Kalman filter to estimate a parametric model of the spread.

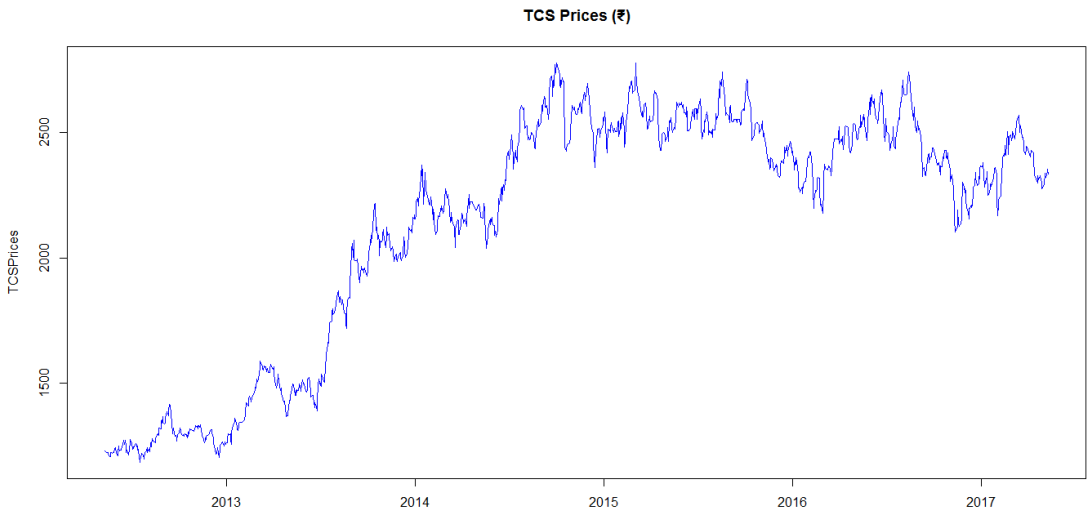
Binh Do, Robert Faff and Kais Hamza (2006) use a general approach to model relative mispricing with reference to mainstream asset pricing theory. They use estimation techniques using state space formulation with Expectation Maximization.

