NUS INVESTMENT SOCIETY

Academic Research Review

An Introduction and Implementation of a Simple Pairs Trading Strategy

EXECUTIVE SUMMARY

Statistical arbitrage is a common trading strategy used by many hedge funds and trading houses in financial markets in an attempt to profit through the corrections of the market on mispriced assets. The strategy works by utilizing computer-intensive algorithms to seek out highly correlated assets/portfolios, and subsequently executing simultaneous trades in opposite assets/portfolios when their price behavior deviates from their expected differences.

In this short paper, we look at how pairs trading, one of the simplest form of statistical arbitrage strategies, works and how it can be implemented a few sets of intuitive pairs of assets. Towards the end of the paper, we evaluate the results we have gotten and state further improvements and extensions to the current method.

INTRODUCTION AND MODEL

As the name suggests, pairs trading strategies are executed on a **pair** of assets, with an expected relationship (usually measured by their correlation coefficient) on their price or returns. When this relationship starts to deviate from their expected relation to a significant degree, two simultaneous trades are executed in anticipation that the underlying price movements will act in a certain way to bring the relationship back to the mean.

Thus, the model is built with the following assumptions and structure:

- Means Reversion: The difference in returns of a pair of assets is assumed to exhibit a mean-reverting behavior relative to their historic mean difference in returns.
- Technical/Filter Indicators: The significance of deviation from the mean is measured by the standard deviation of the returns difference from its mean.

For the strategy presented in this paper, we will be using the **difference in daily log prices** between two assets as the basis.

METHODOLOGY

We derived the return difference for the data in 2007 as *a priori* information to act as our running mean of difference in log returns. $P_{x,n}$ represents the nominal price of asset x at time n. D_n, the difference in log returns, is obtained by normalizing the individual returns:

$$D_n = log\left(\frac{P_{1,n}}{P_{1,n-1}}\right) - log\left(\frac{P_{2,n}}{P_{2,n-1}}\right)$$

The running mean is calculated as the simple average of the last 1 year of returns difference:

$$E(D_{N+1}) = \frac{1}{N}\sum_{i=N-k}^{N} D_i$$

For the purpose of demonstrating the performance of our pairs trading strategy, we have selected 3 different pairs of assets which makes intuitive sense in terms of their correlation:

- **Brent vs. CAD** : As a major exporter of crude oil, we expect that the strength of the Canadian dollar to have a degree of correlation to the price of Brent Crude, the international benchmark for crude oil.
- **PEPSI vs. COCA-COLA** : We expect a degree of correlation in the stock prices of companies in the same industry and selling the same type of products due to the same type of systemic risks they experience.
- **SHELL Type A (NYSE) vs. SHELL Type B (LSE)**: An example of a dual-listed stock, where the stocks of a firm is listed on two or more different exchanges. As both stocks have the same underlying company, we expect changes in prices between the two stocks to be highly correlated.

TRADING STRATEGY

At the start of each trading day, the running mean of the difference in log prices is updated. If there are no existing positions on a specific pair of assets, a check is done on the current returns difference to determine if it has deviate sufficiently (i.e. examining the standard deviation from the running mean) to validate a trade execution.

Trade execution is conducted by dividing available capital evenly between the two assets. If $D_n > E(D_n)$, we place a long position in the first asset and a short position in the second asset. If $D_n < E(D_n)$, we reverse the positions and place a short position in the first asset and a long position in the second asset.

If there are existing positions on the pair, a check is done on the current returns difference to determine whether it has returned sufficiently to the mean to warrant a **take-profit** execution, or conversely moved too far away from the mean to warrant a **stop-loss** execution. The exact parameters for each pair is optimized to accommodate the degree of mean-reversion behavior exhibited.

