# Statistical Arbitrage in South Africa

Jean-Jacques Duyvené de Wit



A minor dissertation submitted to the faculty of Commerce

in partial fulfilment of the requirements for the Master's degree in Mathematical Finance

at the University of Cape Town.

Supervisor: Kevin Kotzé

August 2013

# Table of Contents

## Acknowledgements

I would like to extend a special word of appreciation to my supervisor, Kevin Kotzé. You pushed me when my legs were tired. It was worth it.

I would also like to thank my manager, Albert Geyser. Thank you for your understanding and flexibility around the pressures that a project such as this brings.

Finally, I would like to thank Hennie Bezuidenhout for helping to proof read this document. All errors and omissions are my own.

## Declaration

I declare that the dissertation hereby submitted for the Masters in Mathematical Finance at the University of Cape Town is my own independent work and that I have not previously submitted this work for a qualification at another university / in another faculty. Each significant contribution to this report from the works of others has been attributed and has been cited and referenced.

_________________________ August 2013

Jean-Jacques Duyvené de Wit

# Executive Summary

This study investigates the performance of a statistical arbitrage portfolio in the South African equity markets. A portfolio of liquid stock pairs that exhibit cointegration is traded for a ten year period between the years 2003 and 2013. Without transaction costs, the portfolio has an encouraging Sharpe ratio of 2.1. When realistic transaction costs are factored in, the Sharpe ratio drops to 0.43. The results underline the theoretical profitability of statistical arbitrage as a trading strategy and highlight the importance of transaction costs in a real-world setting.

# Chapter 1: Introduction

## 1.1 Background

Fundamentally, a pairs trading strategy seeks to identify two financial instruments that exhibit similar historical price behaviour. Upon divergence of the two price series, the pairs trading strategy takes positions in the two instruments, betting on their subsequent convergence (Pole, 2007).

Pairs trading is the evolutionary predecessor of statistical arbitrage: statistical arbitrage can be viewed as the extension of pairs trading over a larger number of financial instruments.

The concept of statistical arbitrage covers a large collection of trading strategies. These strategies have a number of common features: (i) the strategies are systematic, or rules based, (ii) the trading portfolio has zero beta with the market, it is essentially market neutral, and (iii) the method used to generate excess returns relies on statistical practices (Avellaneda and Lee, 2010).

Statistical arbitrage is thus a market neutral trading strategy that attempts to profit from short term market dislocation. In contrast with deterministic arbitrage, the profitability of statistical arbitrage relies on the behaviour of favourable bets in expectation – it should profit in the long run after a large number of trades. Statistical arbitrage exploits diversification across a large number of assets to produce an investment strategy that has low volatility and is uncorrelated with the market (Avellaneda and Lee, 2010), since a market-neutral portfolio will be affected only by idiosyncratic returns.

## 1.2 Problem Statement

This author could not find any published academic papers on statistical arbitrage in the South African context. There is a definitive academic void on the performance of statistical arbitrage and pairs trading in the South African financial markets. This dissertation attempts to contribute to the academic pool of knowledge in that void.

## 1.3 Primary Research Objective

In light of the above, the objective of this dissertation is to investigate empirically the performance of a statistical arbitrage trading strategy in the South African equities market. The performance measure that will be used is the Sharpe ratio, in line with other authors that tested statistical arbitrage strategies in international markets (Avellaneda and Lee, 2010). This research intends to answer the question: *Would a statistical arbitrage portfolio have delivered good risk-adjusted returns in a South African context over the last 10 years, as measured by the Sharpe ratio?*

## 1.4 Chapter Summary

This chapter gave a very brief introduction on pairs trading and statistical arbitrage and put the objective of this research project into context.

The rest of this report is structured as follows: Chapter 2 discusses the academic literature relevant to this project. Chapter 3 discusses practical considerations when implementing the model. Chapter 4 discusses empirical results and chapter 5 concludes this report.

# Chapter 2: Literature Review

## 2.1 Introduction

Chapter 2.2 starts the literature review by providing a historical overview of statistical arbitrage. Chapter 2.3 formally defines a statistical arbitrage and chapter 2.4 reviews literature explaining the broad concept of statistical arbitrage.

Chapters 2.5 – 2.11 address a number of fundamental building blocks that are typically used in the construction of pairs trading systems.

Most techniques considered in the report up to that point will have been techniques that are applicable to two securities (i.e. techniques for pairs trading). To move from pairs trading to statistical arbitrage (where multiple securities are traded together), more advanced techniques are often required. Chapter 2.12 will discuss a few popular statistical arbitrage techniques.

Chapters 2.13 and 2.14 discuss the performance of statistical arbitrage systems in an international and South African context, respectively.

## 2.2 Historical Overview of Statistical Arbitrage

The first practical model of statistical arbitrage was developed in the late 1970's at Princeton-Newport partners (Thorp, 2004). In essence, the model ranked stocks by their percentage change in prices over a recent period, say one week, and found that the stocks that had performed strongest recently had a tendency to fall relative to the market in the near future and *vice versa* for weak stocks. This empirical tendency for short-term reversal was key to statistical arbitrage: by selling the top decile and buying the bottom decile one could form a profitable market neutral portfolio. Princeton-Newport partners shelved the idea at the time.

In the early 1980's a programmer at Morgan Stanley, Gerry Bamberger, developed statistical arbitrage from the ideas of pairs trading when he was tasked to help out on their block-trading desk (Bookstaber, 2007:185-190). An important driving force behind pairs trading and statistical arbitrage is the concept of liquidity demand in the market – a liquidity demander (for whom time is more important than price) is willing to sacrifice in execution price to get his immediate need fulfilled. Once this transitory liquidity demand has passed, prices would usually return to an equilibrium level (Bookstaber, 2007:182-194). Bamberger's model contained two key ideas: the short term reversal effect described above and a diversification tool for reducing risk: dividing the market into industry groups and to trade each group separately on a dollar-neutral basis (Thorp, 2004).

By the mid 1980's Bamberger left Morgan Stanley and formed a joint venture with Princeton-Newport partners and profitably traded his model into the late 1980's upon which their partnership was terminated by Bamberger's retirement from Wall Street (Patterson, 2010).

By the late 1980's Princeton-Newport partners had developed another model, making use of factor analysis – common tendencies shared by companies – to drive their statistical arbitrage trading portfolios (Thorp, 2004).

By the late 1980's and early 1990's a number of other statistical arbitrage operations were being developed, including the likes of David E. Shaw and Peter Muller at Morgan Stanley (Thorp, 2004 and Bookstaber, 2007). 1990 also saw the publication of two important articles by Jegadeesh (1990) and Lehmann (1990) on the predictability of the short-term reversal effect of securities, the source of returns for statistical arbitrage.

From those early years statistical arbitrage has become a multi-million dollar industry. Statistical arbitrage was employed not only in the equities market, but across asset classes including fixed income, commodities and currencies.

Recently Avellaneda and Lee (2010) have published a paper showing that statistical arbitrage profits have dropped after 2002 and 2003 in the U.S. equities markets. The natural evolution from a manual block-trading desk to pairs trading to computerised statistical arbitrage had taken its next evolutionary step to high frequency statistical arbitrage.

## 2.3 Definition of Statistical Arbitrage

Formally, Hogan, Jarrow, Teo and Warachka (2004) define statistical arbitrage as a zero initial cost, self-financing trading strategy ($x(t) : t \geq 0$) with cumulative discounted value $v(t)$ such that:

1. $v(0) = 0$,

2. $\lim_{t \to \infty} E^P[v(t)] > 0$,

3. $\lim_{t \to \infty} P(v(t) < 0) = 0$, and

4. $\lim_{t \to \infty} \frac{Var^P[v(t)]}{t} = 0 \text{ if } P(v(t) < 0) > 0 \; \forall t < \infty$

Statistical arbitrage is thus defined intuitively as a strategy that generates positive profits, on average, in present value terms and in which the variation in the profit series averaged across time is increasingly negligible (Do and Faff, 2010).

Inspecting the above definition, one sees that the standard deterministic arbitrage of financial mathematics is simply a special case of statistical arbitrage.

## 2.4 Short Term Reversal Effect

One of the early and most basic statistical arbitrage techniques was studied, amongst others, by Lehmann (1990). Lehmann found that portfolios of securities that had positive returns in one week typically had negative returns the following week, and that portfolios with negative returns in one week typically had positive returns the following week. A typical portfolio would then be constructed by purchasing a set dollar amount of 'losers' and selling short the same dollar amount of 'winners'. Lehman (1990) found that the sizeable return reversal of 'winners' vs. 'losers' mirror apparent arbitrage profits which persist even after correction for bid-ask spreads and plausible transaction costs.

This technique was also used more recently by Khandani and Lo (2007) in their interesting paper that dealt with a massive market-wide shock to statistical arbitrage / market neutral portfolios in August 2007.

## 2.5 Statistical Arbitrage Building Blocks

When reviewing statistical arbitrage models, one may observe that these models are often built up using more fundamental building blocks. A few of these building blocks deserve our attention:

i) One usually sees that statistical arbitrage trading systems attempt to find financial instruments that exhibit 'similar' behaviour. A popular method to measure 'similarity' will be discussed in section 2.6 of this literature review.

ii) One finds that some authors want to filter the similar securities, i.e. find similar securities that are 'more similar' or find securities that provide more consistent mean reversion. Some of these filter criteria will be discussed in section 2.7.

iii) One also finds that entry and exit points for a statistical arbitrage strategy can be determined in different ways (Pole, 2007). Some authors use simple statistical measures such as averages and ranges to determine extremities, whilst other authors attempt modelling the components of a statistical arbitrage using state space models and mean reverting Ornstein-Uhlenbeck processes. Methods to determine entry and exit points will be discussed in sections 2.8 to 2.11 below.

iv) Finally, one finds that statistical arbitrage systems make use of techniques to group financial instruments together. Market neutral portfolios are then constructed from these groups. The portfolios constructed as such are expected to have reduced variance and the techniques are therefore applied in a risk-management context as well (Thorp, 2004). Some of the more prevalent techniques will be discussed in section 2.12.

## 2.6 Cointegration

As was described above, statistical arbitrage seeks to find financial instruments that display 'similar' historical behaviour. Cointegration is a statistical technique that can be employed to determine whether financial instruments display 'similar' behaviour. The concept of cointegration has seen a lot of application in academic papers on statistical arbitrage and pairs trading. See for example Gatev, Goetzmann and Rouwenhorst (2006), Avellaneda and Lee (2010), Govender (2011) and Caldeira and Moura (2013).