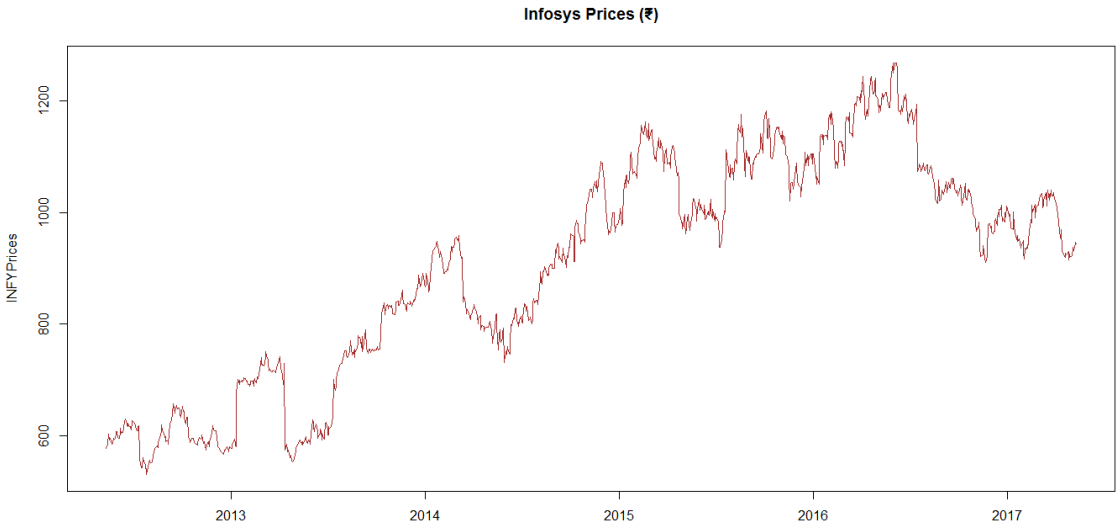
3. Data and Methodology

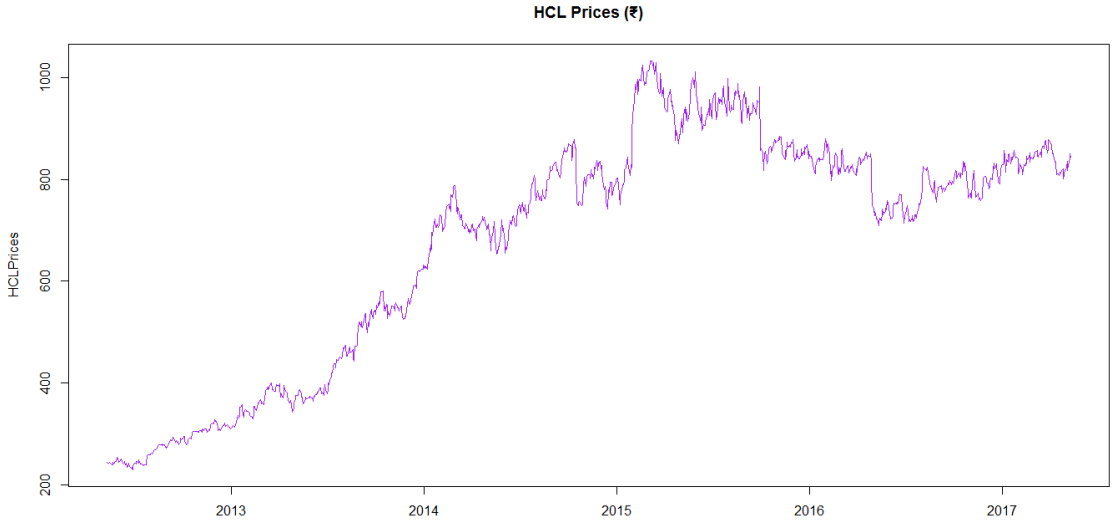
The data used in this study to test pairs trading include 3 prominent IT stocks that prima facie are likely co-move since their performance depends heavily on the USD-INR exchange rate, a favorable import-export environment besides similar market microstructure. The companies under consideration are TCS, Infosys and HCL, which happen to be similar in their form and function. In addition, these stocks have historically shown high correlation. Correlated series aren't necessarily cointegrated, but that shall be put to test here.

3.1. Data

5 years' data from 11 May, 2012 to 10 May, 2017 for TCS, Infosys and HCL have been sourced from Yahoo Finance's repository of historical data.

