Equity Pairs Trading on the JSE Using Single Stock Futures

Sean Paine

A research report submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Business Administration

Johannesburg, 2012

ABSTRACT

This research report investigates the key factors which contribute to the risks and potential profits of pairs trading. It then expounds on the relationship between these key factors and the success of pairs trading. Although previous research has investigated the success of pairs trading in general, this research delves into how pairs trading may be "fine-tuned" to produce the greatest risk-adjusted returns. It is also the first known study of this nature on the JSE.

The problem identified by the research is to define and analyse a hedge fund trading strategy based on the concept of pairs trading using cointegration. This includes determining the key independent variable parameters that are then used to define the pairs trading strategy, and then testing which values are optimal for the key independent variable parameters.

Data were collected from a secondary source, a commercial provider of daily stock market data. This was the only source of data for the research. The data were then run through an application which simulated trades based on certain quantifiable conditions. This software was written by the researcher.

The key findings of the research are that there are two key independent variables which can be used to influence the success of pairs trading. These variables are the cointegration t-statistic, a measure of how cointegrated a pair of shares are, and the opening trading signal variable.

The key message of the research is that by adjusting the two identified variables to suit the risk profile of a fund, it is possible to produce meaningful returns from pairs trading.

DECLARATION

I, Sean James Paine, declare that this research report is my own work except as indicated in the references and acknowledgements. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Signed at …………………………………………………

On the …………………………….. day of ………………………… 2012

DEDICATION

To Sonje

ACKNOWLEDGEMENTS

Thank you to my wife Sonje for her continual support, to the members of Wits Business School Whisky Club for their constant encouragement, and to Chris Muller, my research supervisor, who did not give up on me.

Thanks are also due to Dr. Teri Carmichael for her guidance on the structure and wording of the research, and to Wendy for her general encouragement and proofreading of the research.

TABLE OF CONTENTS

CHAPTER 3: RESEARCH METHODOLOGY 21

CHAPTER 4: PRESENTATION OF RESULTS 35

CHAPTER 5: DISCUSSION OF THE RESULTS.........................53

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Purpose of the study

The purpose of the study is to evaluate the effectiveness of various pairs-trading strategies on the JSE based on the principles of Arbitrage Price Theory and selection of pairs using cointegration. It builds on similar studies done by (Gatev, Goetzman, & Rouwenhorst, 2006).

The study proves whether or not pairs trading is profitable using the principles of cointegration, and if so, what the optimal trading parameters for the key independent variables are to make it so. Due to the secretive nature of most pairs trading strategies (Vidyamurthy, 2004), a strategy was proposed based on information gathered from the literature.

1.2 Context of the study

Pairs-trading is a trading strategy where two derivatives are traded simultaneously. In its simplest form, pairs trading consists of taking a short position on one derivative while going long on the other, in the hope that the short share will rise in price, while the long share will drop, at least relative to each other. At the end of the trading period, or when a pre-determined profit or loss has been realised, the positions on the futures are closed-out simultaneously (Do, Faff, & Hamzal, 2006).

Pairs trading dates back to the 1980's in the United States (Huck, 2008). A number of statistical, mathematical and trading theories have been applied to pairs trading with the intention of making it as profitable and risk-free as possible. Huck states that these methods include mean reversion, contrarian strategies, correlation, and cointegration. In addition, Kalman filtering, and principles of price arbitrage have also been used to drive pairs trading strategies (Vidyamurthy, 2004).

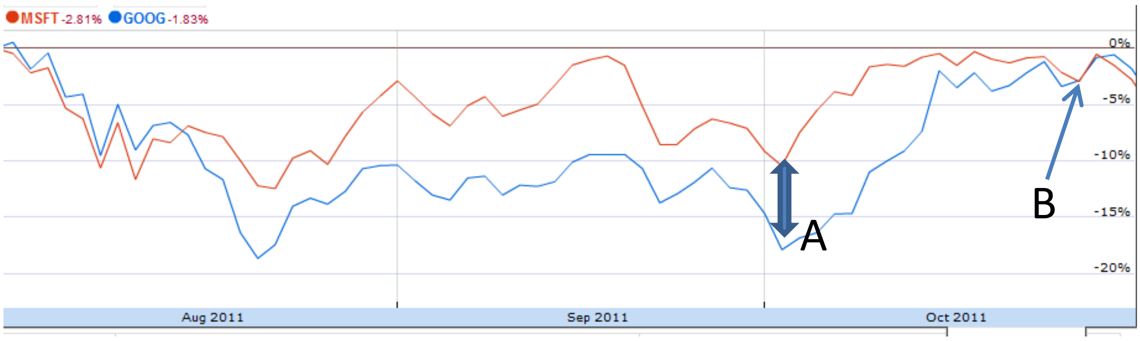

Figure 1 - An Example of Pairs Trading

The following gives an example of how pairs trading is executed. By analysing the share price history of the two shares, it can be determined that they are mathematically co-integrated. Only co-integrated pairs of shares should be traded.

Point A – Indicates the point where the relative distance measure between the two shares increases to a point where the decision to enter into a Pairs Trade is triggered.

GOOG, the lower of the two shares is purchased (a long trade), while MSFT, the higher of the pair is sold (a short trade).

Point B indicates where the two share prices converge. At this point, the trade is reversed, resulting in a profit. Even though the trader had gone short on the MSFT trade (which resulted in a loss), the overall trade was still profitable because the profit on the GOOG trade more than made up for it. This set of trades reduced overall risk whilst still returning a profit. This is one of the benefits of an effective pairs trading strategy.

This study focuses on the use of cointegration as a principle for creating a pairs trading strategy. Cointegration has been selected as it is one of the few principles that can be mathematically proven to theoretically "predict" changes in the relationship between share prices over time. Although correlation has been used for pairs trading, it is not sound theoretically, due to inherent issues with autocorrelation causing spurious correlations between share prices (Yule, 1926).

1.3 Problem statement

1.3.1 *Main problem*

Define and analyse a hedge fund trading strategy based on the concept of pairs trading using cointegration. This includes determining the key independent variable parameters that will be used to define the strategy, and then testing which values are optimal for the key independent variable parameters.

A stratified, deterministic simulation was run to test the optimal values for the key independent variables as identified in the literature. This method has been tested by a number of researchers, and is believe to be superior to the Monte-Carlo simulation method for high-dimensional finance problems. (Joy, Boyle, P., & Tang, 1996) , (Ackworth, Broadie, & Glasserman, 1997).

1.4 Significance of the study

The study fills a gap in current research into the practice of pairs trading. No studies exist on the effectiveness of pairs trading on the JSE. This study provides insight into both the effectiveness of the pairs-trading technique, and determines which parameters are most likely to be the most profitable and the least risky to the investor.

The study provides insight into the debate on market efficiency on the JSE. The strategy works on the principle of finding points of theoretical arbitrage in the current market. Opportunity for arbitrage is considered a factor in determining a market's efficiency (Simons & Laryea, 2005). As pairs trading is effectively about exercising an opportunity for arbitrage, if the JSE has become more efficient over time, then the results of this study hint at that. As this is not the main problem contained in the research, it is not directly studied, but this research nevertheless offers insights into this question.

The study provides guidance to hedge fund managers and academics interested in the concept of Arbitrage Price Theory, cointegration, the Law of

One Price and Pairs Trading (Vidyamurthy, 2004). This study assists hedge fund managers in understanding how to reduce the risk associated with any funds they may be running using the pairs trading approach. It assists academics in better understanding the nature of markets, and the extent to which share prices “move together” over time.

1.5 **Delimitations of the study**

1.5.1 ***Delimitations***

The study includes an investigation of the key independent variables used in the pairs-trading algorithm using all Single Stock Futures available on the JSE South African Futures Exchange (SAFEX) from inception of the market up to June 2010. Pairs trading is typically implemented by using financial derivatives as the trading instrument (Vidyamurthy, 2004). The chosen derivative for this study is the single stock future.

1.5.2 ***Limitations***

The daily closing stock price per single stock future, the cost of trade and trading volume were included as inputs in the study.

All other costs were ignored as they are immaterial. The trading cost was calculated as a percentage of the value of each trade using a pre-determined percentage. This cost did not change during the course of the study, and was based on the current trading cost of Single Stock Futures on the JSE.

Share beta, returns and the mean square average difference between share prices were calculated using the daily closing stock price. Vidyamurthy recommends that the volume-weighted average price be used as it produces a result which more closely mirrors the liquidity of the derivative in the market. However, for this study the daily closing price was used due to availability of data (Vidyamurthy, 2004).

The success of the various parameters applied to the trading strategy was measured by using the Sharpe Ratio. The Sharpe Ratio provides a measure of the risk-reward trade-off of financial performance of a portfolio by comparing returns to volatility (Anderson, 2009) (Bodie, Kane, & Marcus, 2009). A higher Sharpe Ratio indicates more favourable performance.

The nature of pairs trading is that it may be used with any derivative which allows the trader to take a long or a short position. The study is limited to shares linked to single stock futures available on the JSE.

Any trade containing a derivative where the underlying share either de-lists or splits during the course of the trade was closed out immediately. This is a reflection of what would occur in practice, and was the only way to take account of the fact that unpredictable events can result in the suspension of trade of a single stock future.

Pairs trading may be executed by combining more than two shares into a mix of long and short trades. This is referred to as multivariate pairs trading (Perlin, 2007). This research is limited to a bivariate pairs trading strategy, or the concept of having a maximum of two shares considered as being part of a "pair" at any one time.

1.6 Definition of terms

SAFEX: The South African Futures Exchange. An exchange for the trade of both agricultural and equity derivatives. It was acquired by the JSE in 2001 (Johannesburg Stock Exchange, 2010).

SETS: Stock Exchange Trading System. An electronic trading system used to ensure near real-time exchanges and transparency on the JSE (Standard Bank, 2010).

Share Valuation: There are two methods for pricing futures – relative and absolute pricing. Absolute pricing is the pricing of a company based on financial methods for valuing the company based on expected future performance. These methods include pricing a company based on discounted future cash flows, or options pricing (Vidyamurthy, 2004).

Relative pricing is based on the concept that companies are valued according to market substitutes. This form of pricing is immune to the effects of market cycles, as at least theoretically, companies which are substitutes for each other will both be equally affected by market movements, whether those are positive or negative. (Vidyamurthy, 2004).

Margin Call: A demand from a broker to settle a balance on an investment account. Depending on the performance of the underlying shares, a trade using single stock futures can result in a margin call if the net balance of the investment drops sufficiently (Finance Glossary, 2012).

Volume Weighted Average Price: An artificial price benchmark calculated by adding up the value traded on a particular share over the course of a day, and dividing it by the number of transactions. When used in financial modelling, this gives a fairer reflection of what the cost price would have been than using the closing share price (Madhavan, 2010).