Most students of stochastic processes are familiar with the analogy between a random walk and the path of a drunken man. Murray (1993) extends this intuitive analogy to cointegration by considering the paths of a drunk and her dog. Their paths diverge from time to time, but an error correcting mechanism prevents them from diverging too far. The distance between a drunk and her dog is a random variable, but it is stationary despite the nonstationarity of the two paths (Murray, 1993). Their paths can be viewed as being cointegrated.

Before cointegration can be formally considered, the concept of integration needs to be clarified: A univariate time series $P_t$ is integrated if it can be brought to stationarity through differencing. The number of times the series has to be differenced to achieve stationarity is referred to as the order of integration and is indicated $I(d)$, with $d$ the order of integration (Murray, 1993). Stationary time series are denoted $I(0)$.

Now, if two nonstationary time series are integrated of the same order and there exists a nontrivial linear combination of these two time series that is stationary, the two time series are said to be cointegrated (Murray, 1993).

Formally, Engle and Granger (1987) define cointegration as follows: the components of vector $\boldsymbol{P}_t = (P_{1t}, P_{2t}, \dots, P_{nt})'$ are said to be cointegrated of order $d, b$ denoted by $\boldsymbol{P}_t \sim CI(d, b)$, if

1. All components of $\boldsymbol{P}_t$ are $I(d)$ and

2. There exists a vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_n)$ such that the linear combination

$$\boldsymbol{\beta}\boldsymbol{P}_t = \beta_1 P_{1t} + \beta_2 P_{2t} + \dots + \beta_n P_{nt} \sim I(d-b), b > 0.$$

The vector $\boldsymbol{\beta}$ is called the cointegrating vector (Enders, 2010:359).

Through the use of cointegration, it is possible to detect stable long-run relationships among non-stationary variables.

Stock and Watson (1988) observed that two $I(1)$ stochastic variables that are cointegrated of order $(1,1)$ share the same stochastic trend (as in Enders, 2010:363-364). This is fundamentally a very satisfying result for pairs trading: Consider for example two stocks, Anglo Platinum and Impala Platinum. Seen that they have fundamentally very similar businesses (both mine platinum group metals in Southern Africa) it would be logical for them to be driven to a large extent by a common stochastic trend. If these stocks are cointegrated, there exists a specific linear combination of these two stocks that would purge the common stochastic trend from the combination. This is essentially the same insight that is used in the derivation of the Black Scholes option pricing formula.

Another important feature of cointegration is encapsulated by the Granger representation theorem. It states that for any set of $I(1)$ variables, cointegration and error correction are equivalent representations (Enders, 2010:370). Without going into details of an error correction model, this means that changes in prices of two variables will be influenced by past deviations from their long term equilibrium.

Considering the information in the above two paragraphs, it is easy to see why cointegration has found application in pairs trading.

The Engle-Granger methodology is a step-by-step test procedure that can be followed to determine whether two price series are cointegrated (Enders, 2010:373). Consider a situation where one would want to determine whether two stock prices $P_{1t}$ and $P_{2t}$ are cointegrated. The first step in the Engle-Granger methodology would be to test the order of integration of each of the series. An augmented Dickey-Fuller test can be used to determine the number of unit roots in each variable. The two price series need to be integrated of the same order. The next step would be to estimate the long-run equilibrium relationship of the form

$$P_{1t} = \beta P_{2t} + e_t$$

Denote the residuals from an OLS regression as

$$\hat{e}_t = P_{1t} - \hat{\beta} P_{2t}$$

If these residuals are stationary, the two price series are cointegrated. The stationarity of the residuals can be verified using an augmented Dickey-Fuller test.

For pairs trading, the above process can be used to achieve two important outcomes: (i) To identify a cointegrating relationship and (ii) to obtain the stationary cointegrating residual (in statistical arbitrage literature, this residual is also referred to as the *spread*).

Since we expect the residual to oscillate near some equilibrium level, we can use it to generate trading signals (Avellaneda and Lee, 2010).

A typical contrarian (or mean reverting) investment strategy suggests going long 1 South African Rand worth of stock $P_1$ and shorting $\hat{\beta}$ Rand worth of stock $P_{2t}$ if $\hat{e}_t$ is small, and on the other hand going short $P_{2t}$ and long $P_1$ if $\hat{e}_t$ is large.

Once cointegrating pairs have been identified, the statistical arbitrageur is left with the decision as to which of these cointegrating pairs to trade. Many authors use additional filter criteria to identify pairs that are 'more similar' or to find pairs that should provide more consistent mean reversion. Their work is discussed in the next section.

## 2.7 Additional Filter Criteria

Gatev, Goetzmann and Rouwenhorst (2006) match stock pairs in the spirit of cointegration. They assume that a sum of squared deviations statistic on the pairs that reverts to zero implies cointegration. Whilst this is usually one of the early steps to determine whether cointegration exists, they do not actually test for stationarity of the residuals (using for example an augmented Dickey-Fuller test) as is done in the Engle-Granger methodology.

Once they have identified pairs that display similar behaviour, they attempt to identify the best behaving pairs amongst them. So, as filter criteria, they choose the 20 stocks from an in-sample period with the smallest sum of squared deviations and trade these stock pairs out-of-sample (Gatev *et al,* 2006).

Gatev *et al* (2006) notice that as much as 78% of their pairs come from stocks pairs where both stocks lie within the same broad industry groups as defined by Standard & Poors: Utilities, Transportation, Financials and Industrials. This makes sense: securities from similar industries should be exposed to the same fundamental risk factors and therefore should behave in a similar fashion (barring idiosyncratic shocks). Gatev *et al* (2006) however find that the large majority of stock pairs come from the Utilities group. The question then arises whether most of the profitability of their strategy is driven by the pairs from Utilities? As another filter criterion, Gatev *et al* (2006) then group pairs using the above four group industry classification scheme and trade only pairs from within the same industry group. They find that pairs from all industry groups are significantly profitable, with Utilities being the most profitable.

Do and Faff (2010) also investigated additional algorithms to help filter and select trading pairs. Like Gatev *et al* (2006) they grouped trading pairs using an industry classification scheme. Do and Faff however went further than Gatev *et al* (2006): They used a finer 48-industry classification scheme from Fama and French (1997) and found that it generated immediate improvement in their results over the cruder four group classification scheme of Gatev *et al* (2006).

Do and Faff (2010) also investigated the number of zero crossings of the residual in the in-sample formation period. They defined the number of zero crossings as the number of times the residual crossed the value zero. Do and Faff postulated that the number of zero crossings seemed to have some usefulness in predicting future convergence in that the trading pair has a "track record" of mispricing that was subsequently corrected by market participants. Their analysis shows that this was indeed the case.

Do and Faff (2010) furthermore combine the metrics (the sum of squared differences, the industry homogeneity criteria and the number of zero crossings) and found that the combination of these metrics enhanced trading profits considerably.

Avellaneda and Lee (2010) also used filter criteria: They obtained cointegrating residuals and modelled these as mean reverting Ornstein Uhlenbeck processes with a parameter that estimated the characteristic time scale of mean reversion. They selected stocks with mean reversion times less than half the in-sample period.

Avellaneda and Lee (2010) also proposed another filter criterion: to generate signals that took volume information into account. Their criteria favoured mean-reverting price signals that took place on low volume and mitigated signals that took place on high volume. These signals showed improvement with most of the strategies they tested and significant improvement in performance with their Exchange Traded Fund strategies.

Caldeira and Moura (2013) also make use of an additional filter criterion: they compute and in-sample Sharpe ratio and trade out-of-sample the 20 pairs that had the best Sharpe ratio.

It would seem that the use of additional filter criteria is an area rich for research, bounded only by the ingenuity of the researcher.

## 2.8 Archetypal Residuals: Sinusoidal and Popcorn processes

Once a statistical arbitrageur has identified asset series' displaying acceptable levels of 'similarity' and has selected pairs by using filter criteria, he typically attempts to identify entry and exit points by observing the residual. There are two fundamental ways in which entries and exits are typically made (Pole, 2007: 18-20). This is easiest illustrated by an example:

Some statistical arbitrageurs view the archetypal residual process as a sine wave, and they enter a short position in the residual when large positive deviations from the mean occur. They then hold these positions until a large negative deviation from the mean occurs, upon which they close out their short position and enter a long position in the residual. These traders effectively close and reverse their positions at local maxima and minima. This sinusoidal strategy is therefore always in the market. The sinusoidal strategy is illustrated in figure 2-1 below, with a green arrow illustrating a long entry in the residual and a green block the closure of a long position. A red arrow indicates a short entry in the residual and a red block is indicative of covering the short.

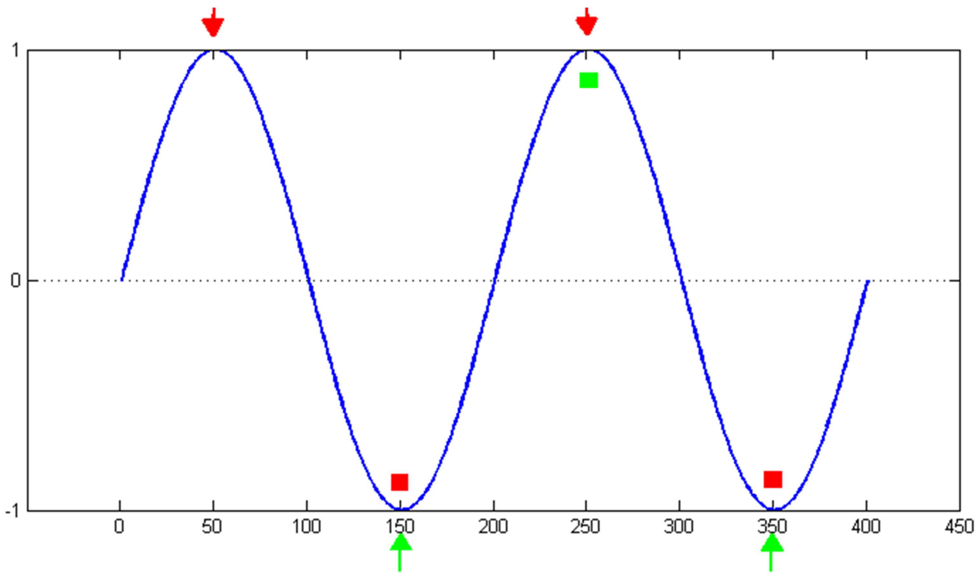


Figure 2-1: Sinusoidal entry and exits

Other statistical arbitrageurs however prefer to exit their positions upon reversion to the mean, and can accommodate a process that is not strictly sinusoidal (i.e. where troughs are not necessarily followed by peaks and so on). Pole (2007) refers to this as a popcorn process, and his convention is adopted here. An immediate consequence of such a strategy is that the statistical arbitrageur is not in the market the entire time.