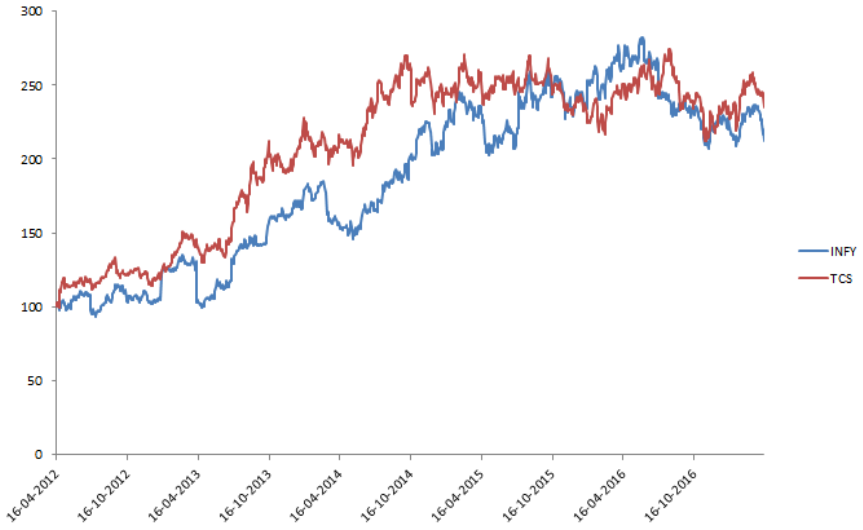




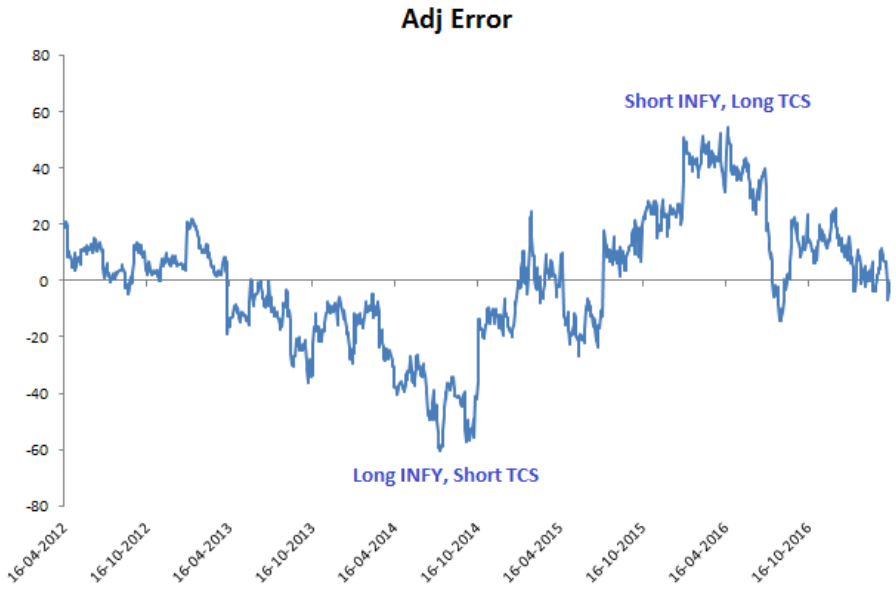


3.2. Methodology

A cursory look at the TCS and Infosys stock prices over 5 years with each stock's starting price re-based at 100 units indicates co-movement:



The mean adjusted difference between the two stocks turns out to be mean-reverting, hinting at possible cointegration (which needs to be tested) and a possible trading strategy may look something like the following:



3.3. Pairs Trading based on Cointegration

Stock prices are assumed to follow a Geometric Brownian Motion (GBM).

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$$

Here S_t is the stock price, μ is average return, σ is volatility of returns and W_t is the standard Brownian Motion process.
From Ito's Lemma, $\log(S_t)$ follows a Brownian Motion:

$$d\log(S_t) = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t$$

Using the fact that the **discrete form of Brownian Motion is a Random Walk**, logarithm of prices are modeled as a Random Walk, so that the idea of cointegration can be readily applied.
The spread δ is therefore modeled as:

$$\delta = \log(Y_t) - \alpha - \beta \log(X_t)$$

where X_t and Y_t are the stock price pair, β is the cointegrating vector (henceforth to be referred to as the **hedge ratio**), while α will be referred to as the **premium**. Once cointegration has been established, in theory, the hedge ratio is expected to stay constant, but in practice, it is practical for the trader to model the spread on a moving-window basis, which implies constantly updating the hedge ratio using trader-specified number of lags (taken to be *180 days* as shown in the code provided in the *Appendix*). Note that the results show that for a cointegrated series, the evolution of the hedge ratio is such that it stays within a narrow band, if not exactly constant throughout.

4. Results and Inferences

The 3 chosen IT stocks, namely, TCS, Infosys and HCL are highly correlated. What needs to be tested next, is whether they are cointegrated in pairs.

Correlation Matrix

ρ	TCS	Infosys	HCL
TCS	1.00	0.90	0.96
Infosys	0.90	1.00	0.89
HCL	0.96	0.89	1.00

Results indicate strong evidence for cointegration between HCL and TCS stocks, moderate cointegration between TCS and INFY stocks and no cointegration between HCL and INFY at the 10% levels of significance upon testing for stationarity of the spread with both Phillips Perron and Augmented Dickey Fuller tests.