Gearing: The ratio between the value of the investment and the amount of capital invested. An investment with high gearing is one where the investment has been structured using substantial debt relative to the size of the investment (Finance Glossary, 2012).

1.7 **Assumptions**

Availability and accuracy of share price information – reasonable, highly sensitive. Share price information was obtained electronically from a commercial source, and should therefore be accurate.

Transaction Cost Accuracy. Transaction cost for a futures trade consist of a Broker’s Commission, and “Market Maker’s fee”. No STRATE or UST fees are payable on a futures trade (Standard Bank, 2010). The study uses a transaction cost of 1% of the total value of each trade. Based on current retail pricing of single stock futures trades (Standard Bank, 2010), this is a reasonable and sensitive assumption. It is sensitive, because the transaction cost directly affects the financial returns from the fund. Changing the transaction cost could alter the results of this study.

Margin Calls: Margin Calls were honoured at the close of the trade of a particular pair of shares. The alternative was to honour the margin call at the end of each day's trade. The JSE takes the total cash position of the investor into account when determining margin calls (Johannesburg Stock Exchange, 2012). The fact that this strategy is cash-neutral and the stop-loss limitations are rigorous means that honouring margin calls on a daily basis is irrelevant to the outcome of the study. This is a reasonable and slightly sensitive assumption.

Interest on Margin Account is ignored. Due to the maximum length of time the money is held (weeks, not months). This is a reasonable and only slightly sensitive assumption.

Liquidity in the Futures Market: The initial assumption was that all futures are liquid, and a buyer or seller would be available at any point where the model requires a transaction. After some anomalous results from the simulation, this assumption was revised. The assumption became that all futures are able to be traded up to their actual historical maximum volume, but not over that value. This is a reasonable and slightly sensitive assumption.

It is important to note that it is on this assumption that the trading success or failure of the simulation model presented in this research rests. All other variables can be taken into account within a typological study, but the application of this model to a practical, real-world test rests heavily on whether or not there would be willing buyers and willing sellers to participate in transactions with the trader executing this strategy. It is beyond the scope of this research to speculate or build a model that takes trading liquidity into account to the point where the research could be expected to exactly match real-world conditions.

As per guidelines from the JSE, Initial Margin Deposit is set at 10% of the derivative being traded. For this study, exactly 10% will be used as the initial margin amount. This is a reasonable and moderately sensitive assumption (Johannesburg Securities Exchange, 2010).

Interest on the cash account is included as a part of the investment. Interest is at 3% per annum. This is a reasonable and moderately sensitive assumption.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter contains a literature review of the key areas of theoretical background knowledge required to understand pairs trading. The first part gives a background and definition of pairs trading, explaining the mechanics of the trading strategy. The second section details how the trading strategy has been applied in various experimental simulations. Finally, the review goes into more detail on the theories of time series analysis, and how cointegration has developed as a time-series-analysis theory since 1981 (Granger, 2004), (Engle & Granger, 2003).

2.2 Definition of topic or background discussion

2.2.1 *Definition of Pairs Trading*

Pairs trading is an investment strategy commonly used by financial institutions such as hedge funds and investment houses (Chng, 2007). The term covers all strategies which consist of the following general steps:

a) Identify a pair of derivatives which "move together". This relationship does not necessarily mean that the derivatives are correlated. In fact, correlation is only one method of determining whether the derivatives should be paired (Alexander, 2001).
b) When a divergence between the two derivative prices is noted, go short, or sell, the derivative which has risen in price relative to the other derivative, and go long, or buy, the derivative which has dropped (Vidyamurthy, 2004).
c) Hold this position until a pre-determined profit, loss or timeframe has been exceeded (Andrade, di Pietro, & Seasholes, 2005).

This strategy may be applied over a variety of time periods, from milliseconds to months (Nath, 2003).

Pairs trading is a speculative, often market-neutral strategy based on quantitative principles. Typically, pairs trading refers to the practice where a trader identifies a pair of equity derivatives which are believed to be relatively mispriced. The trader then applies the principles of Arbitrage Pricing Theory (Ross, 1976) by simultaneously “going long” on the share which is believed to be relatively under-priced, and “going short” on the share which is believed to be relatively over-priced. The trade is concluded after one of the following three criteria is met:

a) The two share prices return to a pre-defined distance, often the point of convergence, indicating that the relative mispricing has been resolved, and the opportunity for further profit has been reduced.

b) A pre-determined time period elapses, suggesting that perhaps the share prices were not relatively mispriced after all.

c) The futures continue to diverge to such a point that it is no longer financially prudent or possible to maintain the position.

Pairs Trading is a medium- to short-term trading strategy, with some practitioners engaging in the trade for as little as milliseconds to minutes, (Nath, 2003) to the more commonly practiced period of days to months (Gatev, Goetzman, & Rouwenhorst, 2006).

Pairs trading as a strategy has been used since the 1980’s by various traders and hedge fund managers. Nunzo Tartaglia, a Wall-Street analyst working for Merrill Lynch, is commonly attributed to having assembled a team who pioneered work in the field of arbitrage trading, including pairs trading. Since then, the strategy has been widely adopted by many traders. Although the strategy has become prolific, the exact implementation of the strategy tends to vary greatly (Gatev, Goetzman, & Rouwenhorst, 2006; Vidyamurthy, 2004; Perlin, 2007).

2.2.2 *Application*

Due to the proprietary nature of the various pairs trading strategies, the majority of the details for application of Pairs Trading are from the academic literature. Exact details of Pairs Trading techniques tend to remain a closely-guarded secret by those who practice them commercially (Alexander, 2001).

Academic studies into Pairs Trading are relatively recent phenomena. The most-referenced studies include those by Gatev, Goetzman, & Rouwenhorst (2006), Vidyamurthy (2004), Herlemont (2004), Elliot, van der Hoek, & Malcolm (2005), and Do, Faff, & Hamzal (2006).

The studies tend to vary in nature with regard to implementation technique. Broadly, however, all researchers performing a study into a pairs trading strategy are confronted by a common set of issues:

a) Identification of tradable pairs
b) Trading time period
c) Ratio of short to long derivative value
d) Signal Identification for opening a trade
e) Signal Identification for closing a trade

In practice, they may be confronted by the following:

a) Methods to raise debt capital
b) Credit to cover margin payments that may be exceeded during a trade.
c) Relationships with owners of equity capital when and if the fund is under pressure from a profit-making point of view (Nath, 2003).

Gatev, Goetzman, & Rouwenhorst (2006) and Nath (2003) both use the “distance method” to identify tradable pairs. This method attempts to quantify the traders’ strategy for pairs trading where one identifies futures which “move together”. This entails identifying pairs which “minimise the sum of squared deviations between them." This method is detailed in the formula below, where *a* and *b* represent the two normalised share price series.

$$\sigma = \sum_{t=1}^{T} (\mathrm{a_t} - \mathrm{b_t})^2$$

In addition to using this measure, Gatev, Goetzman, & Rouwenhorst (2006) presented results by sector, restricting the pair to stocks which belong to the same categories as defined by Standard and Poors. This is a common theme in the literature, where researchers do not purely accept the statistic reference as a signal to enter the trade, but rather seek to avoid futures which may indicate co-movement or cointegration from a statistical point of view, but fundamentally are highly unlikely to be driven by the same economic factors (Gatev, Goetzman, & Rouwenhorst, 2006; Vidyamurthy, 2004).

Gatev, Goetzman, & Rouwenhorst (2006) is in the minority, however, when it comes to using this method for finding tradable pairs. A far more common approach in the literature is to identify pairs of futures which exhibit cointegration. Various statistical tests may be used to identify cointegration between two price series. Vidyamurthy and Herlemont both use the Engle-Granger cointegration test statistic to determine tradability.

The cointegration test as detailed by Engle and Granger is a relatively time-consuming procedure, requiring a large amount of computing power. In the interests of efficiency, Vidyamurthy proposes that one should first screen out unsuitable futures before applying the Engle-Granger test. Vidyamurthy suggests that instead of testing all possible share combinations to find cointegrated pairs, the trader should first rank the pairs by using a distance measure he or she proposes. This distance measure is a measure of the degree of co-movement between futures, represented by the following formula.

$$\mathrm{d}(A,B) = |\rho| = \left| \frac{cov(r_A, r_B)}{\sqrt{var(r_A)var(r_B)}} \right|$$

(Vidyamurthy, 2004, p. 94)

However, Vidyamurthy does concede that it is better to eliminate the use of this formula when it is possible to apply the Engle-Granger two-step cointegration test to the entire population of futures. Herlemont (2004) and Do, Faff, & Hamzal (2006) both concur, with neither of them using the distance measure as an indicator of co-movement.

ECM for trading where A is the first share, and B is the second share.:

$$\log(p_t^A) - \log(p_{t-1}^A) = \alpha_A \log(p_{t-1}^A) - \gamma \, log(p_{t-1}^B) + \varepsilon_A$$

$$\log(p_t^B) - \log(p_{t-1}^B) = \alpha_B \log(p_{t-1}^A) - \gamma \, log(p_{t-1}^B) + \varepsilon_B$$

(Vidyamurthy, 2004, pg. 18)

2.2.3 *Trading Summary Strategy Summary*

This matrix gives a summary of the various studies into pairs trading. It shows the researchers involved, the method used for identification of derivative pairs, the period over which the trades are conducted, the distance between normalised derivative prices used to signal the opening of a trade, and the signal used to determine the closing of a trade.

Researchers	**Pairs Identification**	**Trading Time Period**	**Signal to Open**	**Signal to Close**
(Do, Faff, & Hamzal, 2006)	Engle-Granger cointegration test	Days to Months	Two standard Deviations	Convergence to 0
(Gatev, Goetzman, & Rouwenhorst, 2006)	Sum of Squares Distance Measure	Days to Months	Two standard Deviations	Convergence to 0

Researchers	Pairs Identification	Trading Time Period	Signal to Open	Signal to Close
(Herlemont, 2004)	Engle-Granger Cointegration test	Days to Months	Two standard Deviations	Convergence to 0
(Nath, 2003)	Only US Treasury Securities with sufficient liquidity	Seconds to Minutes	Variable percentile of distance	Variable percentile of distance, or end of trading period
(Vidyamurthy, 2004)	Engle-Granger cointegration test combined with co-movement statistic	Days to Months	0.7 standard Deviations	0.7 standard deviations in the opposite direction

Table 1 - Summary of Trading Strategies

2.2.4 *Table of Critical Values*

The Table above is a set of critical values applied to various methods for determining cointegration. The study uses the Advanced Dickey Fuller method (indicated by ADF in Table 2). Therefore a value of 2.91 had been selected as the critical value for the ADF t-statistic for this trading strategy. However, as a part of the research, the actual values for cointegration were calculated, and a different set of critical values were determined. This is discussed further in the Description of Results Chapter.