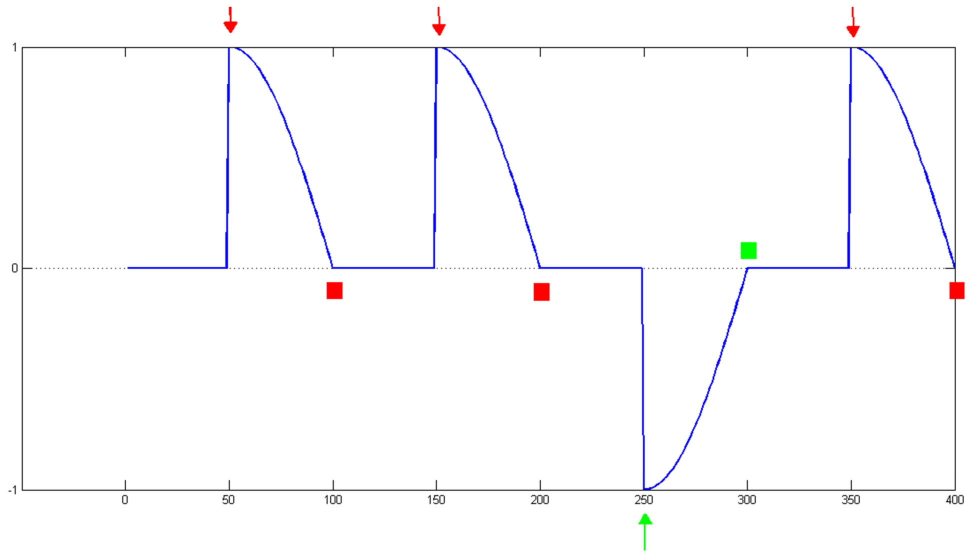


Figure 2-2: Popcorn entry and exits

In figure 2-2 above a statistical arbitrageur would enter a short position in the residual at a local maximum and cover his short at the mean. He would then be free to enter a short again at the next local maximum and cover upon mean reversion; he would not have to hold out for a trough in between peaks. In figure 2-2, entries and exits are indicated using the same colour blocks and arrows as in figure 2-1.

## 2.9 Statistical Measures to Determine Entry and Exit Points.

Once the cointegrating residual has been identified and an archetypal entry and exit strategy has been chosen, the precise question of *when* to enter and exit the residual (and by implication the underlying assets) needs to be answered. It is the task of the statistical arbitrageur to determine entry and entry and exit points for the trading system.

A typical starting point would be to calculate a moving average over a rolling time window and place entry points at a set number of standard deviations above and below the rolling (local) mean (Pole, 2007 and Do and Faff, 2010).

Figure 2-3 below illustrates the residual on an AGL-BIL pair for a period from July 2008 to July 2009 as well as a rolling moving average as well as dotted lines 1 standard deviation above and below the mean. Assuming an archetypal popcorn trading process, short entries in the residual would be made above the dotted line and closed when reaching the mean. Equivalently, long entries would be made below the dotted line and exits would be made at the mean.

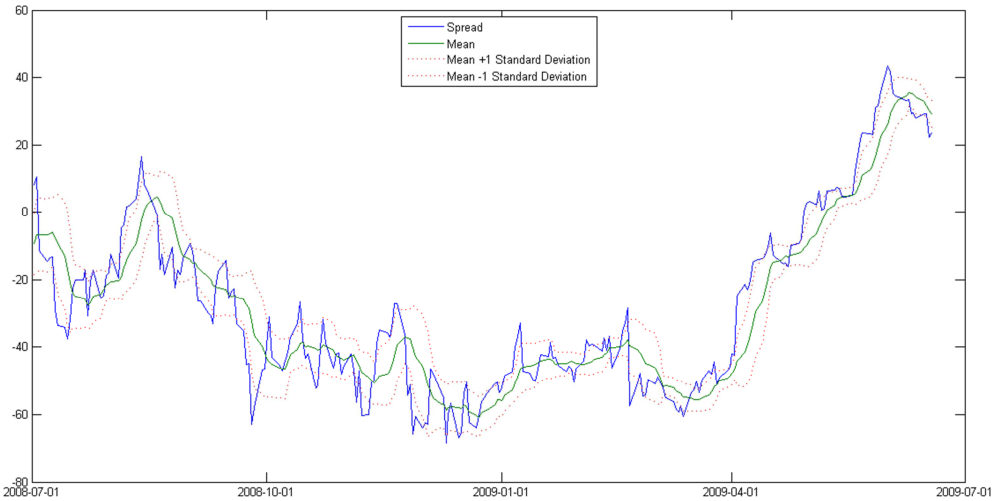


Figure 2-3: The residual on AGL-BIL with a rolling mean and 1 standard deviation band to indicate entry points.

The implicit assumption when using the standard deviation is that the residual is distributed normally. Care has to be taken here when working with financial data.

Another way to determine entry and exit points would be to calculate a rolling window maximum and to add and subtract from it a percentage of the local range (Pole, 2007).

Figure 2-4 below illustrates the AGL-BIL residual over the same timespan with a rolling mean as well as *max* residual – 20% range line to indicate short entries and another line with a *min* residual +20% range to indicate long entries.

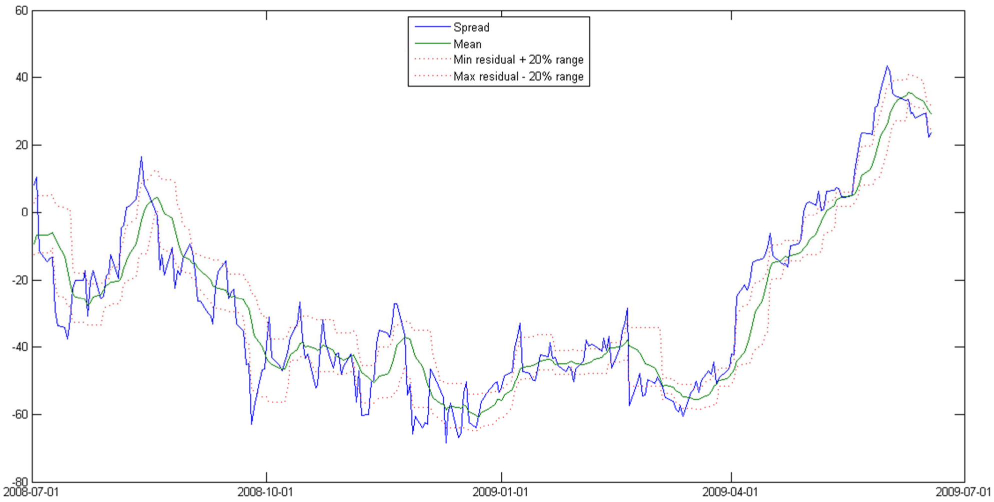


Figure 2-4: The residual on AGL-BIL with a rolling mean and *max* residual – 20% range and *min* residual +20% range lines to indicate entry points.

This approach has the advantage of not having to assume normality of the residual but does also force one to use outliers in the data. Working with outliers and extremes can become a complicated affair, especially when attempting to model them (Pole, 2007).

A third alternative is to use percentile rankings of the local distribution. Figure 2-5 below shows the same residual as above with a local mean and lines indicating the local 80th and 20th percentile points.

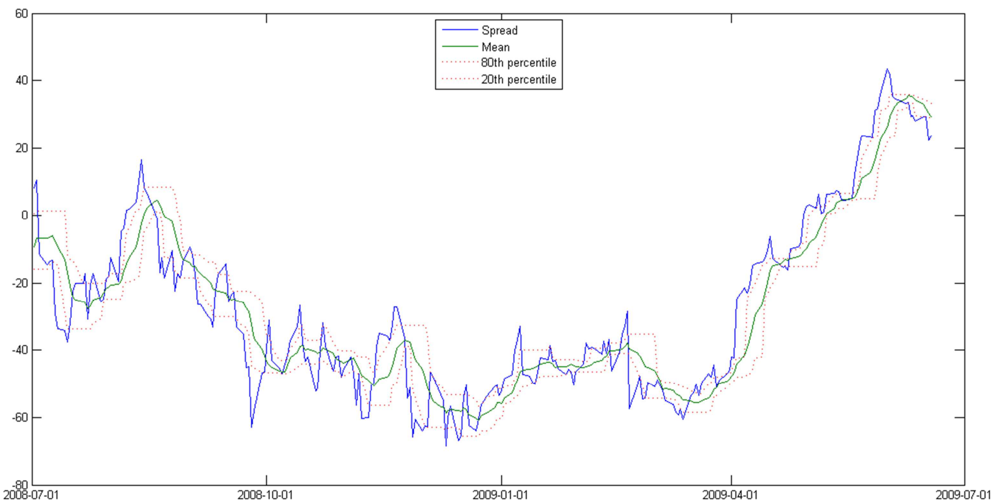


Figure 2-5: The residual on AGL-BIL with a rolling mean and local 20th and 80th percentile points indicating long and short entry points.

To the naked eye, it would seem that there is not much difference between the three graphs above.

As another alternative, Pole (2007) suggests the use of an Exponentially Weighted Moving Average (EWMA) from time series statistics as opposed to a rolling moving average as a point estimate for the mean.

The advantage in using the EWMA becomes evident when the residual makes step changes or there is a trend in the residual. The EWMA easily allows for one time use of an intervention discount factor to quickly adjust to step changes in the residual. The interested reader is referred to Pole (2007) for further information.

## 2.10 State Space Modelling

Some state space models can be used as an alternative to a rolling (local) mean or an EWMA. State-space models may be used to describe the behaviour of an unobserved component (the true residual) by defining its relationship with the observed process (the estimated/observed residual).

Triantafyllopoulos and Montana (2009) suggested such a model, expanding on the work of Elliot, van der Hoek and Malcolm (2005). Triantafyllopoulos and Montana's model involves a Gaussian linear state-process for modelling mean-reverting spreads arising in pairs trading. In their model, they view the observed residual as a noisy realization of the true residual, suggesting that discrepancies between the true and observed residual could be temporary market inefficiencies. This is similar to the using a mean or an EWMA as discussed in section 2.9 above. The model suggested by Triantafyllopoulos and Montana (2009) was unsuccessfully employed by both Govender (2011) and Duyvené de Wit, de Alessi and Moodliyar (2012).

Another model that can be considered to estimate a local mean is the Kalman filter – an algorithm that produces estimates of unknown variables from observed measurements (that may include noise and other inaccuracies).