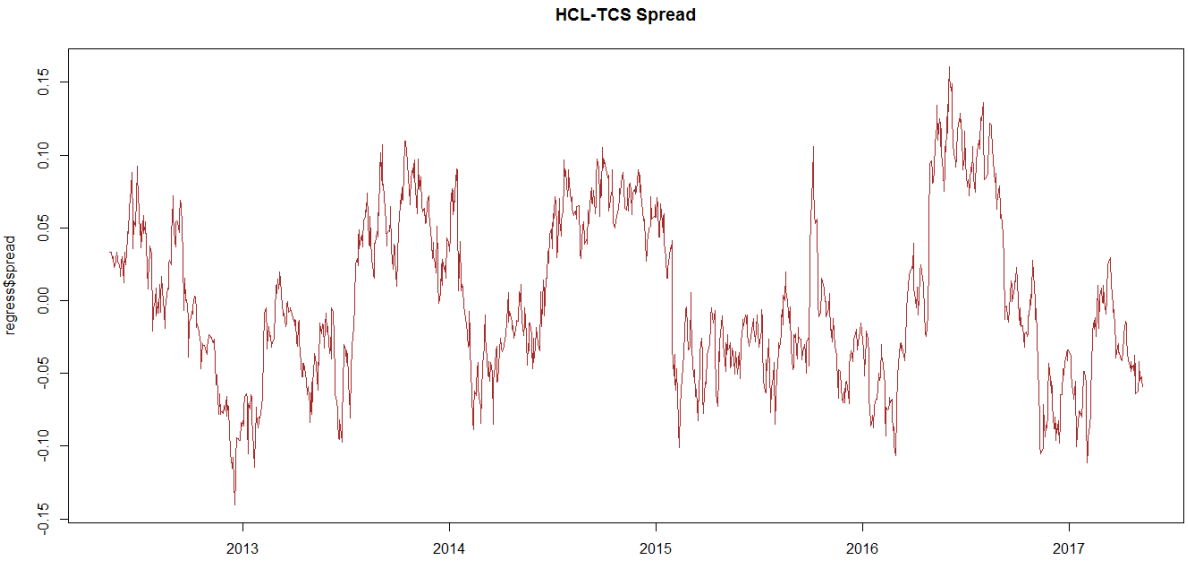
Tests for Stationarity of the Spread (10% Level)

	PP Test	ADF Test
HCL - TCS	TRUE	TRUE
TCS - INFY	TRUE	FALSE
HCL - INFY	FALSE	FALSE

Hence, we have three possible scenarios on the level of cointegration between stocks and the cumulative returns when following pairs trading strategy given each scenario. Given below are the results for the 3 pairs.

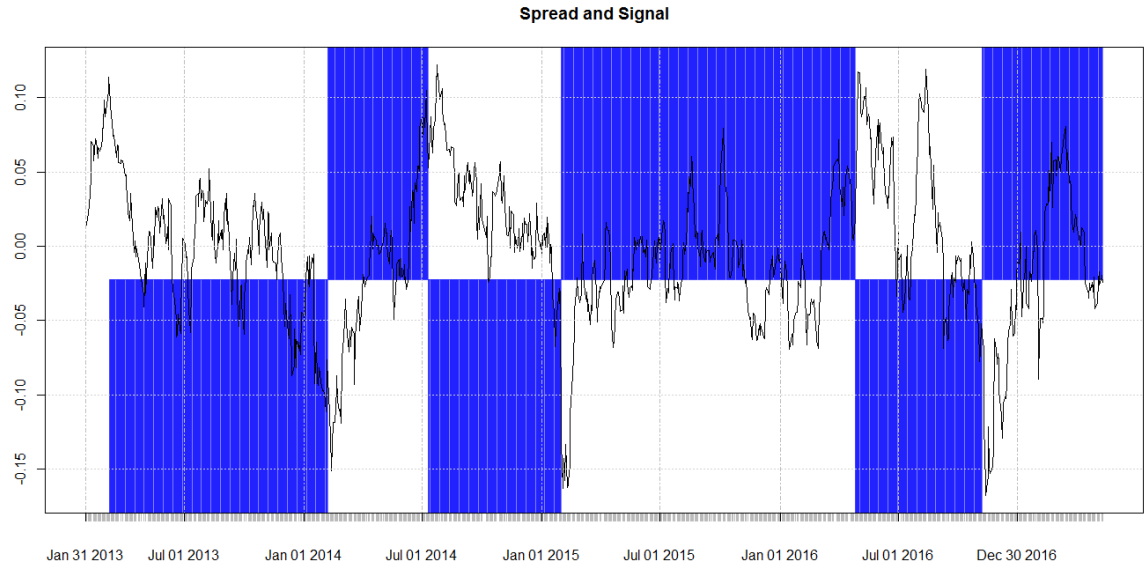
4.1. TCS-HCL (Strongly Cointegrated)