	Critical Values			
Statistic	Name	1%	5%	10%
1	CRDW	.455	.282	.209
2	DF	3.90	3.05	2.71
3	ADF	3.73	3.17	2.91
4	RVAR	37.2	22.4	17.2
5	ARVAR	16.2	12.3	10.5
6	UVAR	59.0	40.3	31.4
7	AUVAR	28.0	22.0	19.2

Table 2 - Critical Values

(MacKinnon, Haug, & Michelis, 1995)

2.3 **Factors Affecting Pairs Trading**

The research problem has been defined as:

Define and analyse a hedge fund trading strategy based on the concept of pairs trading using cointegration. This includes determining the key independent variable parameters that will be used to define the strategy, and then testing which values are optimal for the key independent variable parameters.

Based on the literature, the research proposition is that there exists a set of key independent variables which can be controlled to produce the most profitable, least volatile returns using pairs trading as a trading strategy. These key independent variables are:

1) The squared distance at which the trade is opened
2) The ADF t-test cointegration score

The research problem is defined along the following topics, or discussion points:

What are the key independent variables one should control when running a pairs trading fund to produce the best results? Given that time series analysis does not respond well to methods that directly test for the correlation of stock prices over time (Yule, 1926), what other options are available? The literature suggests that cointegration is a reliable method for determining the relationship of share prices over time. If this is the case, which method of determining cointegration should be used for the study?

Once the key independent variable parameters are established from the literature, what are the best values for those parameters to produce the best results? The best results are defined as those which return the most favourable Sharpe Ratio over the duration of the study. As this study is based on the principles of arbitrage pricing theory, and arbitrage pricing theory can be used to determine whether a market is efficient (Huberman & Zhenyu, 2005), will this study reveal any shifts in market efficiency over the life of the study?

2.3.1 *Trading on the JSE*

The efficiency of the JSE changed to become more strong-form efficient with the introduction of STRATE (Share Transactions Totally Electronic) in 1998 and SENS (Stock Exchange News Service) in 1997 (Mabhunu, 2004). This efficiency was further bolstered by the scrapping of JET (Johannesburg Equities Trading) replaced by SETS (Securities Exchange Trading System) on May 13, 2002 (Mkhize & Msweli Mbange, 2006). As the study is based heavily on principals of arbitrage pricing theory, testing a problem on a market with a different efficiency to the one we currently experience may produce spurious results. This is not a factor in this study, as the trading period starts after the events noted above.

2.3.2 *Time-Series Analysis*

A time series is a series of values measured over time (Vidyamurthy, 2004). The values are either arrived at via either a deterministic or probabilistic formula. This research is concerned with the theoretical implications of the probabilistic time series. Shares trading data is considered probabilistic.

When engaging in time-series analysis, the starting question is whether there is a relationship between the current value and the value one step previous. If a relationship exists, this is referred to as autocorrelation. It is identified by measuring the correlation between time series one value apart (Vidyamurthy, 2004). The correlation is estimated by using the formula below:

$$\hat{p}(\tau) = \frac{\frac{1}{T}\sum_{t=\tau+1}^{T}[y_t - \bar{y}][y_{t-\tau} - \bar{y}]}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Autocorrelation allows the analyst to identify various types of time series, and to understand that even though a time series may exhibit correlation, it would be incorrect to assume that the time series itself has a trend. It was with this in mind that Engle and Granger devised the theory of cointegration.

Stock prices in general have been shown to exhibit a random walk. (Vidyamurthy, 2004). Mathematically, any time series which follows a random walk is impossible to predict consistently, as it is a non-stationary time series. The random walk is defined as:

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots = \varepsilon_t + \gamma_{t-1}$$

Although the logarithm of a share price time series may be non-stationary, the differences between two different share prices often exhibit a stationary behaviour. This pattern may be identified using a test for cointegration (Vidyamurthy, 2004).

2.3.3 ***Cointegration***

Cointegration is a commonly-used measure for the identification of pairs. For this research, cointegration testing is used to establish the suitability of two derivatives for a pairs trade.

The first step when performing a pairs trade is the selection of the two futures which will be used for the actual trade. There are a number of means for selecting these futures. In the literature, pairs are usually selected by reviewing their relative past performance (Do, Faff, & Hamzal, 2006), (Gatev, Goetzman, & Rouwenhorst, 2006), (Vidyamurthy, 2004).

The method for actual selection tends to differ throughout the literature, and some traders prefer to look for futures which are highly correlated, either positively or negatively. However, the bulk of the research suggests that a better approach is to look for futures that are not correlated, but rather cointegrated (Vidyamurthy, 2004).

The literature shows that a number of methods may be used to prove whether a pair of time series data exhibits cointegration. Multivariate series may also exhibit cointegration, but this is beyond the scope of this research. This research is limited to a bivariate pairs trading strategy.

There are a number of methods discussed in the literature for the construction of the cointegration statistic. These include the Engle-Granger, Phillips, Johansen and Stock Methods (Vidyamurthy, 2004).

Based on the methods most commonly used, this research will focus on the Engle-Granger cointegration test using the Cointegration Regression Augmented Dickey-Fuller (CRADF) method to determine the level of cointegration. The critical values for this test have been determined in previous literature (MacKinnon, Haug, & Michelis, 1995).

An Ordinary Least Squares regression must be estimated with stationary data and residuals. If not, the regression is spurious. The data in the time series must therefore be tested for a unit root to determine whether it is stationary. If both time series are I(1) (non-stationary), and the regression produced is an I(0) error term, then the two series are cointegrated (Agarwal & Narayan, 2004).

A time series that is identified as a random walk is therefore a non-stationary time-series. The Dickey-Fuller Test tests for the presence of a random walk:

$$y_t = y_{1-1} + u_t$$

Where y represents the value of the item being tested, and u represents the expected error.

However, the Dickey-Fuller test suffers from false positives due to autocorrelation. To overcome this problem, the Augmented Dickey-Fuller Test was introduced. By adding lagged dependent variables, Dickey-Fuller arrived at the following formula:

$$\Delta y_t = (\rho - 1)_{y_{t-1}} + \alpha_i \sum_{i=1}^{m} \Delta y_{t-i} + u_t$$

According to Noriega and Ventosa-Santaularia, the ADF method is preferred over the DF method to eliminate spurious cointegration (Noriega & Ventosa-Santaularia, 2006). The full Cointegration Regression Augmented Dickey Fuller formula is presented below:

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

$$H_0: \gamma = 0$$

$$H_1: \gamma > 1$$

Where $\hat{\gamma}$ is the is the alpha (intercept) of the regression of the residuals calculated in the standard Dickey Fuller Test.

2.4 Conclusion of Literature Review

The literature review covered the definition of pairs trading, a strategy where two shares are traded in conjunction with each, the aim of which is to realise a net profit by the expected relative change in price between the two shares. As the trade involves going short on one of the shares, it is necessary to use derivatives to facilitate the trade. Single stock futures are one of the derivatives which may be used for pairs trading (Chng, 2007).

There are a number of ways to determine the suitability of pairs for trading. Cointegration has been identified as a quantitative method of determining suitability. There are a number of ways to test for cointegration. The Engle-Granger cointegration test is one of the most common (Engle & Granger, Co-integration and Error Correction: Representation, Estimation and Testing, 1987).

Several teams of researchers have created simulated models to study the performance of pairs trading. The list of researchers is summarised in Table 1 in Section 2.2.3.

A number of factors affect the success of pairs trading. These include the squared distance at which the trade is opened, and the cointegration t-test score determined by using the Engle-Granger cointegration test. Previous research has tended to use a static value for these variables which may affect the success of a pairs-trading fund. No published research has been done on the effect of tweaking these key independent variables to produce the most optimal trading strategy on the JSE. This discovery from the literature creates the base for the research proposition.

2.4.1 ***Research Proposition***

Based on the literature, the research proposition is that there exists a set of key independent variables which can be controlled to produce the most profitable, least volatile returns using pairs trading as a trading strategy. These key independent variables are:

1) The squared distance at which the trade is opened.
2) The ADF t-test cointegration score. The score shows the extent to which the pair of shares displays a level of cointegration. As the ADF t-test is a null-hypothesis test ($H_0 : \gamma = 0$)), correctly stated, the ADF t-test gives an indication of the chances that two shares are not co-integrated.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Research Paradigm

The research methodology is a time-series analysis of secondary financial data using a deterministic simulation to test performance of a trading strategy by altering key independent variables. A trading strategy was simulated using the closing prices of single stock futures over a seven-year period. The trading strategy was simulated multiple times, always with the same deterministic rule set, but with different input parameters for the key independent variables.

The results of the various trading strategies were then analysed and classified in terms of both return and volatility over time. The most favourable trading strategy was then identified.

The research was conducted in order to explain the relationships between several of the key independent variables involved in the concept of a pairs-trading cointegration strategy. The research was not designed to prove or disprove the concept of pairs trading itself. This type of research is therefore classified as *nomothetic.* (Bennett, 1995), the aim of which is to partially explain the success factors of pairs trading by means of a causal model (Groenewald, 1989).

According to the Groenewald Matrix, (Groenewald, 1989), a nomothetic explanation from documentary data sources is therefore classified as a *typological study*. Groenewald states that the aim of a typological study is to establish a theoretical construct or model which may be used in comparison of empirical cases.

The theoretical construct in the case of this research is a universe where single stock futures are traded on a market where prices are historically the same as the SAFEX. The theoretical construct sets out to determine that there exists a set of parameters which may be optimised to produce the highest possible returns from a pairs trading strategy. This research is an empirical study analysing past data by applying a trading strategy to that data. The results of the trading strategy are then statistically analysed to determine which input parameter values produce the most favourable outcome.

3.2 Research Design

The time series analysis was done using data from the inception of retail single stock futures in the South African market. Single stock futures were introduced to the South African market in 1999 (Standard Bank, 2010).

For each of the 30+ trading quarters under review, the following actions were performed:

1. Find futures that are cointegrated, and should therefore be paired.
2. Run through all combinations of parameters for that trading period, using the futures identified in 1), and analyse which parameters produce the most profitable portfolio using the trading rules for pairs trading. The first two parameters were tested in conjunction. They are the:
    a. Distance between the two futures prices required for the opening of a trade (The opening standard deviation)
    b. The cointegration t-statistic of each pair
3. Compare the various trading strategies using the different input parameters, or key independent variables to determine which portfolio produces the most desirable Sharpe Ratio.
4. Describe the results of the various simulations to give more insight into the validity of the assumptions made in the study.