## 2.11 Ornstein-Uhlenbeck Processes

Section 2.9 above discussed a few statistical measures (standard deviation, range and percentiles) that can help to determine entry and exit points. Other alternatives also exist: Bertram (2010) suggests modelling the residual as a mean-reverting Ornstein-Uhlenbeck process, and he derives analytical formulas to maximise the in sample mean return and Sharpe ratios, taking trading costs into consideration.

Bertram's (2010) derivation loosely proceeds along these lines (the interested reader is referred to his original work):

Model the spread $w_t$ as a mean-reverting Ornstein-Uhlenbeck process represented by the following stochastic differential equation:

$$dw_t = -\alpha w_t dt + \sigma dB_t$$

where $\alpha > 0$, $\sigma > 0$ and $B_t$ is a Wiener process. The above stochastic differential equation is essentially an Ornstein-Uhlenbeck process with its mean set to zero. Through inspection of the above equation, one can see the pull (mean-reversion) towards zero.

Denote the point at which a trade is entered as $a$ and the exit point for a trade as $m$. Define transaction costs as $c$.

The profit per trade can then be defined as $r(a, m, c) = m - a - c$ and the expected profit per unit time is shown by Bertram (2010) to be:

$$\mu(a, m, c) = r(a, m, c)/\mathbb{E}[\tau]$$

where $\tau$ represents the time taken to complete a trade cycle.

Bertram (2010) then splits the trading cycle $\tau$ into two first-passage times (going firstly from $a$ to $m$ and then from $m$ to $a$), and derives an expression for the function $\mu(a, m, c)$:

$$\mu(a, m, c) = \frac{\alpha(m - a - c)}{\pi\left(Erfi\left(\frac{m\sqrt{\alpha}}{\sigma}\right) - Erfi\left(\frac{a\sqrt{\alpha}}{\sigma}\right)\right)}$$

In order to maximise this function, its derivative with respect to $a$ and $m$ are taken and the resulting equations solved. Bertram (2010) shows that $m = -a$, indicating that the bands are symmetrical around zero. Bertram (2010) approximates the solution with a Taylor series:

$$a = -\frac{c}{4} - \frac{c^2\alpha}{4\left(c^3\alpha^3 + 24c\alpha^2\sigma^2 - 4\sqrt{3c^4\alpha^5\sigma^2 + 36c^2\alpha^4\sigma^4}\right)^{\frac{1}{3}}} - \frac{\left(c^3\alpha^3 + 24c\alpha^2\sigma^2 - 4\sqrt{3c^4\alpha^5\sigma^2 + 36c^2\alpha^4\sigma^4}\right)^{\frac{1}{3}}}{4\alpha}$$

Bertram (2010) thus derived an analytical formula to maximise the in sample mean return of a mean-reverting spread process, taking trading costs into consideration. Bertram's derivation is applicable to a sinusoidal archetypal trading process (as discussed in section 2.8 above).

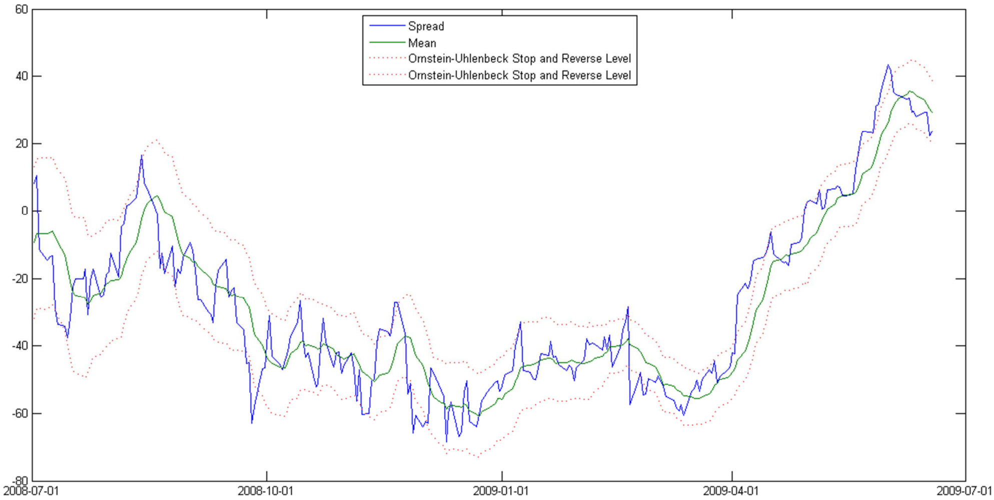


Figure 2-6: The residual on AGL-BIL with a rolling mean and Ornstein-Uhlenbeck stop and reverse levels as per Bertram (2010).

A derivation similar to Bertram, but for a popcorn trading process was done by Duyvené de Wit, de Alessi and Moodliyar (2012) for their semester project for the subject Applied Time Series Analysis. Their derivation follows:

Use the same equation for $\mu(a, m, c)$ from above but let $m = 0$. This is done because trades should be closed upon mean-reversion of the archetypal popcorn process, i.e. when the spread $w_t = 0$. The partial derivative of $\mu$ with respect to $a$ is given by:

$$\frac{\partial \mu}{\partial a} = \frac{\alpha}{\pi Erfi\left(\frac{a\sqrt{\alpha}}{\sigma}\right)} - \alpha(a+c)\left[\pi Erfi\left(\frac{a\sqrt{\alpha}}{\sigma}\right)\right]^{-2} \pi \frac{2}{\sqrt{\pi}} e^{\frac{a^2\alpha}{\sigma^2}} \frac{\sqrt{\alpha}}{\sigma}$$

$$= \frac{\alpha}{\pi Erfi\left(\frac{a\sqrt{\alpha}}{\sigma}\right)} - \frac{2\alpha^{\frac{3}{2}}}{\pi^{\frac{3}{2}}\sigma}(a+c)\left[\pi Erfi\left(\frac{a\sqrt{\alpha}}{\sigma}\right)\right]^{-2} e^{\frac{a^2\alpha}{\sigma^2}}$$

This partial derivative can be set equal to zero and a numerical solution found for $a$. Note that the parameters $\alpha$ and $\sigma$ in the model need to be estimated: Yurek and Yang (2007) provide a simple estimation methodology.

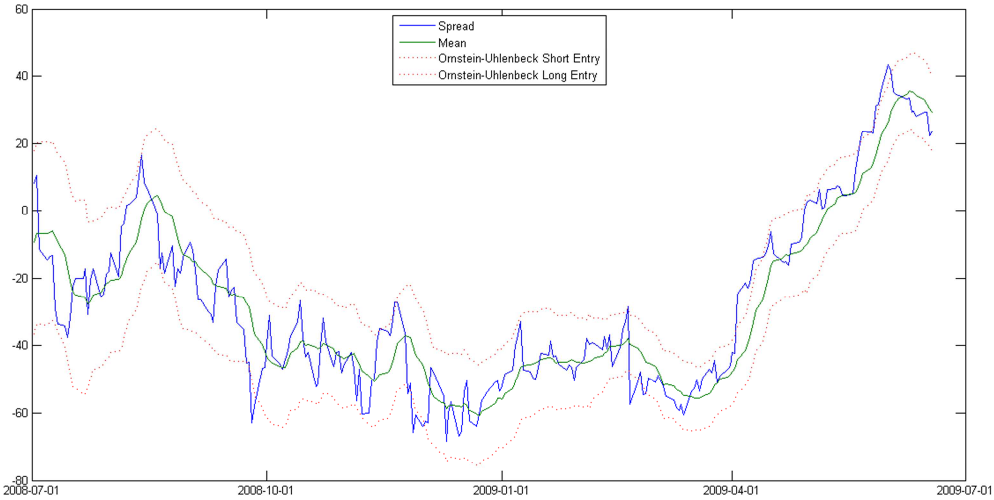


Figure 2-7: The residual on AGL-BIL with a rolling mean and Ornstein-Uhlenbeck short and long entry levels for an archetypal popcorn trading process.

To the naked eye, figures 2-6 and 2-7 would seem fairly close – they do however differ: the OU levels indicated in figure 2-7 are wider than those indicated by 2-6. This makes sense intuitively, since an archetypal popcorn process would have to account for similar transaction costs compared to a sinusoidal process, but with a smaller per-trade profit margin.

## 2.12 Moving from Pairs Trading to Statistical Arbitrage

The literature review has thus far mostly considered techniques that are applicable to two securities, i.e. techniques for pairs trading. To move from pairs trading into statistical arbitrage (where multiple securities are traded together), more advanced techniques are often required. This section will briefly mention a few popular techniques. These techniques were not used in this study, but are mentioned here in the context of a literature review.

Trading groups of stocks together is the natural extension of pairs trading. Consider the financial services sector in South Africa. Perform a cointegration / regression analysis on a number of stocks in the sector against the sector benchmark / index e.g. the Satrix FINI Exchange Traded Fund (ETF). The analysis will find that some stocks are cheap relative to the index, some are fairly priced and that others are expensive. A statistical arbitrage portfolio would hold a collection of trades relative to the ETF (or more generally, factors explaining the systematic stock returns). Some stocks will be held long against a short position in the ETF, whilst other stocks will be held short against a long position in the ETF. When netting off a large number of the positions, one would expect the holdings in the ETF to be a small fraction of the total holdings. The overall portfolio will therefore look like a long/short portfolio of single stocks.

A popular technique to handle a multitude of securities is the statistical technique of Factor Analysis. Factor Analysis attempts to describe the movements of a number of correlated observed variables using a (usually) lesser number of unobserved factors (Pole, 2007). These unobserved latent factors are in themselves weighted linear combinations of the observed variables. Factor Analysis has a long history in financial academic literature. In 1976, Ross proposed his well-known Arbitrage Pricing Theory (APT). APT proposed that asset returns can be explained using an *n*-factor model, with the unspecified factors left to empirical study. Later, Roll and Ross (1980) used maximum likelihood factor analysis to estimate factor loadings as well as the number of factors generating returns in Ross' APT-model.

In their 2010 paper, Avellaneda and Lee use a single factor model to test for statistical arbitrage profits. They use capital weighted ETFs as factors and trade similar industry groups of stocks against their respective ETFs.

Principal Component Analysis, which on the surface is very similar to Factor Analysis (Hull, 2012), is another technique that is often used as a building block in statistical arbitrage trading systems. "The central idea of Principal Component Analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables" (Jolliffe, 2002). The main goal of PCA is to find an orthogonal matrix with the property that the principal components are uncorrelated and are arranged in order of decreasing variance (Lay, 1997). PCA is especially valuable in applications where most of the variation in the data is due to variations in only a few of the principal components (Lay, 1997).