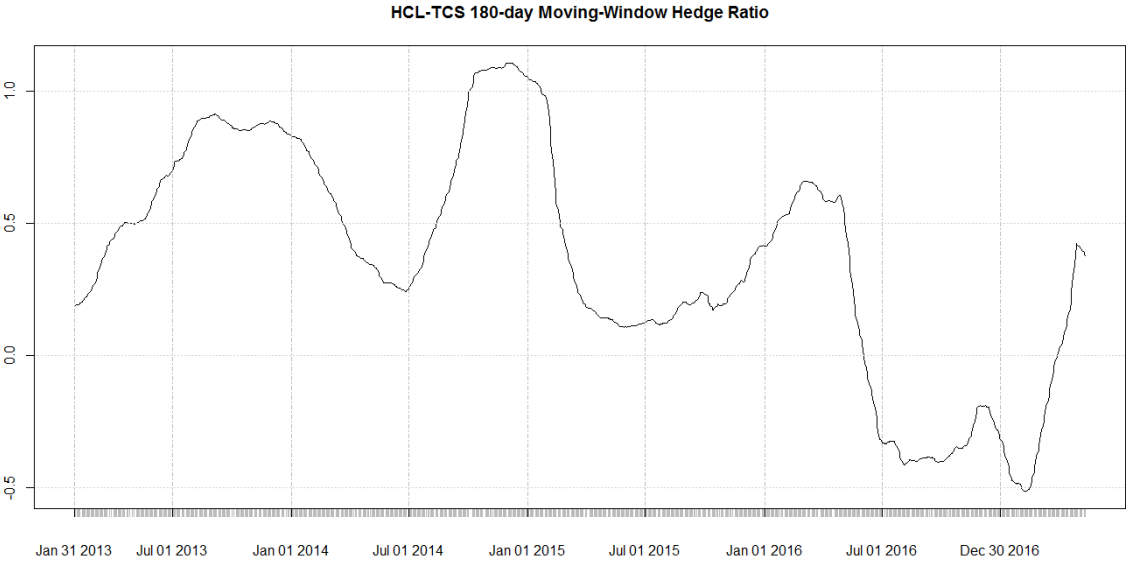
The TCS-HCL spread is shown to be I(0) at approx. 2% level of significance according to the Phillips Perron test and approx. 5% level of significance according to ADF test.



A divergence threshold is specified according to which opposite positions are taken in the divergent stocks. The intuitive threshold-rule used for the stocks throughout this analysis is that the threshold should be as close to the minimum of the maximum positive and

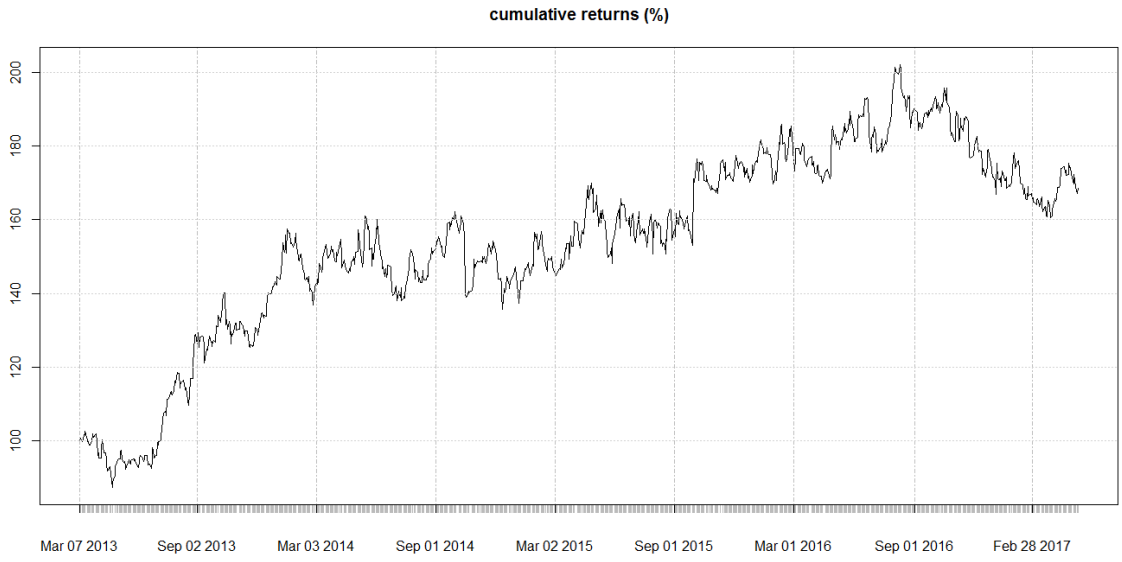
maximum negative divergences from mean. As mentioned earlier, the threshold rule can be modeled in sophisticated ways based on the specific process that fits the I(0) spread. The threshold rule used here on the other hand is basic and is *not* on a moving-window basis. The Signal-Spread for trade execution based on the trade-trigger rule is as follows:





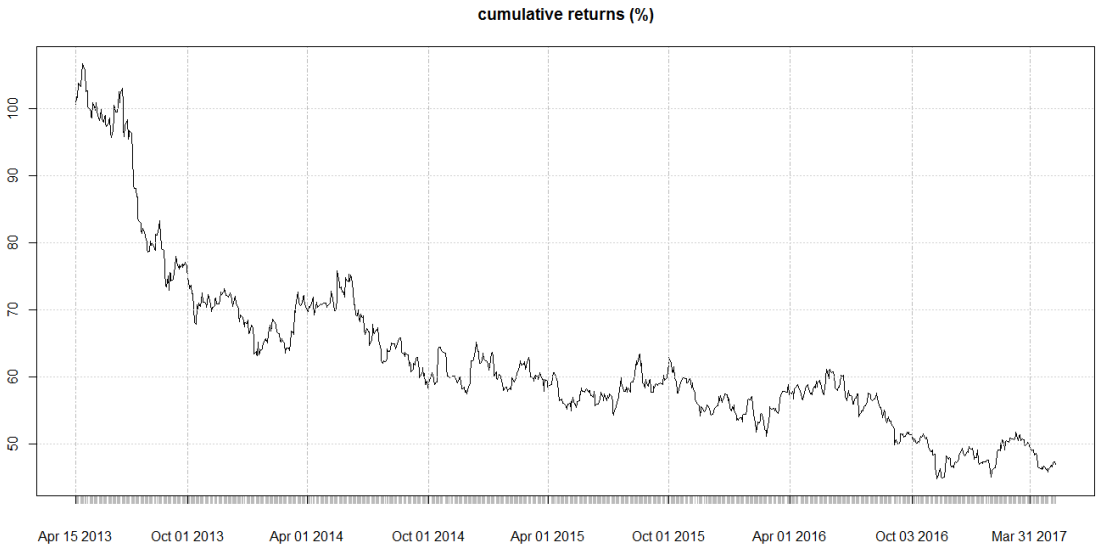
The evolution of the cumulative returns[1] over time based on the employed pairs trading strategy is the last point of interest, and is shown in the figure below. It amounts to 68% cumulative returns spanning over 5 years. Also, the probability of beating 20% returns exceeds 80%.

[1]The workhorse for this analysis consists of 7 functions written in the R programming language - *Signal*, *ParameterEstimates*, *ParameterEstimatesHistorical*, *Stationarity*, *Weights*, *Returns* and *.returns*. These have been included in the Appendix.



4.2. TCS-INFY (Moderately Cointegrated)

Since the spread[2] between TCS and Infosys fails to reject the null of unit root according to the ADF test and is found to be I(0) only according to the PP test, that too at a level of significance of 10%, the pairs-trading strategy is not expected to work. The cumulative returns are along expected lines and are shown below. Clearly, pairs-trading is ineffective when a pair of stocks show little or no cointegration.