The details for these steps follow in the rest of the chapter:

3.2.1 *Identify Pairs*

The Engle-Granger Method for cointegration was used. Identification of pairs was performed across all possible single stock future combinations. For example, all single stock futures that closed out in March 2010 were measured against all other single stock futures that closed during the same time period. This process was repeated across all time periods.

The Engle-Granger two-step method is used for testing for cointegration between two or more non-stationary time series. It consists of running an Ordinary Least Squares (OLS) regression, and then using the resulting residuals as inputs into either the Dickey-Fuller test, or the Augmented Dickey-Fuller test to determine if they are stationary. The time series is said to be cointegrated if the residual is itself stationary (Engle & Granger, Co-integration and Error Correction: Representation, Estimation and Testing, 1987).

(Noriega & Ventosa-Santaularia, 2006) recommend that the Augmented Dickey-Fuller (ADF) test and not the Dickey-Fuller test should be used to test the residuals from the first step of the Engle-Granger two-step method. The ADF test has been proven by (Elliot, Rothenberg, & Stock, 1996) to be the most uniformly powerful test for the analysis of residuals in this type of problem The ADF test was therefore used. This test includes the lagged residuals to avoid errors from residual correlation (Herlemont, 2004). The end result of the ADF test is a t-statistic. Pairs with the highest negative ADF t-statistics are considered the "most" co-integrated. This research set out to prove whether pairs that are more co-integrated are more likely to be profitable and stable than pairs that are less co-integrated.

The step-by-step process for the Augmented Dickey Fuller test is:

1. Reduce the distance between futures using the Solver algorithm built into Microsoft Excel. Given futures a and b, the objective is to reduce the value of σ to the minimum possible. By solving for the lowest possible σ, a ratio β will be created which will be used to adjust the trading price of future b. This ratio is known as the cointegration Beta.

$$\sigma = \sum_{t=1}^{T} (\mathrm{a_t} - \beta \mathrm{b_t})^2$$

2. Use the result from 1) to adjust the share futures prices time series data. In other words, the share price of share $\mathrm{b_t}$ is multiplied by the Beta (β).

3. Apply the natural logarithm to the adjusted share futures prices time series data.

4. Run an Ordinary Least Squares regression on the results from 3).

5. Using the results of the regression, Subtract (α from (β times ln(a)) from ln(b)

$$\ln(a) - (\beta(\ln(b) - \alpha)$$

The results of this formula are the residuals of the time series.

6. Run an Ordinary Least Squares regression on the residuals

7. Calculate the ADF t-statistic on the results of the regression. The Augmented Dickey Fuller formula for the t-statistic is:

$$\frac{\alpha}{(se)}$$

where α is the alpha (intercept) of the regression of the residuals, and se is the standard error of the regression of the residuals.

8. Originally, the research planned to compare the t-statistic to the critical values identified by MacKinnon (MacKinnon, Haug, & Michelis, 1995). A t-statistic of less than -3.5 was sufficient for a pair of futures to be considered cointegrated. (Alexander, 2001). Due to the number of pairs which were analysed, the critical value was instead determined by calculating the critical values based on the results of analysing the results of cointegration calculations across the data set. Critical values at the 95th percentile were calculated. This is explained in more detail in Chapter 5.

3.2.2 ***Execute Trades***

The trading rule that was applied was that if the two selected pairs diverged by more than the opening standard deviation, an opening trade was executed, providing the two shares were not already currently “in a trade” with each other. The opening trade consists of taking a long position on the lower-priced future, and taking a short position on the higher-priced future in equal ratios.

This position was held until the futures prices were less than or equal to the closing distance, or the end of the three-month trading period where the futures were closed out, and the trade ended automatically. In addition, the trade was closed out if the two share prices diverged by more than 5 times the standard deviation of the sum of the squares of the average distance between the two shares.

The closing distance for all trades was zero. A zero closing distance was selected as reasonable, based on this being the standard for prior pairs-trading studies. These include the studies by (Do, Faff, & Hamzal, 2006), (Gatev, Goetzman, & Rouwenhorst, 2006) and (Herlemont, 2004).

3.2.3 ***Optimise the first two parameters with all cointegrated futures.***

The data were then trimmed down to include only pairs that have been identified as having a t-statistic of less than -1.5. A value of -1.5 is far more than the critical value for cointegration. As a result, many pairs were included that did not show a tendency to cointegration. This was important to prove whether or not the degree of cointegration had an effect on the average profitability of a particular pair of shares. The next step was to optimise the key independent variable of the trading strategy. The test was run for the opening standard deviation (Parameter 1). The test was run with values between .5 and 2.5 with intervals of 0.5 for a total of five variables.

The most effective combination of parameters was deemed as those that produce the highest Sharpe Ratio. The Sharpe Ratio is defined as

$$S = \frac{R - R_f}{\sigma} = \frac{E[R - R_f]}{\sqrt{\mathrm{var}[R - R_f]}}$$

Where:

R is the asset return, R_f is the return on the risk free rate, $E[R - R_f]$ is the expected value of the excess of the asset return over the risk free rate, and σ is the standard deviation of the excess of the asset return (Sharpe, 1994). Although Sharpe states that this formula is not perfect, it is at least a measure that brings both return and risk to account when attempting to determine the success of a portfolio.

3.2.4 ***Cash***

The method assumed a working pool of R1 000 000 in cash. If the portfolio went into bankruptcy (achieved a balance of less than zero), the fund was considered a failure, and simulated trading ended as of that point.

3.2.5 ***Stop Loss***

If the mean square distance $(a_t - \beta b_t)^2$ exceeded 5 times the standard deviation (σ), the trade was closed to prevent the one trade from destroying the entire value of the portfolio. Due to the geared nature of single stock futures, tremendous losses, orders of magnitude greater than the original investment, can be created from just one trade.

3.2.6 ***Total list of key independent variables:***

The total list of key independent variables identified was:

1. The opening standard. deviation
2. The cointegration t-statistic

3.2.7 ***List of Portfolios and other variables***

The gearing was set based on general guidelines from the JSE. It was set to 10% of the value of the derivative. This is a deviation from reality, where derivative gearing is based on volatility and liquidity of the underlying share. As this information was not available, and the gearing ratio tends to have a median of 10%, the number was set as such (Standard Bank, 2010).

The quantity of single stock futures purchased was set to one for the first of the two shares in the share-pair. The quantity of the second pair in the transaction is equal to the rounded value of the beta calculated by Solver in Step 1 of 3.2.1 if beta is greater than 1, and 1 divided by beta if beta is less than 1. For example if the beta is calculated at 0.2, then 5 shares (1 divided by 0.2) would be used for the trade. This produced a close to cash-neutral trade.

3.2.8 *Conclusion*

The various portfolios were benchmarked against each other using the Sharpe Ratio as the measure of success.

3.3 Population and sample

3.3.1 *Population*

The population was all closing prices of all single stock futures available on the JSE since inception in 2000.

3.3.2 *Sample and sampling method*

The sample data were made up of the daily closing prices of all single stock futures linked to shares on the JSE from inception of single stock futures on the SAFEX to June 2010. The sample is all single stock futures traded from October 2003 to December 2011.

3.4 The research instrument

The research instrument is a deterministic simulation model designed and developed by the researcher. The first component of the simulation model determines the level of cointegration between each share pair on the SAFEX. This part of the model is important to the research, because it calculates the cointegration statistic, which is used as one of the key independent variables for the study.

The second component of the simulation model is the deterministic “engine” which processes all trades as they would have occurred historically, day-by-day, pair-by-pair, using the input parameters as determined from the literature review. The output of this component (the profit or loss of each trade) , when combined with the input parameters, produced a data set which was analysed to determine which combination of input parameters produced the most profitable, least volatile returns over the historic period of 2004 to 2011.

The technical tools used for creation of the instrument were Microsoft SQL Server, a referential database, Microsoft Excel 2010, a calculation engine with broad application, and QlikView, a data analysis and data presentation tool produced by Qliktech.

The instrument was designed and written by the researcher. The pros to this approach included the following:

- The instrument was custom-built to the exact requirements of the research.
- Costs around the research were reduced. Purchasing a product with sufficient functionality would have increased the cost of the research.
- The instrument could be adjusted to cater for new ideas, and changed to filter out spurious information. It allowed for a more exploratory approach to the research.

The cons to this approach included the following:

- A lot of effort was spent in designing and developing the instrument, perhaps detracting from a focus on the other requirements of the research.
- The research instrument may have to be re-designed and re-developed for even the slightest variation in the research, should another researcher build on this research.

3.5 Procedure for data collection

Data were provided by Sharenet, a commercial provider of share and derivative information on the JSE. The data were downloaded using the tools provided by Sharenet, and collated into a database using software written by the researcher.

The data were checked for consistency and accuracy, and inaccuracies and irrelevant data were then removed from the database. Details around the removal and cleansing of the data are further detailed in Chapter 4.

3.6 Data analysis and interpretation

The various trading scenarios were ranked according to their success. Success is determined by applying the Sharpe Ratio to the returns of the strategy.

A surface analysis chart was used to show the returns of the various strategies, and how these were affected by the cointegration t-statistic, and the opening standard deviation.

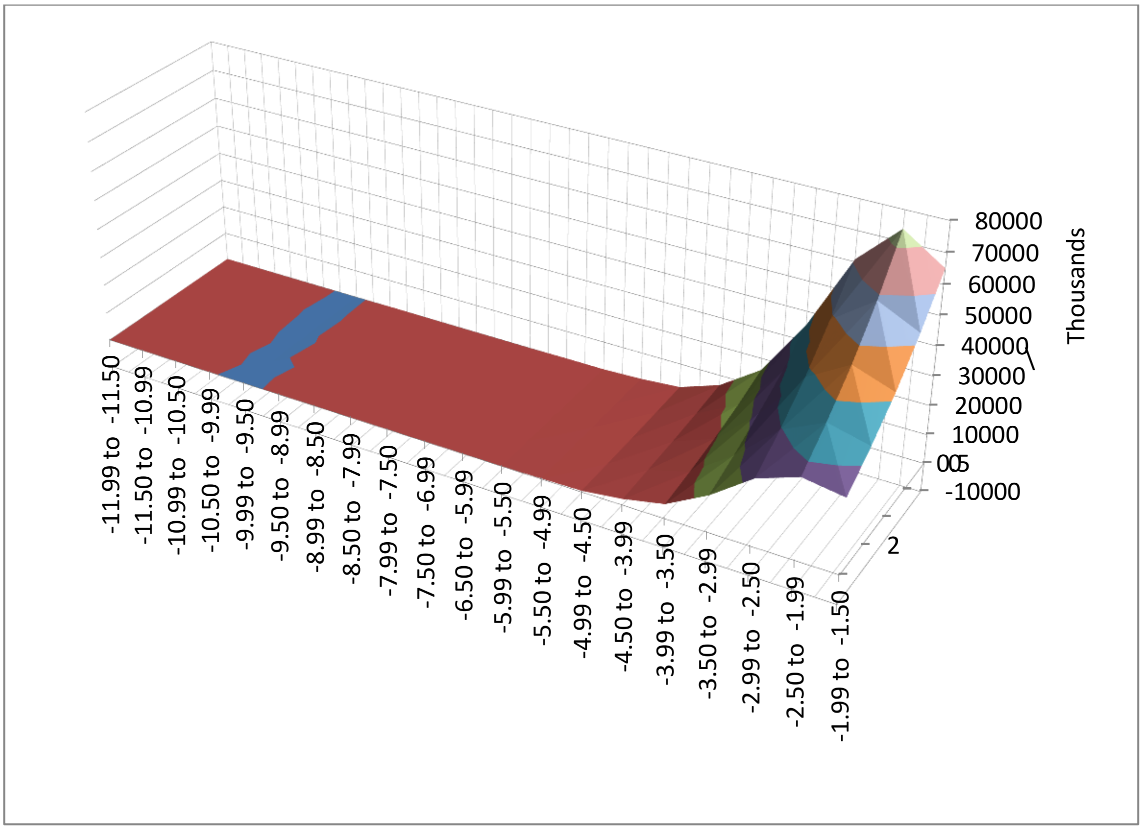

Figure 2 : Closing and Opening Standard Deviation Surface Chart

A line graph was used to show profits/losses from the various portfolios compared to each other.