Avellaneda and Lee (2010) also made use of PCA in determining factors (and hence in determining residuals) for statistical arbitrage systems. They conclude that the number of factors needed to explain stock returns should be equal to the number of eigenvalues needed to explain approximately 50% of the variance of the empirical correlation matrix. They found that this number varied across time and that it lay somewhere between 10 and 30 factors. They also observed that this number varied inversely to the VIX Option volatility index, "… suggesting more factors are needed to explain stock returns when volatility is low, and less in times of crisis…" (Avellaneda and Lee, 2010).

Avellaneda and Lee's (2010) main contribution was to show how different sets of factors lead to different residuals and to different profit and loss profiles for statistical arbitrage models.

## 2.13 The Performance of Pairs Trading Globally

This section reviews international academic literature on pairs trading and highlights their findings in terms of the profitability of pairs trading strategies.

Caldeira and Moura (2013) test a straightforward statistical arbitrage strategy on Brazilian stocks between 2005 and 2012. Their strategy delivers excess returns of 16.38% per year with a Sharpe ratio of 1.34. They report that their strategy has a low correlation with the market.

Avellaneda and Lee (2010) measure statistical arbitrage strategies in the US Equities market between 1997 and 2007. Their PCA-based strategies have a Sharpe ratio of 1.44 and their ETF-based strategies have a Sharpe ratio of 1.1 over the period. An interesting observation from their study is that the performances of both their ETF and PCA strategies on U.S. stocks degrade after 2002. They do not allude to the reason for the degradation in performance.

In their 2010 paper, Do and Faff confirm the downward trend in profitability of pairs trading in the U.S. since the mid 1990's, affirming the findings of Avellaneda and Lee. Do and Faff's results furthermore suggest that the main driver in the reduction in pairs trading profitability is the worsening of general arbitrage risks (including fundamental risks like an unexpected disturbance in the relationship between two stocks) and not by an increase in market efficiency (e.g. the hedge fund industry competing away opportunities, as was mentioned by Gatev *et al* (2006)). Do and Faff's analysis shows that the increase in market efficiency did play a role, but that it was not the *main* driver for the decline in pairs trading profits.

Interestingly, Do and Faff (2010) also report significantly increased performance in times of turbulence (the bear markets between 2000-2002 and 2007-2009), which they attribute to a less efficient market during those times.

Gatev *et al* (2006) report annualized excess returns of about 11% for their portfolio of pairs. They study a very long timespan (from 1962 to 2002) which provides substantial support to the historical profitability of pairs trading. In their paper, they also study a true out-of-sample period of four years (1999-2002) and find it to be consistent with their long term results.

## 2.14 The Performance of Pairs Trading in South Africa

In 2011, Govender submitted a master's dissertation dealing with Statistical Arbitrage in South African Financial Markets. He made use of the recursive Bayesian algorithm suggested by Triantafyllopoulos and Montana (2011) to estimate his time-varying state-space parameters. Govender modelled deviations from the true spread as an OU process, deriving optimal entry and exit strategies analytically, as in Bertram (2010). Govender suggested that it was possible to derive large arbitrage profits from the strategy.

In 2012, Duyvené de Wit, de Alessi and Moodliyar submitted a semester project for the subject Applied Time Series Analysis at the University of Cape Town. In the project, they attempted to improve on Govender's approach: Govender made use of a sinusoidal entry and exit strategy, whereas Duyvené de Wit *et al* attempted to use a popcorn process for entry and exits. In reviewing Govender's work, they found that Govender might have accidentally set the level component of his state space model equal to zero – having a significant impact on the performance of his model. Duyvené de Wit *et al's* results were inconclusive due to difficulties in implementing the recursive Bayesian algorithm that estimated their state-space parameters.

Other than the above, this author could not find any published academic papers on statistical arbitrage in the South African context. It is thus clear that there is a significant academic void on the performance of statistical arbitrage and pairs trading in the South African financial markets. This dissertation attempts to contribute to the academic pool of knowledge in that regard.

## 2.15 Chapter Summary

This literature review set the scene for the empirical work that is to follow. It highlighted different components of statistical arbitrage trading strategies and compared and contrasted approaches that were implemented in prominent academic papers. It reviewed international papers on statistical arbitrage profitability and identified a void in South Africa that this dissertation will start to address.

# Chapter 3: Research Design

## 3.1 Introduction

This section discusses important decisions that were made when designing the pairs trading strategy. Broadly speaking, the following theoretical building blocks from the literature review above were used to construct a statistical arbitrage trading strategy: Liquid stocks on the JSE were selected for pairs trading by testing for cointegration. The top 20 pairs, based on their in-sample number of zero crossings, were selected for trading in an out-of-sample portfolio. Entry and exit points for an archetypal popcorn process were determined using a moving average, in combination with modelling the market inefficiencies as Ornstein-Uhlenbeck processes and maximising for their mean return in-sample.

More details are discussed in the sections that follow and the full model is discussed in chapter 3.6.

## 3.2 Choice of Trading Pairs, Data Vetting

The FTSE/JSE Top 40 Index was constructed and launched by FTSE on 24 June 2002 with a base value of 10300.31. This index made for a good choice as investable universe for this study: the stocks are both broadly monitored and acceptably liquid on an end-of-day basis. The stocks constituting the index are also announced quarterly *ad hoc*, meaning a reasonable investor following the markets will know what stocks will be included in the index for any upcoming date.

For the company to be included in the universe of investable stocks, the criterion is set that the company must have been in the top 40 index at the trade date (as opposed to the date when the research was undertaken). This decision avoids survivorship bias.

A convenient time period for consideration is 11 years since 1 July 2002, allowing for one year of in-sample data to set up models. Out of sample testing of the strategy will be undertaken from 1 July 2003 to 28 June 2013 (10 years). This date range was an arbitrary choice made when the research design was done.

One should note that the major bear market associated with the global financial credit crisis is contained in this sample, which is a good thing; one would prefer a strategy to handle extreme events rather than breaking down under extreme volatility / drawdowns.

Prices of the individual assets have been adjusted for dividends and corporate actions like stock splits, etc. This financial information was obtained from Bloomberg.

## 3.3 Instruments and Trading Costs

The instruments used for trading were the raw stocks themselves (not futures, contracts for difference or options). It is assumed that a statistical arbitrageur would have to pay in cash for the long positions in his holdings, and that he would have to post margin equal to the value of the assets he holds short. This choice strips the strategy from leverage and allows for baseline comparisons. Leverage can simply be applied afterwards. Leverage (position sizing) has the ability to turn winning strategies into losing strategies when too much leverage is applied. The application of leverage is left for another research project[1].

Transaction costs play a significant role in statistical arbitrage portfolios due to the high turnover in the portfolio. Because of this, three scenarios will be analysed: a scenario where transaction costs are set equal to zero, a scenario where transaction costs are set equal to the costs used in U.S. academic papers and a scenario where transaction costs are set equal to more realistic South African costs.

Slippage and costs that are used in U.S. academic studies (Avellaneda and Lee, 2010), amount to 0.05% (5 basis points) per trade, giving a round trip transaction cost of 10 basis points. Note that 'a position' is in actual fact a market neutral position in *two* assets:

$$y_t = p_{1t} - \beta_1 p_{2t}$$

Hence we calculate transaction costs as 0.05% of each asset holding at entry and 0.05% again for each asset holding when closing a position.

Through consultation with a local fund manager, a more realistic figure of 25 basis points (50 basis points round trip) will be used to emulate transaction costs for a South African institutional trader.

The trade entry and exit strategy undertaken in this study is discrete; there is no continuous trading. Positions are entered into when an entry is signalled, and unwound on an exit signal. Although it would be technically correct to adjust positions on a daily basis (due to changing betas), the simpler approach is taken in this study.

## 3.4 Sample Periods, Walk Forward Testing

During testing, an in-sample size of 60 trading days was selected for testing cointegration. This in-sample duration is of arbitrary length and was chosen to coincide with Avellaneda and Lee's (2010) choice. This in-sample period is defined in this report to be in-sample period $a$. Hence, using 60 historical values of $p_{1t}$ and $p_{2t}$, values for $\hat{e}_t$ are calculated.

For the moving average, one requires a number of values for $\hat{e}_t$ in order to calculate $x_t$. This is referred to as in-sample period $b$ and was chosen to be 10 days.

[1] For more information on position sizing, see Duyvené de Wit (2008) and Tharp (2008).

Thirdly, the OU section of the model requires a number of data points for $w_t$ in order to make an optimal choice for the value of $a_t$ (see section 2.11 for the relevant formulas). This period is referred to as in-sample period *c* and was also chosen to be 60 days long.

The net effect is that we need in-sample data of 130 days before the strategy can start trading; less than a year's worth of data is therefore needed to set up the model.

Like with Avellaneda and Lee (2010), estimation and trading rules are kept simple to avoid data mining.

Using the in-sample data (in our case 130 days), one then tests the strategy on the following day using the parameters obtained using data from the prior 130 days. The in-sample window of 130 days is then shifted forward by one day and updated parameters are obtained and used to generate buy/sell signals for the next day, and so on. This procedure is known as step-forward or walk-forward testing; the interested reader can refer to Kaufman (2005: 871-874). Like Avellaneda and Lee (2010:762), estimation is always done looking back from the trade date (130 days in our case), thus simulating decisions which may have been taken in real time. The goal of this is to avoid data snooping and as a consequence, unrealistic results.

## 3.6 The Model

Let the prices of two assets be denoted by $P_{1t}$ and $P_{2t}$ respectively. Let of $p_{1t} = \ln(P_{1t})$ and $p_{2t} = \ln(P_{2t})$. To determine whether these two assets are cointegrated, run a regression of the form:

$$p_{1t} = \beta p_{2t} + e_t$$

and estimate

$$\hat{e}_t = p_{1t} - \hat{\beta} p_{2t}$$

Verify the stationarity of the residuals $\hat{e}_t$ by performing an Augmented Dickey-Fuller test. Stationarity implies cointegration.

Signals are generated using a moving average of the residuals as a point estimate for the true value of the spread:

$$x_t = \frac{1}{n} \sum_{i=0}^{n-1} \hat{e}_{t-i}$$

where n = 10. Define

$$\hat{e}_t = x_t + w_t$$

Therefore

$$w_t = x_t - \hat{e}_t$$

Perform the above for all possible combination of stock pairs in the investable universe and select from the cointegrated pairs the top 20 pairs with the best in-sample number of zero crossings for trading.

Take entry and exit signals from these top 20 pairs.