4.3. HCL-INFY (No Cointegration)

The null hypothesis of unit root in the spread[3] is failed to be rejected for HCL-INFY pair even at the 10% level of significance with both PP and ADF tests. Therefore pairs trading isn't applicable here. For this case too, the cumulative returns over 5 years (code in the

[2]Signal and Spread for TCS-INFY pair shown in Appendix
[3]Signal and Spread for HCL-INFY pair shown in Appendix

Appendix) turn out to be negative if the strategy is employed[4].

4.4. Conclusion

This paper tests the famous pairs trading strategy on 3 IT stocks that are first tested for pair-wise cointegration, after which a pairs trading strategy is formulated based on the spread between each pair of stocks. It can be seen from these examples that pairs trading is ineffective for stocks that show little or no cointegration and have the potential to yield significant returns for a strongly cointegrated pair of stocks.

This idea has been heavily tested, but due to its proprietary nature, it is hard to find literature demonstrating this strategy in the real world. This paper attempts to make the most of the principle of pairs trading by employing a simple trading rule to evaluate the strategy.

This analysis however does not factor in transaction-costs adjusted returns and uses a simple threshold-based signalling rule to position trades. More sophisticated models can be coded up and tested to better simulate the stock-specific spreads and associated trading rules.

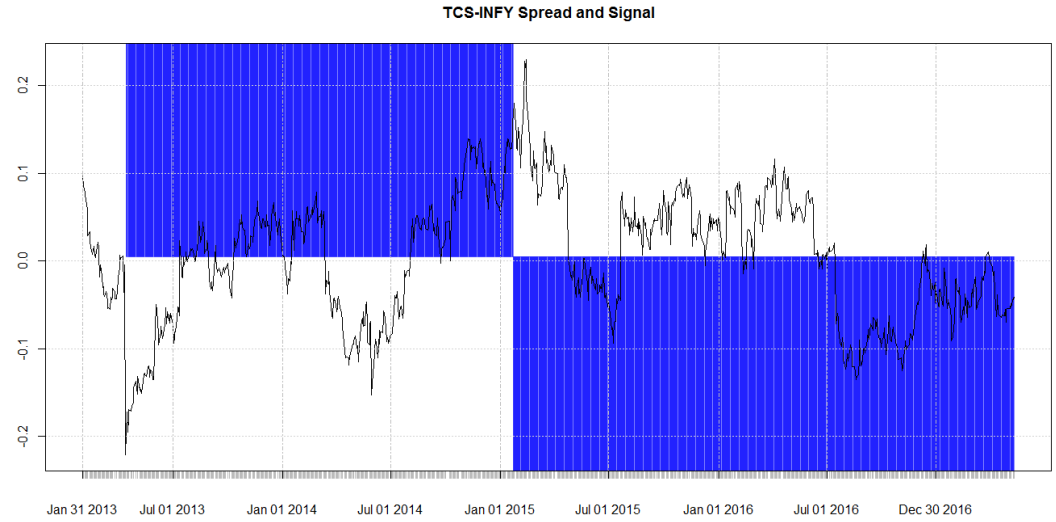
The caveat being that pairs trading requires position sizing and good market timing where opportunities are scarce and traders must compete to be the proverbial early bird.