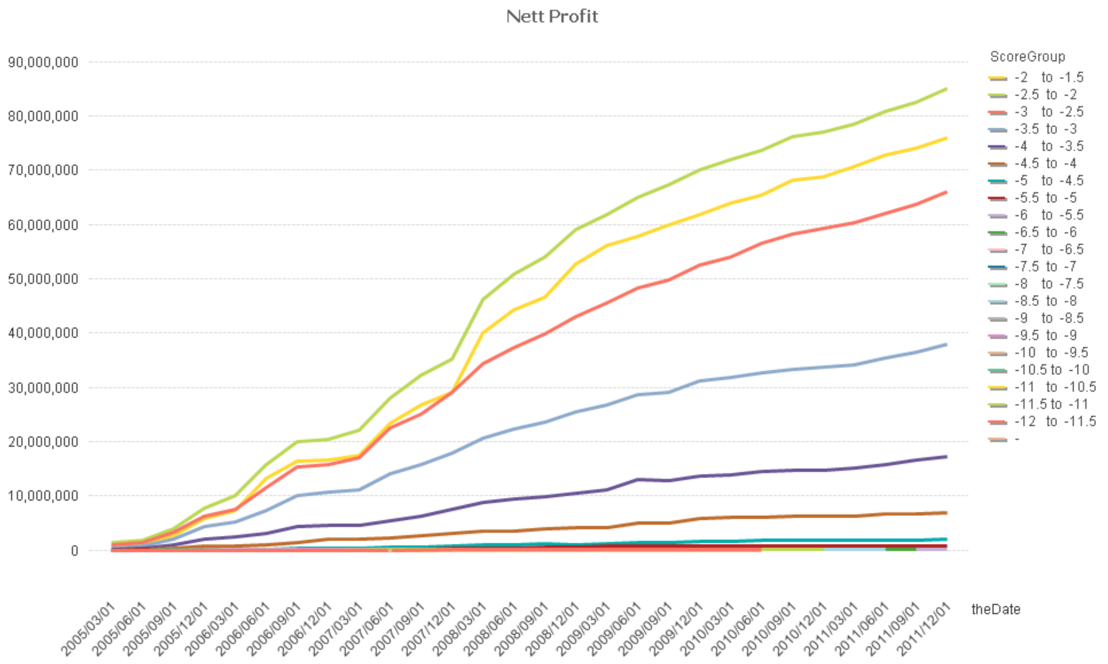

Figure 3 : Cumulative Return Percentage Line Graph

A scatter chart was used to compare the Sharpe Ratios of the various portfolios generated by the parameters to the ALSI.

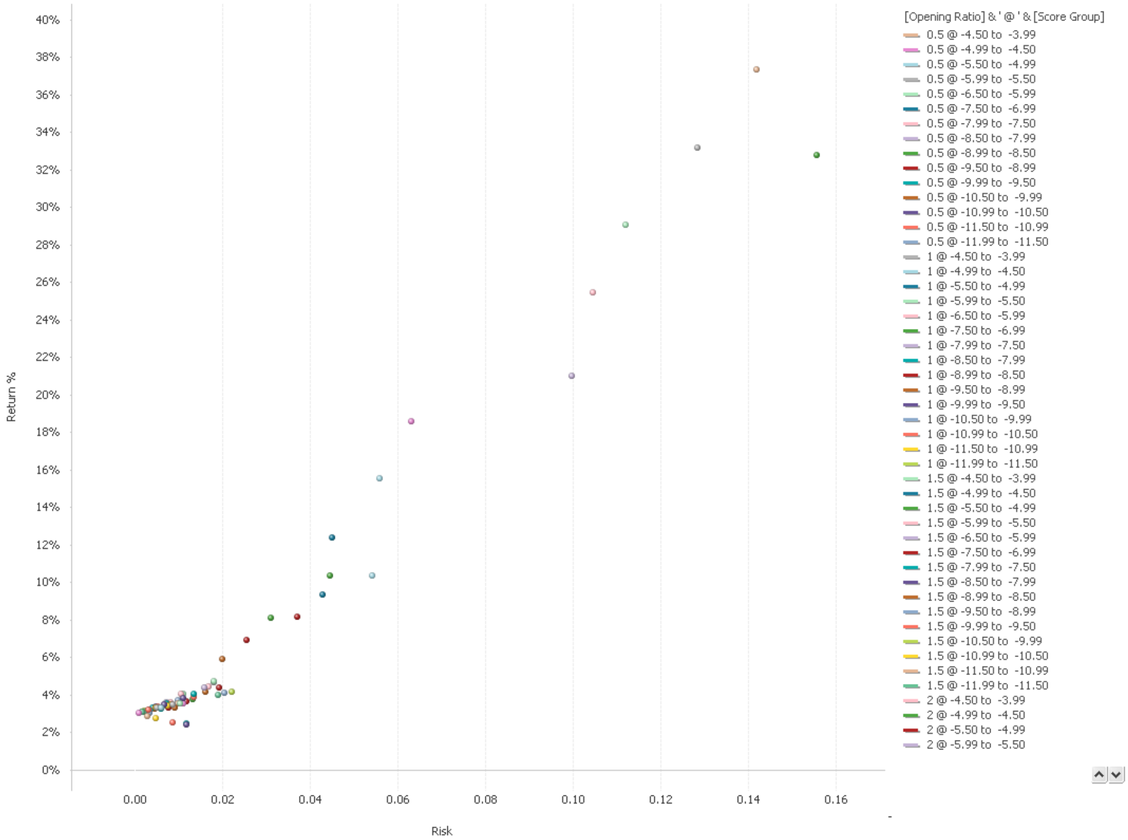

Figure 4 - Chart of Sharpe Ratios

The Sharpe Ratio was used to determine success of the various hypothetical portfolios in the theoretical simulated construct.

3.7 Limitations of the study

1) Costs were an estimate. True costs could have resulted in the study having a different outcome.
2) The closing price was used for all trades. In reality, the traded price may differ slightly from the closing price.
3) The research assumed availability of willing buyers and sellers. In reality, there may not always be a willing buyer and seller.
4) The methodology assumed that the key independent variables were successfully identified from the literature. There may be other variables

which were not identified which may have an effect on the outcome of the trading strategies.

3.8 Validity and reliability

3.8.1 ***External validity***

External validity as defined by (Leedy & Ormonde, 2001) is the extent to which the results of the research apply to situations beyond the study itself. The study attempts to replicate all costs encountered in the real world of trading futures on the JSE with as much detail as is feasible. This leads credence to the assertion that this research is therefore externally valid.

3.8.2 ***Internal validity***

Internal validity is defined as whether the variables which have been selected for the research are valid for the creation of a theoretical model. (Groenewald, 1989) asks the question "Are the chosen variables a valid representation of the theoretical universe of discourse?" This question may be extended to determine whether the variables are exhaustive, mutually exclusive and meaningful. (Leedy & Ormonde, 2001) define internal validity by the extent to which the design of the research, and the data that it yields, allow the researcher to draw accurate conclusions about the relationships within the data.

The variables have been determined by an analysis of the literature. The most common and important variables have been identified, and are therefore deemed to be valid.

3.8.3 ***Reliability***

Reliability may refer to the extent to which the study may be replicated between different times using the same instrument. As the study is based on an exact algorithm, the study will reproduce the same results reliably each time. If the same algorithm were run on another financial derivative instrument or even on another data set of single stock future information from another time period or stock market, there is no guarantee that the study would return the same results. This is the nature of time-series analysis.

CHAPTER 4: PRESENTATION OF RESULTS

4.1 Introduction

A total of 8 954 single stock futures were paired up with each other to create a total of over one million (1 150 266) possible share combinations. All of these were analysed to calculate the t-statistic for cointegration between each pair. The results were then analysed to calculate the critical values for the t-scores.

All shares with a cointegration t-statistic of less than -1.5 were then traded in a simulated environment across the five different values for the controllable independent variable (the opening point for the trade). This resulted in the simulation of a total of 3.5 million hypothetical trading scenarios.

These hypothetical trading scenarios resulted in an average of over 10 trades each, resulting in over 35 million separate simulated transactions.

This chapter presents the results of each of the steps described above, as well as the resulting net profit or loss of the various trading strategies which were calculated from the simulation.

The results have been presented as a series of charts visually demonstrating the outcomes of the research. Each of these charts is presented and clarified in this chapter.

4.2 Data Profile of Sample Population

Initially the full set of all stocks from the inception of Single Stock Futures on the SAFEX through to the present was going to be used for the study. As the study progressed, certain subsets of data were removed from the simulation, and certain results of the simulation had to be excluded from the final analysis. Chapter 5 discusses the reasoning behind why the subsets were eliminated. This section lists the subsets of data that were eliminated without discussing the reasoning behind the elimination.

All single stock futures from before 2004 were eliminated due to a general lack of trading volume. The data sample was further reduced due to questionable data received from the data service provider. Particularly, two subsets were removed from the study: All shares where the share price fluctuated by more than 50% across two days were removed. And all share prices that were recorded after the 15th of the month, in the month that the single stock future theoretically closed out were removed.

Finally, the results dataset was reduced due to concerns around trading volume. All share trade simulations where the simulated trading volume exceeded the maximum actual volume traded were removed from the results of the simulation.

4.3 Results Pertaining to the Simulation as a Whole

Due to the fact that there is one research proposition, the results are all presented in this section of the chapter.