Now model the spread $w_t$ as a mean-reverting Ornstein Uhlenbeck process with the following stochastic differential equation:

$$dw_t = -\alpha w_t dt + \sigma dB_t$$

Following the process in section 2.11 of the literature review above, one has to estimate $\alpha$ and $\sigma$. This is done using Jurek and Yang's (2007) methodology. Finally, one needs to optimise for the entry level $a$ in an archetypal popcorn process by solving for $a$ in the following equation (see section 2.11):

$$\frac{\partial \mu}{\partial a} = 0 = \frac{\alpha}{\pi Erfi\left(\frac{a\sqrt{\alpha}}{\sigma}\right)} - \frac{2\alpha^{\frac{3}{2}}}{\pi^{\frac{3}{2}}\sigma}(a+c)\left[\pi Erfi\left(\frac{a\sqrt{\alpha}}{\sigma}\right)\right]^{-2} e^{\frac{a^2\alpha}{\sigma^2}}$$

## 3.7 Weighting Two Assets in a Pair: The Interpretation of Beta

Since the regression is run on logged prices, the interpretation of $\beta$ in this context is not straightforward.

With logged asset prices $p_{1t}$ and $p_{2t}$ as defined above, also define a percentage change in asset prices as $c$ and holding weights in assets 1 and 2 as $h_1$ and $h_2$ respectively. In any pair, we want to hold the assets such that their weights add up to 1, i.e. $h_1 + h_2 = 1$.

When interpreting regression coefficients where both the dependent and independent variable are logged, one has to interpret the regression as follows: *Ceteris paribus*, a 1% change in $p_2$ should lead to a $100 \times (1.01^\beta - 1)$ percentage change in $p_1$, and thus in general, a $(100 \times c)\%$ change in $p_2$ should lead to a $100 \times ((1+c)^\beta - 1)$ percentage change in $p_1$.

But we want our pair to be market neutral, i.e. unaffected by say a 1% market drop (we would want our short position to gain the equivalent of what our long position loses).

Therefore,

$$(1+c)^\beta h_1 + (1-c)h_2 = 1$$

But since $h_2 = 1 - h_1$ we see

$$(1+c)^\beta h_1 + (1-c)(1-h_1) = 1$$

Solving for $h_1$ we find

$$h_1 = \frac{c}{(1+c)^\beta - (1-c)}$$

Testing our equation for the trivial case when $\beta = 1$ (a 1% change in asset 1 should lead to a 1% change in asset 2), we find that

$$h_1 = \frac{c}{2c} = \frac{1}{2}$$

This is an intuitively satisfying result illustrating that when asset 1 and asset 2 change in a one-to-one fashion, their weight in a pair should each be 50%.

## 3.8 Weighting Pairs in the Portfolio

As mentioned, the portfolio is constructed by taking signals from the twenty pairs with the highest number of zero crossings in sample. These twenty pairs are all equally weighted with 5% capital exposure to each pair trade. If there are not twenty trades open at a given point in time, all available capital is not exposed to the market.

There is another approach to portfolio weighting: scaling trades in line with the number of open positions (Gatev *et al* 2006:805), which is less conservative.

The former, more conservative approach is taken in this study. This will mean that all available capital is not necessarily exposed to the market at a given moment in time.

## 3.9 Chapter Summary

With this information, entry and exit levels are algorithmically explicit and trades can be entered and exited as the model dictates. The statistical arbitrage trading system can now be back tested and its performance measured over the ten year period. Results are discussed in the next chapter.

# Chapter 4: Empirical Results

## 4.1 Introduction

This chapter reports the empirical results of the research that was conducted. To help the reader develop intuition for the process, chapter 4.2 highlights results for a single pair. Chapter 4.3 reports results at the portfolio level.

## 4.2 Results for a Single Pair

The stock pair SAB/SOL (South African Breweries/SASOL) is discussed here for demonstrative purposes. The upper pane of figure 4-1 shows the residuals and the lower pane shows the output from the Augmented Dickey-Fuller test (a value of 1 indicates cointegration, and a value of zero the opposite).

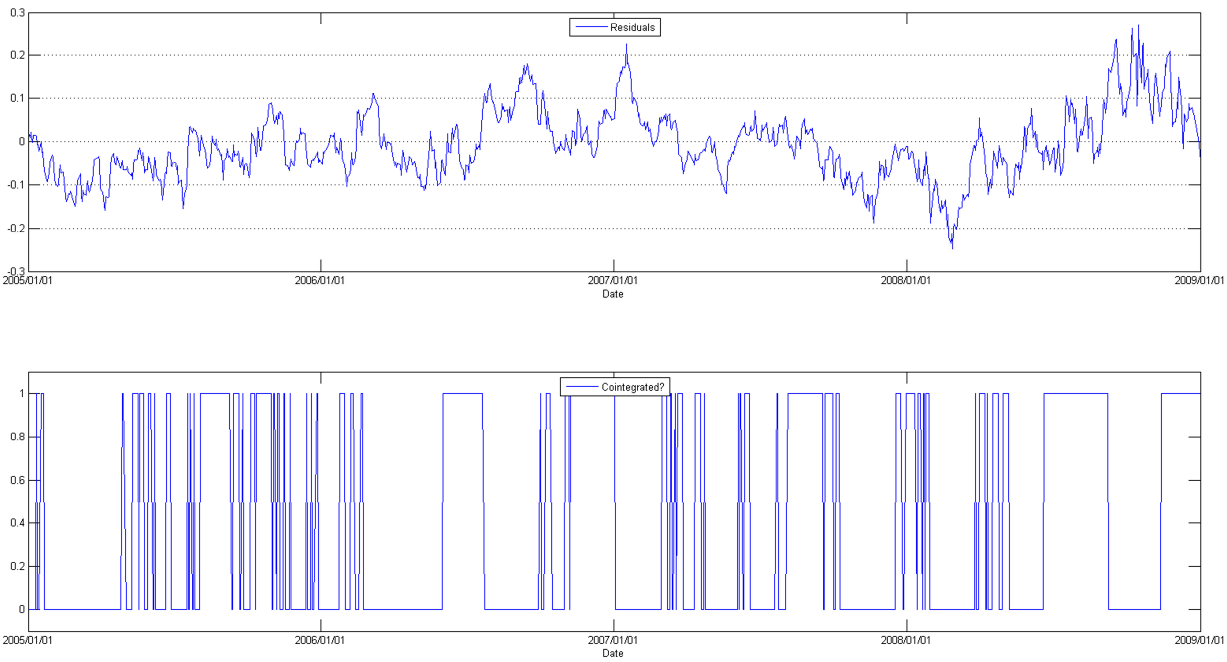


Figure 4-1 (a) Residuals on SAB/SOL and (b) output from the Augmented Dickey-Fuller test.

Figure 4-2 below shows the residuals for the same pair over the same period, including the local mean and Ornstein-Uhlenbeck entry levels. The lower pane in figure 4-2 shows the differenced residual $w_t = \hat{e}_t - x_t$ as well as the Ornstein-Uhlenbeck entry levels (the reader is referred to chapter 3.6 for the relevant variables).

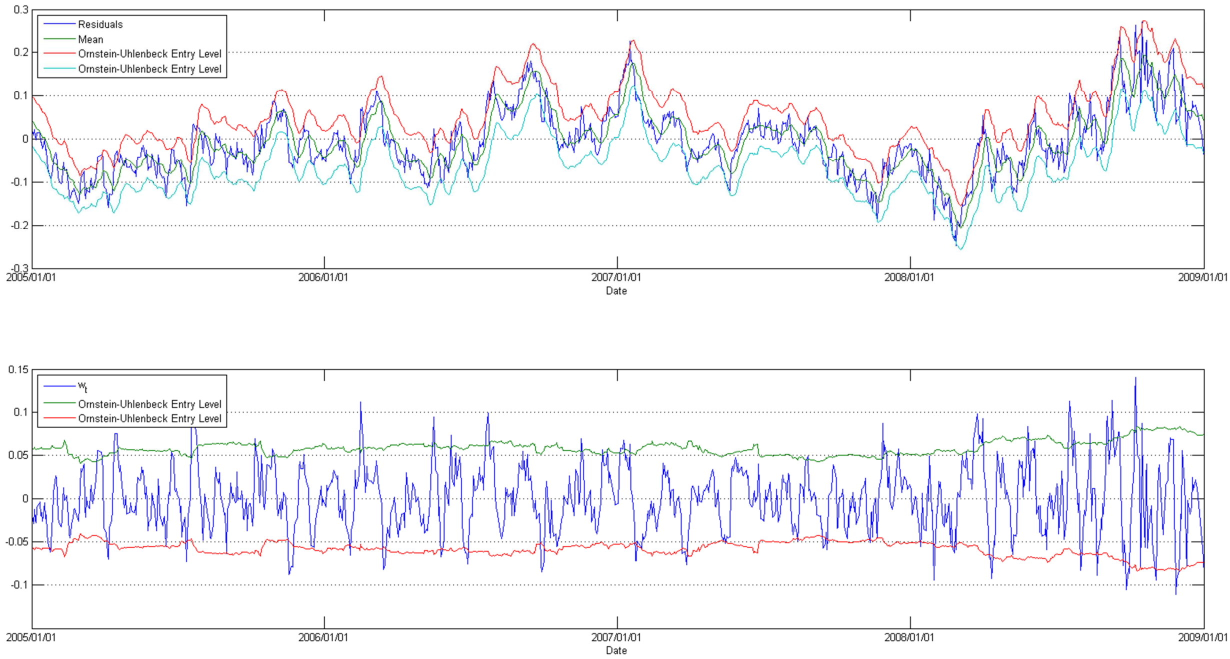


Figure 4-2 (a) Residuals on SAB/SOL with their local mean and Ornstein-Uhlenbeck entry levels and (b) the differenced residual $w_t$ as well as the Ornstein-Uhlenbeck entry levels.

Figure 4-3 below shows trading signals for the pair. For the popcorn process, one would expect to see entry signals upon breach of the Ornstein-Uhlenbeck entry and levels and exit signals upon mean reversion. One should however recall that other criteria (cointegration, the presence of the pair's underlying assets in the Top40 index) also play a role in determining whether signals for the pair are actually put forward for trading. In the bottom pane, a value of 2 indicates a long entry in the pair, 1 indicates closing a long position. A value of -2 indicates entry in a short entry and -1 indicates covering the short position.