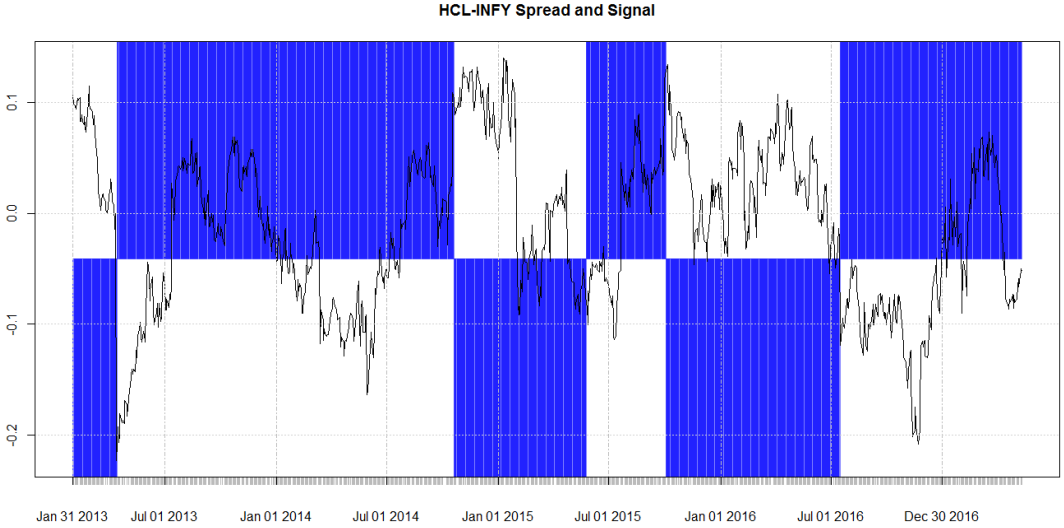
5. References

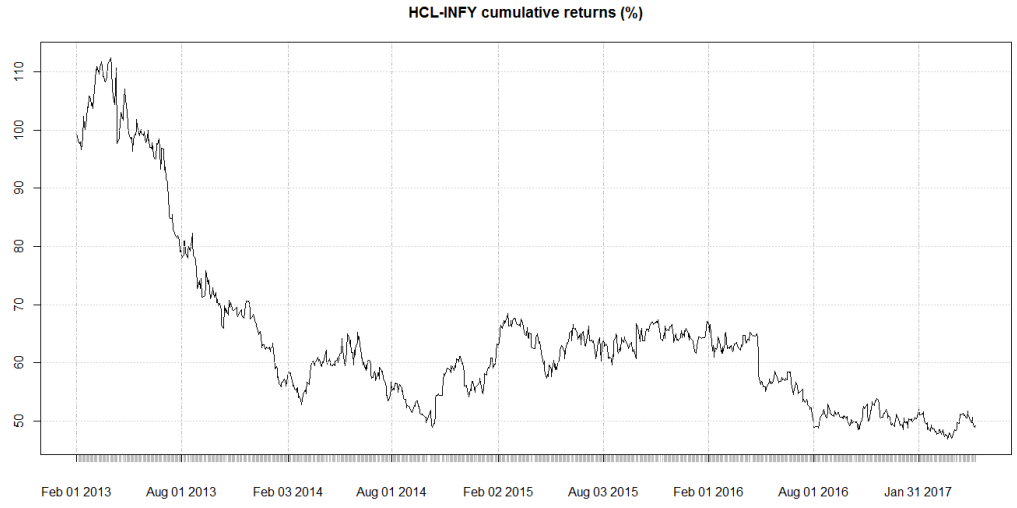
- **Gatev, Goetzmann and Rouwenhorst (1999)**
- **Vidyamurthy (2004)**
- **Eliott, van der Hoek, Malcolm (2005)**
- **Binh Do, Robert Faff, Kais Hamza (2006)**
- **Shinichi Takayanagi, Kohta Ishikawa**, "Introduction to Pair Trading - based on Cointegration"
- **Andrew Neil Burgess**, "A Computational Methodology for Modeling the Dynamics of Statistical Arbitrage"

[4]Evolution of Cumulative Returns for HCL-INFY pair shown in Appendix

6. Appendix


TCS-INFY Spread and Signal
0.2
0.1
0.0
-0.1
-0.2
Jan 31 2013
Jul 01 2013
Jan 01 2014
Jul 01 2014
Jan 01 2015
Jul 01 2015
Jan 01 2016
Jul 01 2016
Dec 30 2016


HCL-INFY Spread and Signal
0.1
0.0
-0.1
-0.2
Jan 31 2013
Jul 01 2013
Jan 01 2014
Jul 01 2014
Jan 01 2015
Jul 01 2015
Jan 01 2016
Jul 01 2016
Dec 30 2016


HCL-INFY cumulative returns (%)
110
100
90
80
70
60
50
Feb 01 2013
Aug 01 2013
Feb 03 2014
Aug 01 2014
Feb 02 2015
Aug 03 2015
Feb 01 2016
Aug 01 2016
Jan 31 2017