4.3.1 *Cointegration Bootstrap Results*

This section contains the presentation of results relating to the cointegration bootstrap. This data were obtained by running a cointegration analysis across more than one million pairs of single stock futures, and calculating the critical score at various percentiles of the data set.

Percentile	Mean value of *n*	Critical Score
98%	80	-5.1118
95%	80	-4.5556
90%	80	-3.9935

Figure 5 Table of Critical Values for ADF t-statistic

The mean number of matching rows for any two possible single stock futures combinations was 80. Across the 1.15 million possible combinations of SSF's, the above bootstrapped values for the cointegration t-statistic were calculated. To state the null hypothesis for the cointegration test, the lower the score, the less likely that the pairs were not cointegrated.

4.3.2 *Opening Standard Deviation Results*

This section contains a presentation of the results pertaining to the opening standard deviation, and the effect it had on the outcomes of the various simulation scenarios.

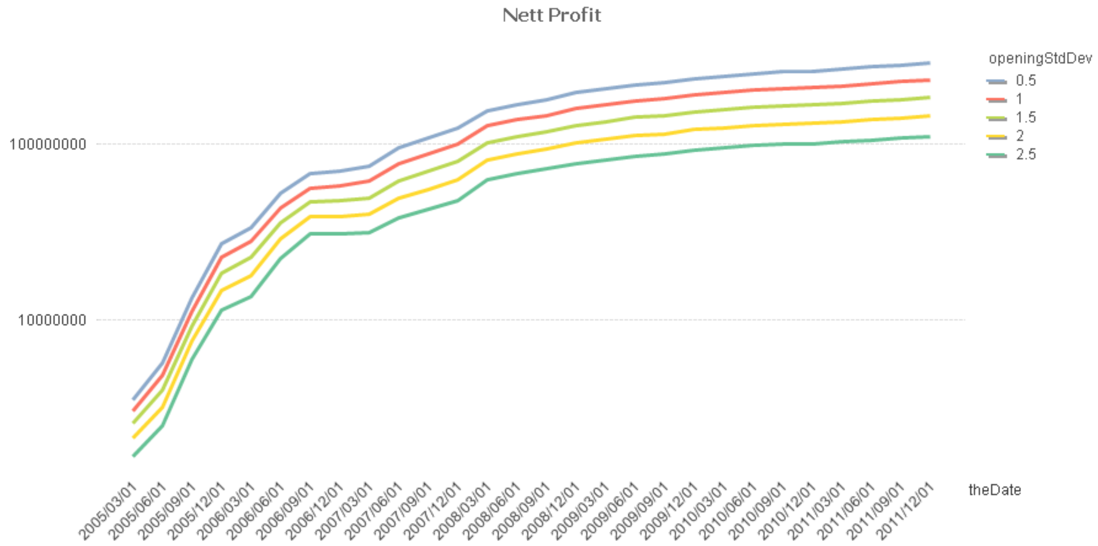

Figure 6 – Returns by Opening Standard Deviation

The figure above represents the accumulated returns of the various trading strategies. In the above figure, an opening Standard Deviation Beta of 0.5 results in a better overall result than any of the other opening standard deviation values.

Figure 7 -- Average Net Profit by Opening Standard Deviation

The above figure represents the average net profit per trade per opening standard deviation. Although an opening standard deviation of 0.5 returned a more profitable total result (See previous figure), an opening standard deviation of 1.0 returned a higher average profit per trade.

4.3.3 ***Simulation Results by cointegration t-score.***

Based on the fact that an opening standard deviation of 0.5 resulted in the highest profits, this opening standard deviation was selected as the starting-point for the next section of the analysis. Showing all co-integration scores by all standard deviations led to a number of data points (100) which are simply not readable or understandable on a line graph showing a time series. The analysis of all 100 data points is included in the section following this one, where the closing standard deviation AND the cointegration t-test are analysed in terms of both volatility and profitability.

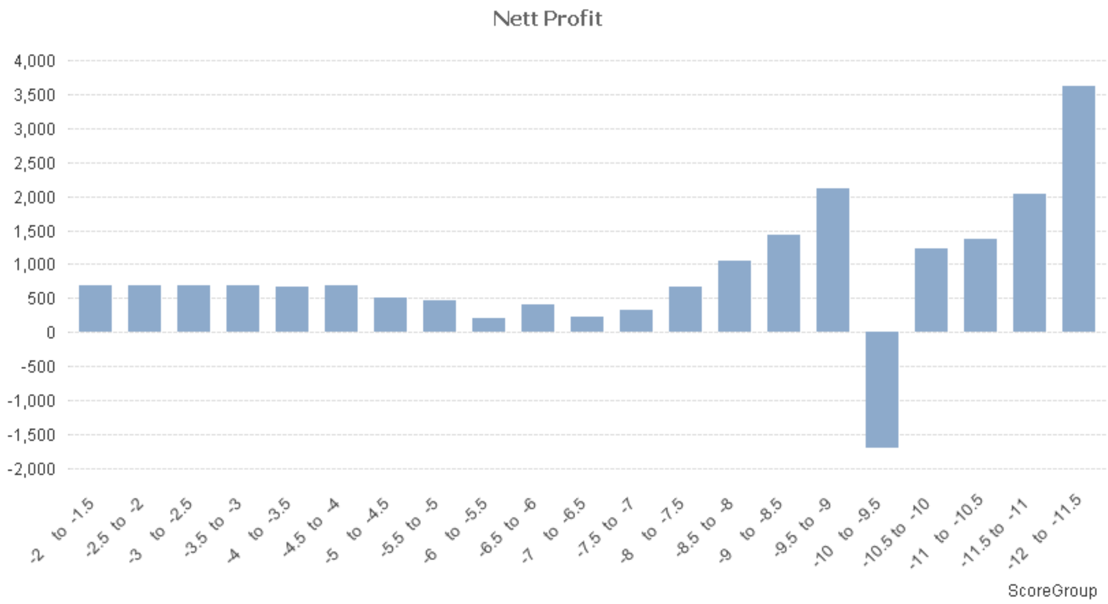

Figure 8 – Average Net Profit by Cointegration T-statistic

The above figure shows the average net profit by cointegration t-statistic. The figure clearly shows that there is a trend towards profitability for pairs that are more cointegrated.

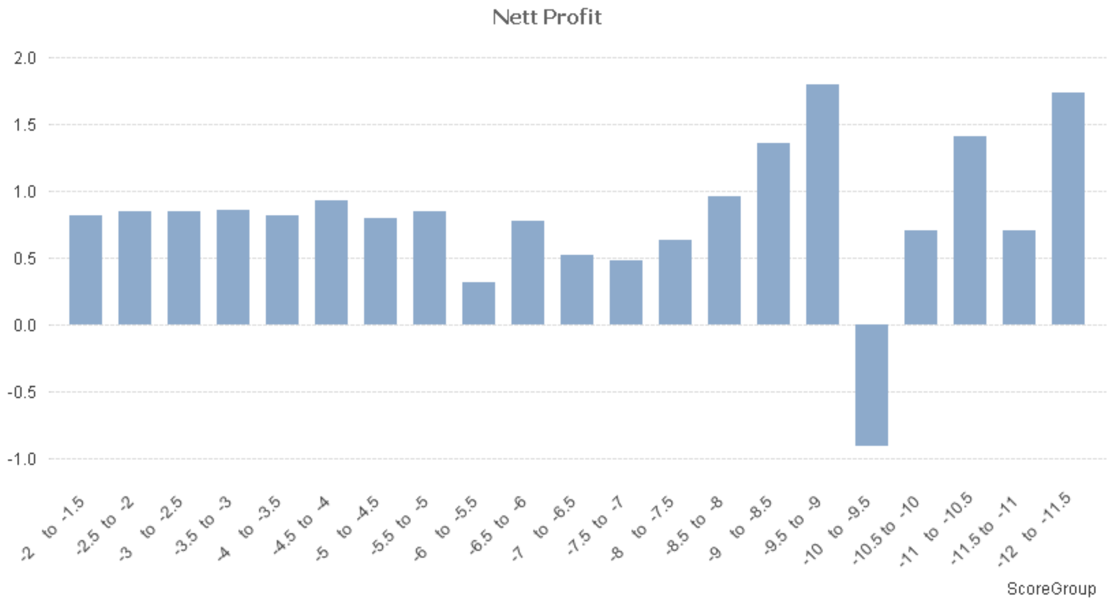

Figure 9 – Net Profit / Average Margin by Cointegration T-Statistic Group

The above figure shows the net profit divided by the average margin required to enter into each pairs trade by cointegration t-statistic grouping.

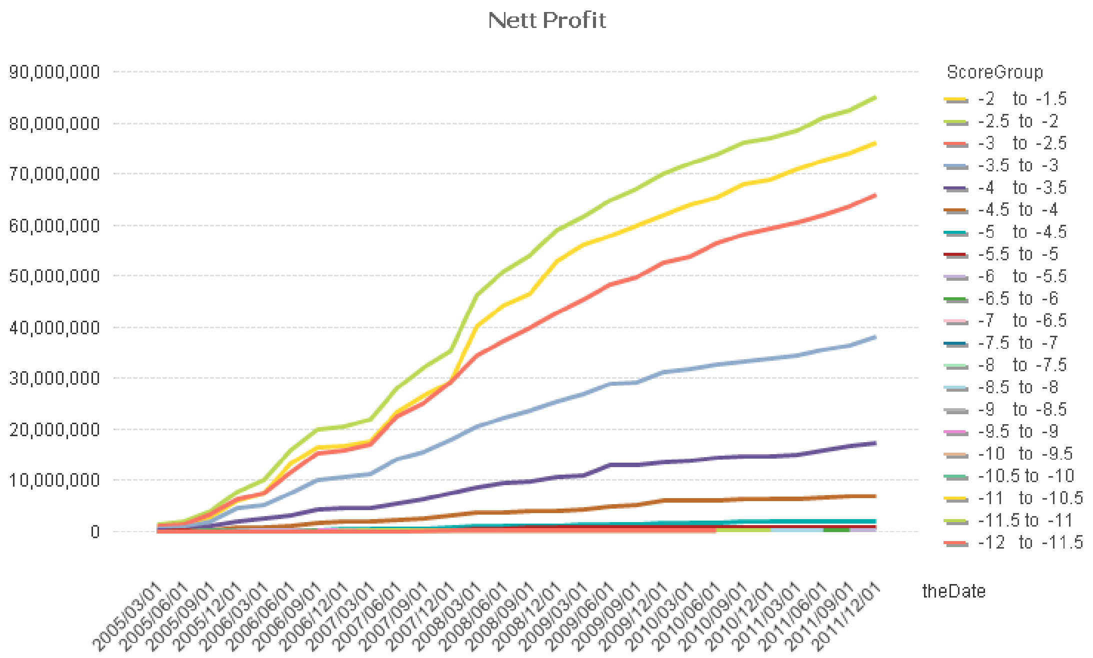

Figure 10 – Net Profit by Cointegration T-Statistic Group

The above figure show the net profit by cointegration t-statistic grouping over time. It indicates that the most profits over the lifetime of the simulation were made by the share pairs that were the least cointegrated.