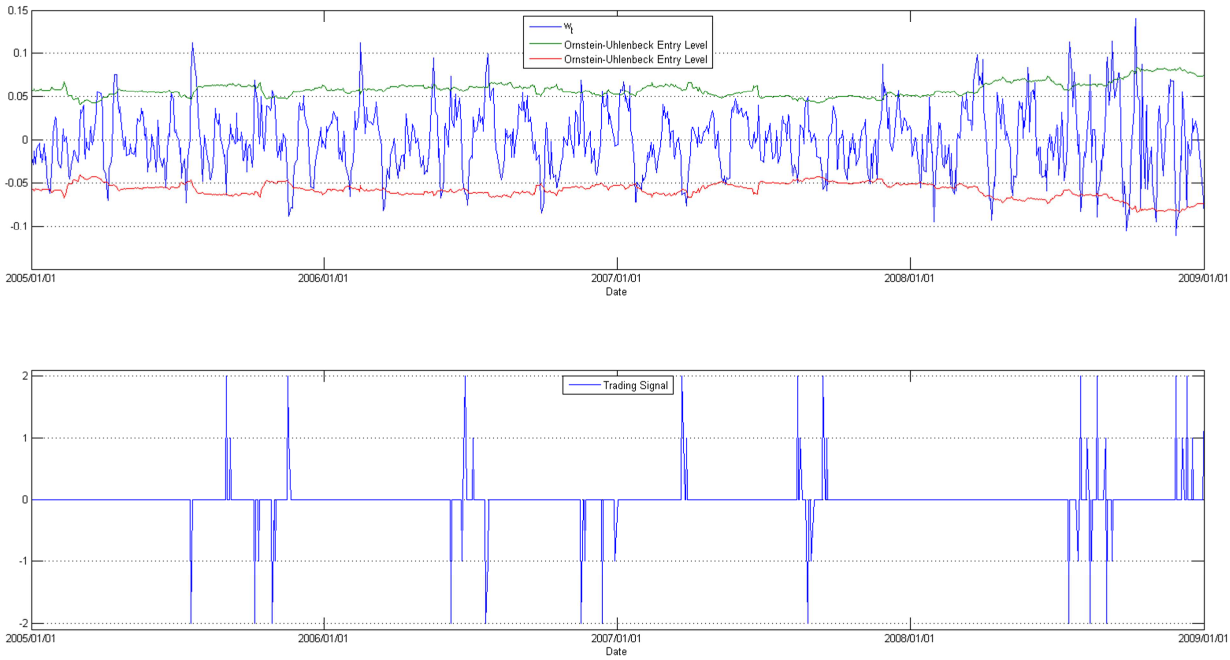


Figure 4-3 (a) Differenced residual $w_t$ as well as the Ornstein-Uhlenbeck entry levels and (b) trading signals.

Looking at the signals, one observes that positions are typically held only for a few days and are sometimes even closed out on the very next day. Holding periods are short and trading frequencies are low when looking at a single pair through time.

The above information is generated for each pair in the universe of pairs. At every point in time (i.e. every day), the in-sample number of zero crossings are tallied. Pairs are then sorted according to the number of zero crossings, with a higher number of zero crossings indicating (theoretically) stronger mean reversion. To construct the portfolio, trading signals are taken from the top 20 sorted pairs to form the statistical arbitrage portfolio. The portfolio-level results are discussed next.

## 4.3 Portfolio-Level Results

This section discusses portfolio-level results. Transaction costs play a significant role in statistical arbitrage portfolios due to the high turnover in the portfolio. Because of this, three scenarios are analysed: a scenario where transaction costs are set equal to zero, a scenario where transaction costs are set equal to the costs used in U.S. academic papers and a scenario where transaction costs are adjusted to realistically reflect a South African context.

### 4.3.1 Scenario 1: Zero Transaction Costs

Figure 4-4 below indicates the capital line of the statistical arbitrage portfolio when transaction costs are set to zero in the final calculation. The portfolio starts off with a capital amount of 100 units. The lower pane of figure 4-4 indicates the portfolio's daily returns.

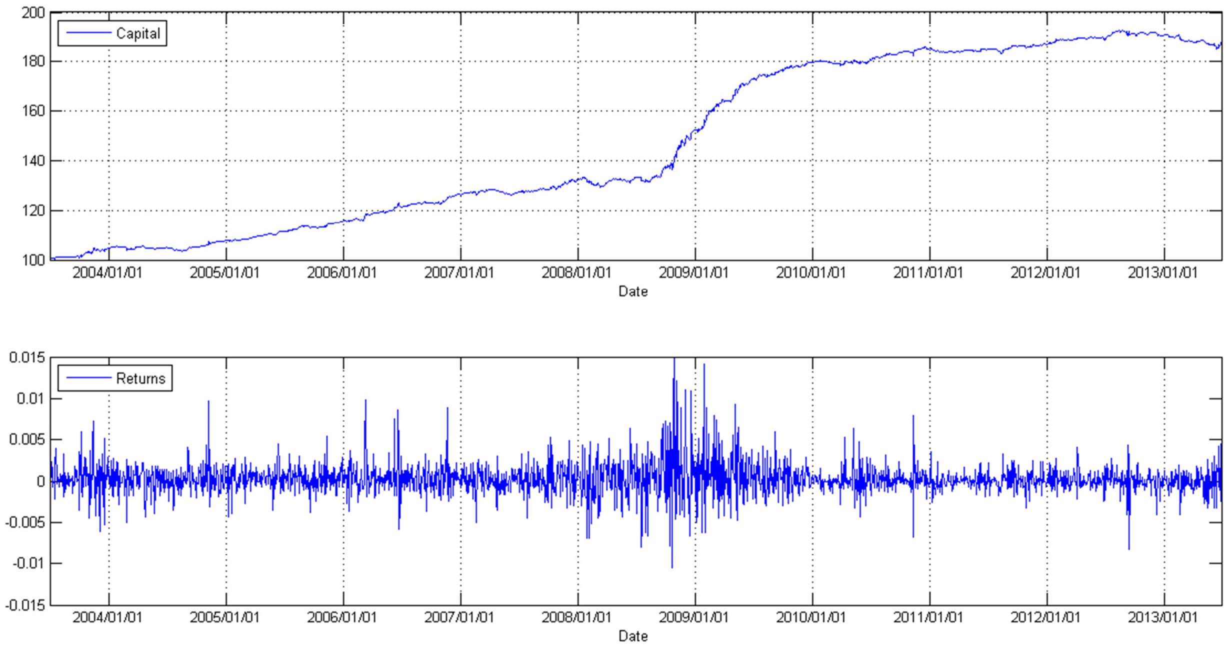


Figure 4-4 (a) Portfolio capital and (b) daily returns with zero transaction costs.

It is interesting to analyse the shape of the capital line over time. Between 2003 and 2008 the capital line increased at a steady pace.

2008/2009 clearly brought increased profitability for the portfolio, driven by increased volatility (as evidenced by the large return volatility). This coincides with the global credit crunch, confirming the findings of Do and Faff (2010) who reported significantly increased performance in times of turbulence, like the bear market of 2007-2009. Do and Faff attribute this performance to a less efficient market.

From 2010 onwards, the capital line still increases, but at a slower pace. It would be interesting to investigate statistically whether the performance from 2010 onwards is significantly different from prior periods. It is noted as a point for further research.

The table below provides information on the portfolio's performance.

| | |
|---|---|
| Transaction costs | 0% |
| Initial value | 100 |
| Final value | 186.95 |
| Most positive daily return | 1.62% |
| Most negative daily return | -1.05% |
| Mean daily return | 0.00025190 |
| Standard deviation of daily returns | 0.0019 |
| Sharpe Ratio[2] | 2.1060 |
| Maximum Drawdown (%) | 3.80% |

Table 4-1: Portfolio performance with zero transaction costs.

The Sharpe ratio exceeding 2 is encouraging, as is a maximum drawdown of only 3.8%. These figures would indicate that the portfolio can be leveraged significantly.

[2] The risk free rate was assumed to be zero.

### 4.3.2 Scenario 2: Academic U.S. Transaction Costs

When transaction costs are set equal to 0.05% (5 basis points), as was done by Avellaneda and Lee (2010), the profitability of the portfolio decreases as can be seen in figure 4-5 below. Once again the portfolio starts off with a capital amount of 100 units. The lower pane again indicates the portfolio's daily returns.

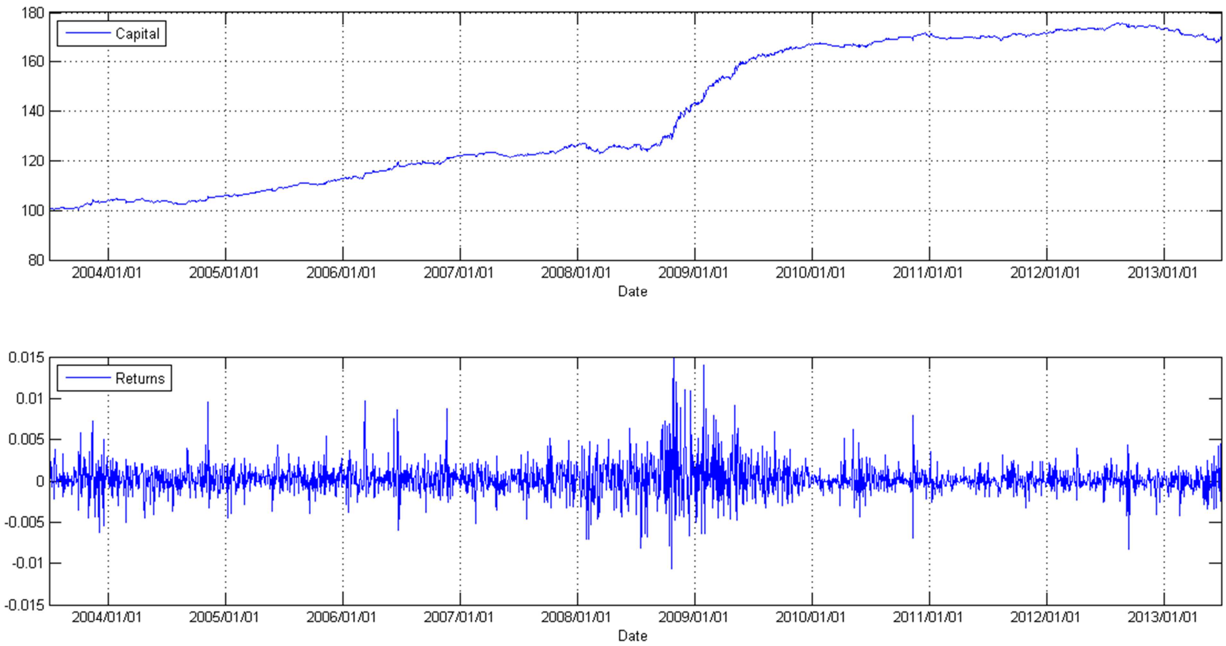


Figure 4-5 (a) Portfolio capital and (b) daily returns with 5 basis points transaction costs.