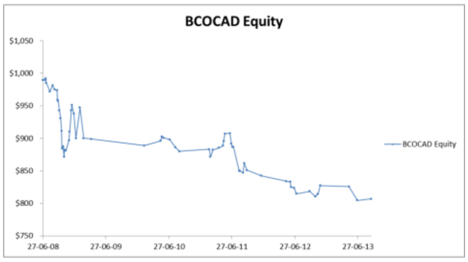
BACK-TESTING RESULTS

The results of back-testing is displayed below in terms of the cumulative equity value over the testing period (2008 – 2013), with key statistics displayed on the left side of each curve. Detailed portfolio statistics are presented in the appended table.

-19.3%
cumulative return

-4.2%
annualised return

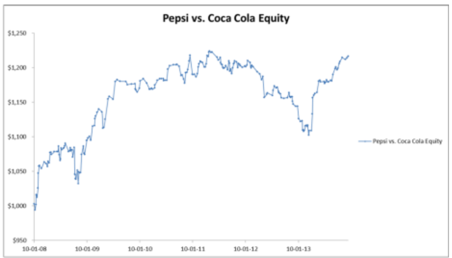
-0.06
Correlation



21.67%
cumulative return

4.00%
annualised return

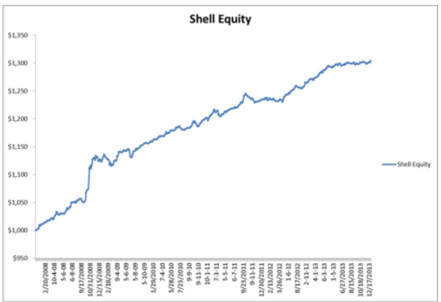
0.69
Correlation



30.4%
cumulative return

5.45%
annualised return

0.98
Correlation



Assets	Brent Crude vs. CAD	Pepsi vs. Coca Cola	Shell Type A vs. Type B
Correlation	-0.06	0.69	0.98
Annualized Return	-4.2%	5.45%	4.00%
Win Rate	38.10%	56.12%	64.06%
Average Trade Return	-0.33%	0.09%	0.07%
Max consec. Win	6	9	11
Max consec. loss	8	6	7
Max Trade Profit	5.22%	2.62%	3.42%
Max Trade Loss	-4.97%	-3.56%	-0.93%
Max DD	-10.19%	-4.84%	-3.80%
Avg Trade Length	1.08 days	1.66 days	1.81 days

CONCLUSION

As can be observed, a high correlation significantly affects how well the strategy does, regardless of our intuitive perception of the relationship between two assets. Examining the **Pepsi vs. Coca-Cola** strategy, we notice a dip in trading performance in the late 2013s. Upon further investigation, we found that the correlation between Pepsi (PEP) and Coca-Cola (KO) has dipped during that period, further strengthening our basis assumption that correlation is key to success in this strategy.

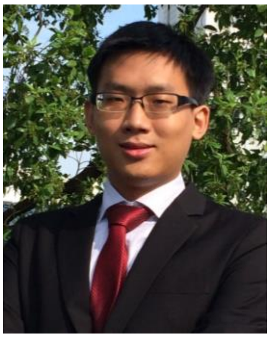
Improvement and extensions

With the increase in market efficiencies in the financial markets, statistical arbitrage strategies such as the pairs trading strategy demonstrated here can arguably perform better if we reduce the timeframe of the execution (i.e. instead of running the algorithm on a daily basis, we move towards tighter timeframes such as hourly or even per-minute trades). This increases the chances for our algorithm to catch potential mispricing, but also increases the model risks.

In the example here, we pre-select the asset pairs based on qualitative understanding. However, in more sophiscated statistical arbitrage trading models, a very important step is to not fixate on specific asset pairs, but rather periodically filter through an asset universe for suitable pairs. This prevents the algorithm from trading on asset pairs whose correlations may have deteriorate due to market factors, as well as improve the chances of trade entries.

RESEARCH ANALYSTS

Jonathan Yeo Tun Seng



Tseng Shih Ying

Wang Huang

Xie Lefan