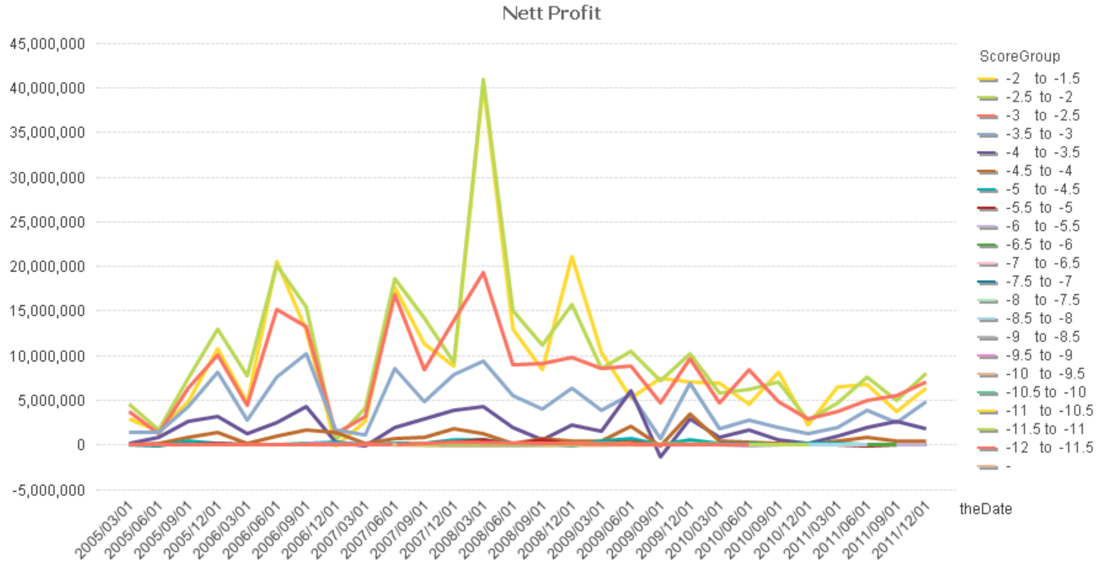

Figure 11 – Net Profit by Cointegration T-Statistic Group

The above figure shows the various net profits grouped by t-statistic, without any sort of accumulation. This chart shows that all groups tend to exhibit a fair amount of volatility. This volatility is quantified in the next section.

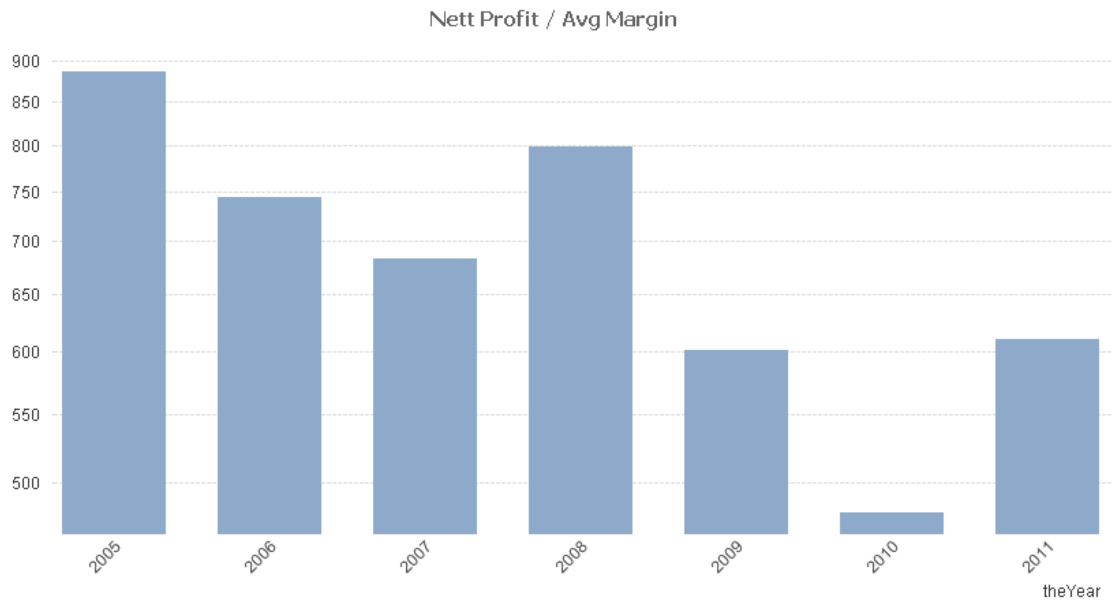

Figure 12 – Average Net Profit by Year

The above figure shows average net profit by year. The trend shows a reduction in average profits over time, as more recent years tended to produce a lower average profit than in the past. This chart is for all pairs with a cointegration score of less than -1.5.

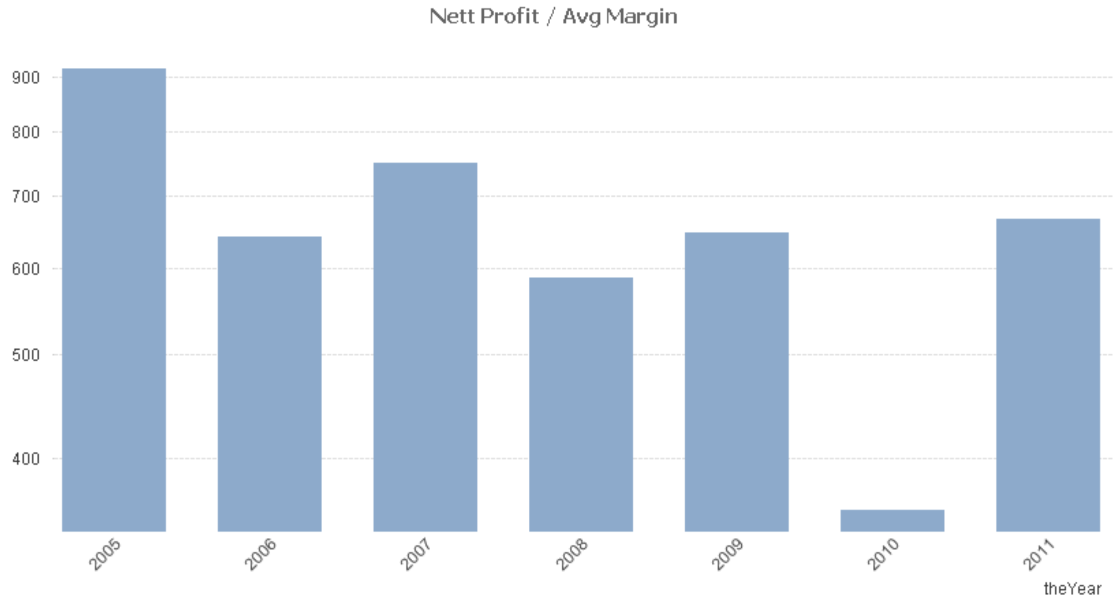

Table 3 - Net Profit / Avg. Margin by Year (Cointegrated Only)

The table above shows only pairs that were cointegrated, and it illustrates that the returns, when viewed as a ratio against the average margin required for the trade. This ratio varied considerably year-by-year. The highest return ratio was 1.3, and the lowest was 0.6.

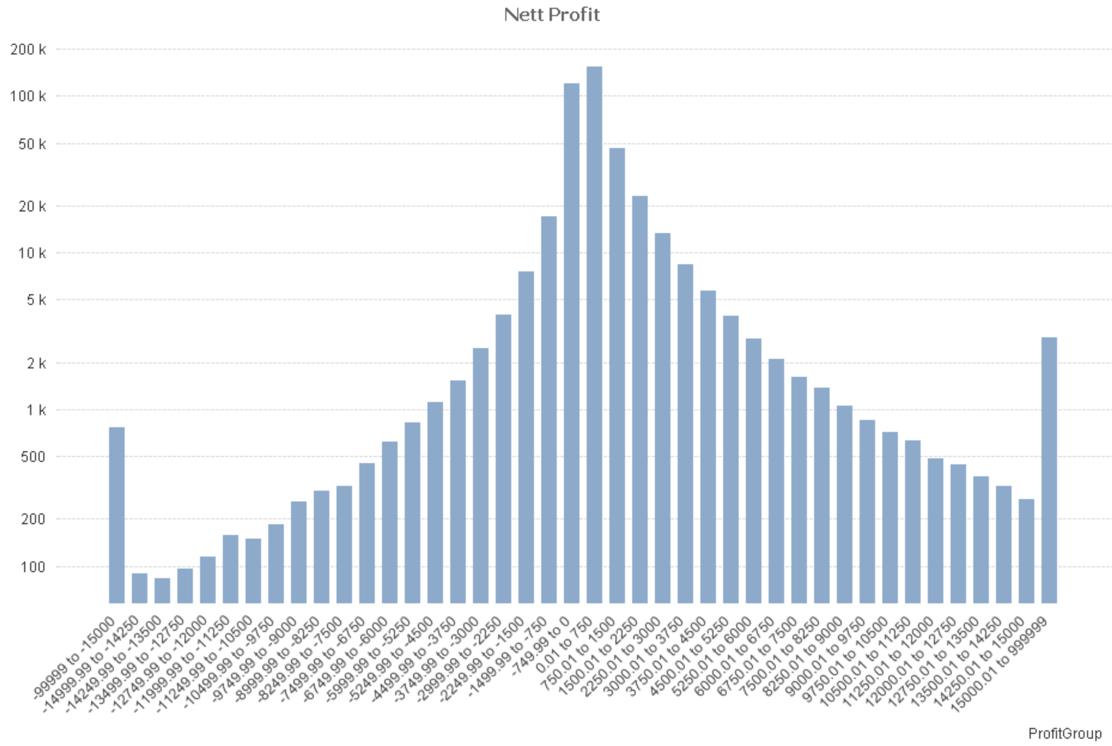

Figure 13 – Distribution of Returns

The above figure shows the distribution of returns across all time for all cointegration groups.

Statistic	Total	Cointegrated	Non-cointegrated
Average	R654.83	R641.09	R689.53
Skewness	5.25	4.96	5.44
Kurtosis	105.91	128.09	103.61

Table 4 - Statistical Summary of Returns

The statistics above relate to the net profit distribution from all trades where the opening standard deviation was 0.5. The distribution of returns shows a positive skewness with high kurtosis. A positive skewness indicates that on average the returns were positive. The high kurtosis is an indication that the returns tended to vary extensively.