The table below provides information on the portfolio's performance.

| | |
|---|---|
| Transaction costs | 0.05% entry/exit (0.1% round trip) |
| Initial value | 100 |
| Final value | 169.10 |
| Most positive daily return | 1.62% |
| Most negative daily return | -1.07% |
| Mean daily return | 0.00021181 |
| Standard deviation of daily returns | 0.0019 |
| Sharpe Ratio[3] | 1.7716 |
| Maximum Drawdown (%) | 4.6% |

Table 4-2: Portfolio performance using academic U.S. transaction costs.

As expected, transaction costs decreased the performance of the portfolio. The size of transaction costs as was used by Avellaneda and Lee (2010) might be realistic for U.S. equities, but they are not realistic in a South African context.

[3] The risk free rate was assumed to be zero.

### 4.3.2 Scenario 3: South African Transaction Costs

Through consultation with a local fund manager, a more realistic figure of 25 basis points (50 basis points round trip) is used in the next analysis.

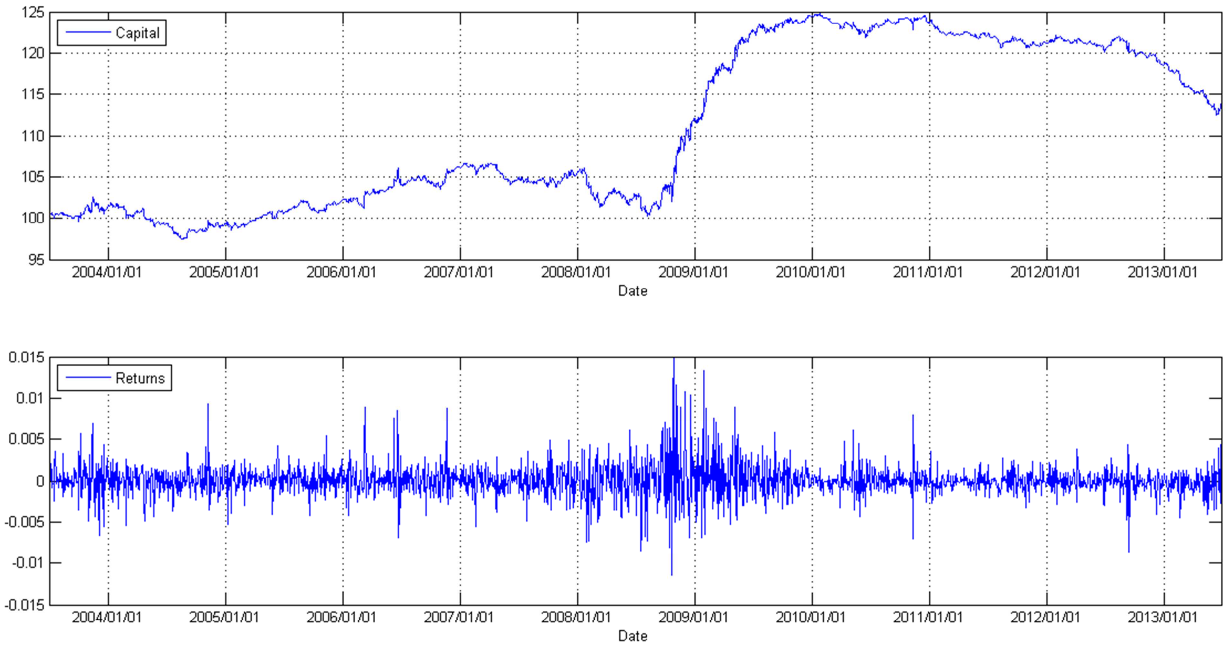


Figure 4-6 (a) Portfolio capital and (b) daily returns with 25 basis points transaction costs.

| | |
|---|---|
| Transaction costs | 0.25% entry/exit (0.5% round trip) |
| Initial value | 100 |
| Final value | 113.22 |
| Most positive daily return | 1.61% |
| Most negative daily return | -1.11% |
| Mean daily return | 0.00005151 |
| Standard deviation of daily returns | 0.0019 |
| Sharpe Ratio[4] | 0.4299 |
| Maximum Drawdown (%) | 9.87% |

Table 4-2: Portfolio performance using realistic South African transaction costs.

The difference between zero transaction costs (scenario 1) and realistic South African transaction costs (scenario 3) highlights the effect that transaction costs have on a statistical arbitrage trading system. Due to its high turnover and low profit margins, transaction costs have a significant impact on profitability.

[4] The risk free rate was assumed to be zero.

In figure 4-6, it is interesting to look at the shape of the capital line again. The period between 2003 and 2008 brought very little returns. Positive performance is again evident in 2008/2009 during the credit crisis. From 2010 onwards, the portfolio loses money.

## 4.4 Chapter Summary

This chapter discussed empirical results of the statistical arbitrage system. By studying a single pair, it demonstrated some of the inner workings of the system. Portfolio-level results were discussed using three different sets of transaction costs.

# Chapter 5: Conclusion

## 5.1 Conclusion

The purpose of the quantitative research conducted in this report was to answer a very narrow question that was laid out at the start: *Would this statistical arbitrage portfolio have delivered good risk-adjusted returns in a South African context over the last 10 years, as measured by the Sharpe ratio?* In theory, it would have. In practice, it didn't. Transaction costs had a major impact on the profitability of the system. Theoretically this portfolio shows potential, but when realistic transaction costs are factored in, its performance is poor.

The system outlined here is not ready to be traded in the markets, but focused research will undoubtedly improve the performance of the system.

Areas for further research and possible improvements are mentioned in the next section.

## 5.2 Recommendations for Further Research

Firstly, the use of leverage in this system can be considered. This should reduce relative transaction costs.

Secondly, optimization of the input parameters (in-sample sizes a, b and c) can be considered. One should be mindful that one now crosses the threshold into data mining: reusing the same data set and choosing the best performing parameters. Statistical adjustments (like White's reality check) will need to be made when continuing down this path.

Thirdly, one can consider adjusting the system to make better use of the available capital, since this system did not expose all capital to the market at all times (see section 3.8).

When looking at the trading system in the absence of transaction costs, it would be interesting to find out whether the slower pace of portfolio growth between 2010 and 2013 was significantly different from prior periods. Could it be possible that, like in the U.S. after 2003, statistical arbitrage profits are dwindling in South African markets?

One could also investigate what the performance of this system is in higher frequency timeframes (e.g. hourly data versus daily data).

# Chapter 6: Bibliography


Avellaneda, M. and Lee, J.H. 2010. Statistical arbitrage in the US equities market. *Quantitative Finance,* Vol. 10:7, pp. 761-782.

Bertram, W. 2010. Analytic solutions for optimal statistical arbitrage trading. *PhysicaA: Statistical Mechanics and its Applications* 389:11, pp. 2234-2243.

Bookstaber, R. 2007. *A Demon Of Our Own Design.* New Jersey: Wiley.

Caldeira, J. and Moura, G. 2013. "Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy." [Online]. Available from: http://ssrn.com/abstract=2196391 [4 May 2013].

Do, B. and Faff, R. 2010. Does Simple Pairs trading Still Work? *Financial Analysts Journal*. Vol 66:4 pp. 1-13.

Duyvené de Wit, J. 2008. *Quantitative position sizing methods for risk-adjusted interday index trading on the JSE.* MBA Dissertation, University of the Free State.

Duyvené de Wit, J., de Alessi, A. and Moodliyar, L. 2012. Pairs trading in South Africa. *Semester project submitted for the subject Time Series Analysis.* University of Cape Town.

Enders, W. 2010. *Applied Econometric Time series*. New Jersey: John Wiley & Sons.

Gatev, E., Goetzmann, W. and Rouwenhorst, K. 2006. Pairs trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies,* Vol. 19:3, pp.797-827.

Govender, K. 2011. *Statistical Arbitrage in South African Financial Markets*. M.Phil Dissertation, University of Cape Town.

Hogan, S., Jarrow, R., Teo, M. and Warachka, M. 2004. Testing market efficiency using statistical arbitrage with applications to momentum and value trading strategies. *Journal of Financial Economics*, Vol. 73, pp. 525-565.

Hull, J.C. 2012. *Options, Futures and other Derivatives*. 8th Ed. Boston: Pearson.

Jegadeesh, N. 1990. Evidence of Predictable Behavior of Security Returns. *Journal of Finance.* Vol. 45:3, pp. 881-898.

Jolliffe I.T. 2002. *Principal Component Analysis, Springer Series in Statistics.* 2nd Ed. New York: Springer.

Jurek, J. and Yang, H. 2007. "Dynamic Portfolio selection in Arbitrage. *EFA 2006 Meetings paper*." [Online]. Available from: http://ssrn.com/abstract=882536. [4 May 2013].

Kaufman, P.J. 2005. *New Trading systems and Methods.* 4th Ed. New Jersey: John Wiley and Sons.

Khandani, A.E. and Lo, A.W. 2007. "What happened to the Quants in August 2007?" [Online]. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1015987 [4 May 2013].

Lay, D.C. 1997. *Linear Algebra and its Applications*. 2nd Ed. New York, Addison-Wesley.

Lehmann, B.N. 1990. Fads, Martingales and Market Efficiency. *The Quarterly Journal of Economics,* Vol. 105:1, pp. 1-28.

Murray, M.P. 1993. A Drunk and Her Dog: An Illustration of Cointegration and Error Correction. *The American Statistician*, Vol. 48:1, pp. 37-39.

Patterson, S. 2010*. The Quants*. London: Random House Business Books.

Pole, A. 2007. *Statistical Arbitrage: Algorithmic Trading Insights and Techniques.* New Jersey: John Wiley and Sons.

Roll, R. and Ross, S.A. 1980. An Empirical Investigation of the Arbitrage Pricing Theory. *The Journal of Finance* 35:5, pp. 1073-1103.

Ross, S.A. 1976. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory,* Vol. 13:3, pp. 341-360.

Stock, J. and Watson, D. 1988. Testing for Common Trends. *Journal of the American Statistical Association.* Vol. 83:4, pp. 1097-1107.

Tharp, V. 2008. *Definitive guide to position sizing.* NC: IITM.

Thorp, E. 2004. A Mathematician on Wall Street (Six-part series). *Wilmott Magazine.*

Triantafyllopoulos, K., Montana, G., (2011). Dynamic modelling of mean-reverting spreads for statistical arbitrage. *Computational Management Science.* Vol. 8:1-2, 23–49.