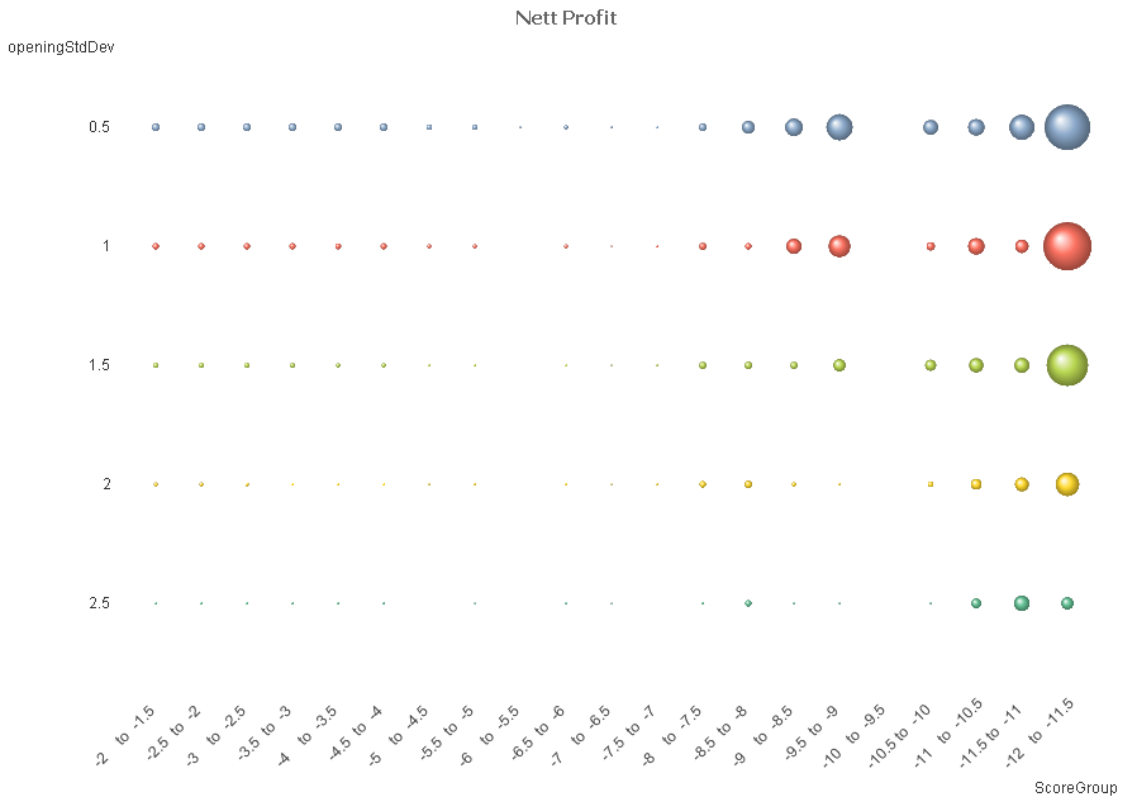

Figure 14 – Average Net Profit by Opening Standard Deviation and Cointegration T-Statistic (Summary)*

The above figure shows the average net profit. The size of the bubble represents the size of the average net profit. It illustrates that there is a strong relationship between the degree of cointegration, and the average net profit. An opening standard deviation of between 0.5 and 1.5 with a low cointegration t-score of -6.5 or lower generally produced the most positive average returns.

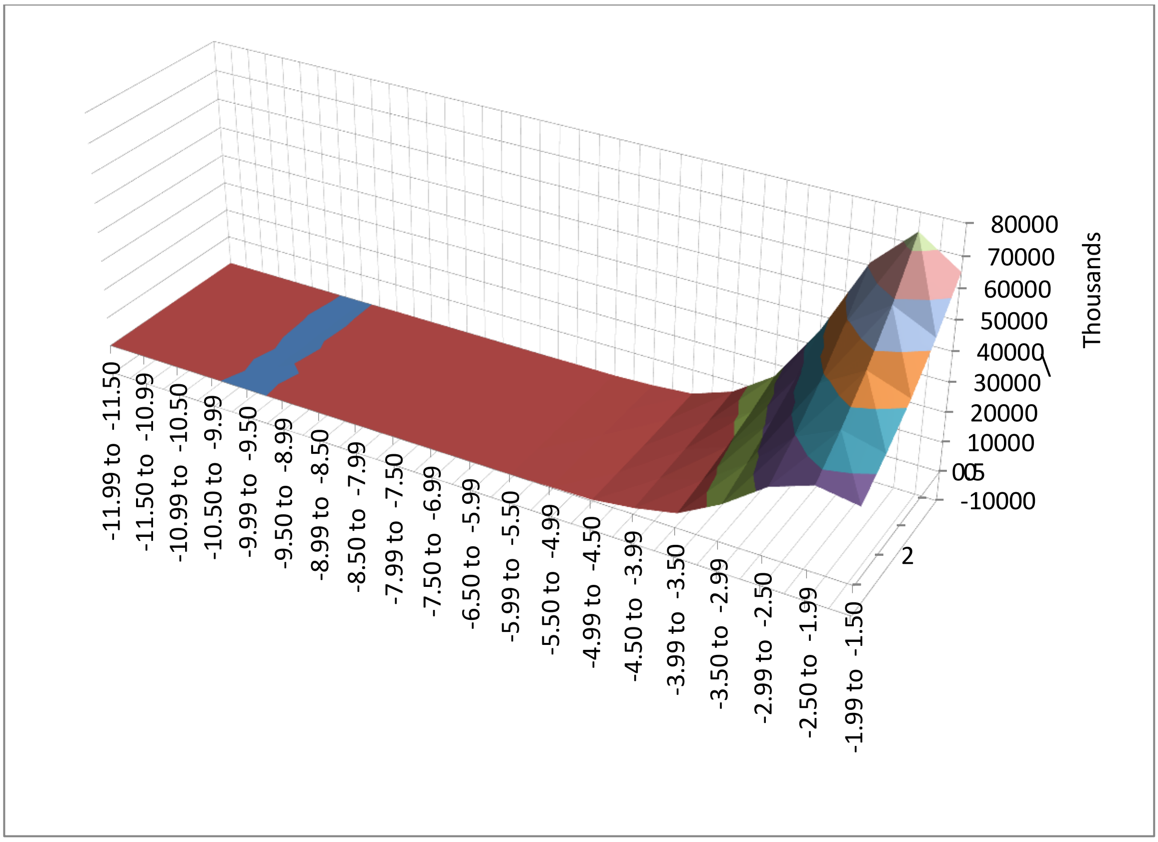

Figure 15 - Total Profit by Cointegration Score by Opening Standard Deviation

The above figure shows total profit by cointegration score and opening standard deviation. As illustrated in the previous chart, the lower the cointegration score, the higher the average profit. However, this chart shows that the opportunity for profit-taking is much lower in the lower cointegration t-test score ranges.

4.3.4 *Sharpe Ratio*

The previous two sections have dealt with the returns from the various possible portfolios which could be modelled from the simulation. This last section explores how both risk and returns interact. An analysis of fund performance is not complete without bringing risk into account (Agarwal & Narayan, 2004). This factor is important when dealing with a hedge fund using derivatives, as it is possible to lose more money than one initially invested (Standard Bank, 2010).

This research uses the Sharpe Ratio to quantify the interplay between risk and return. Stable and high-performing funds have the best Sharpe Ratios (Sharpe, 1994).

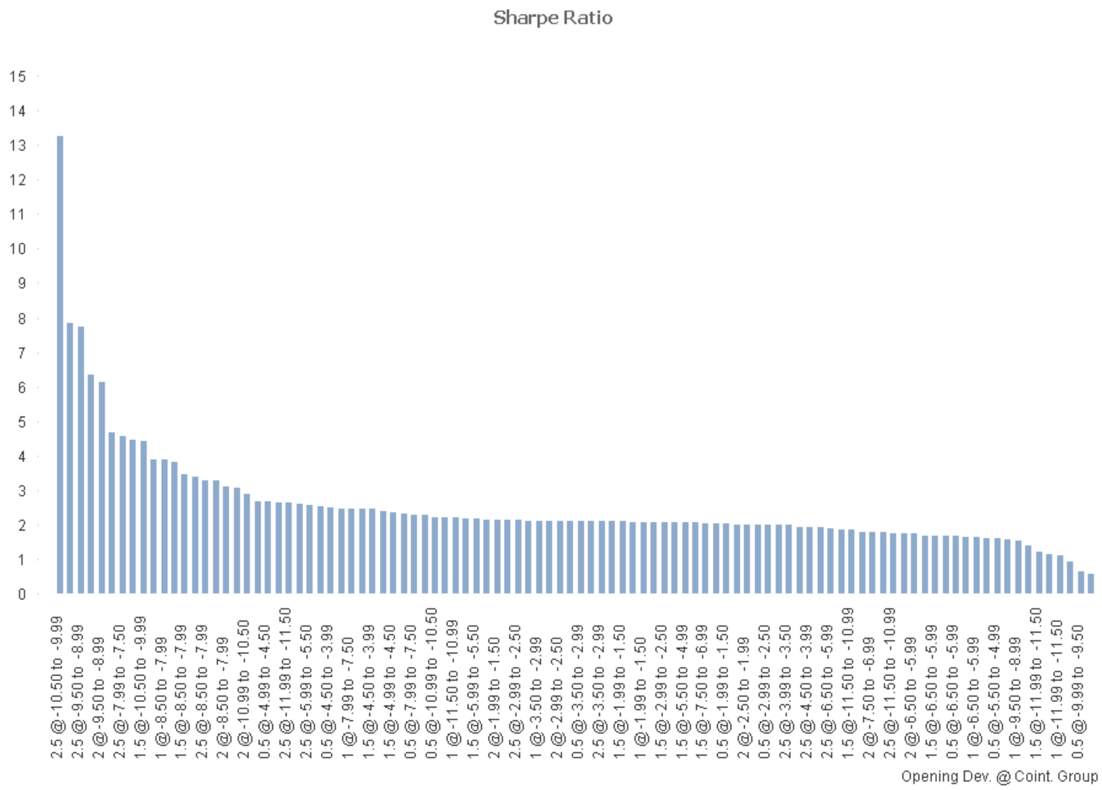

Figure 16 -- Ranked Sharpe Ratios for Top 30 Combinations

This chart displays the top 30 Sharpe Ratios for the various potential portfolios. As previously seen, portfolios having the most-cointegrated pairs had the highest Sharpe Ratios. An opening standard deviation of 2.5, on average, produced the most favourable Sharpe Ratios.

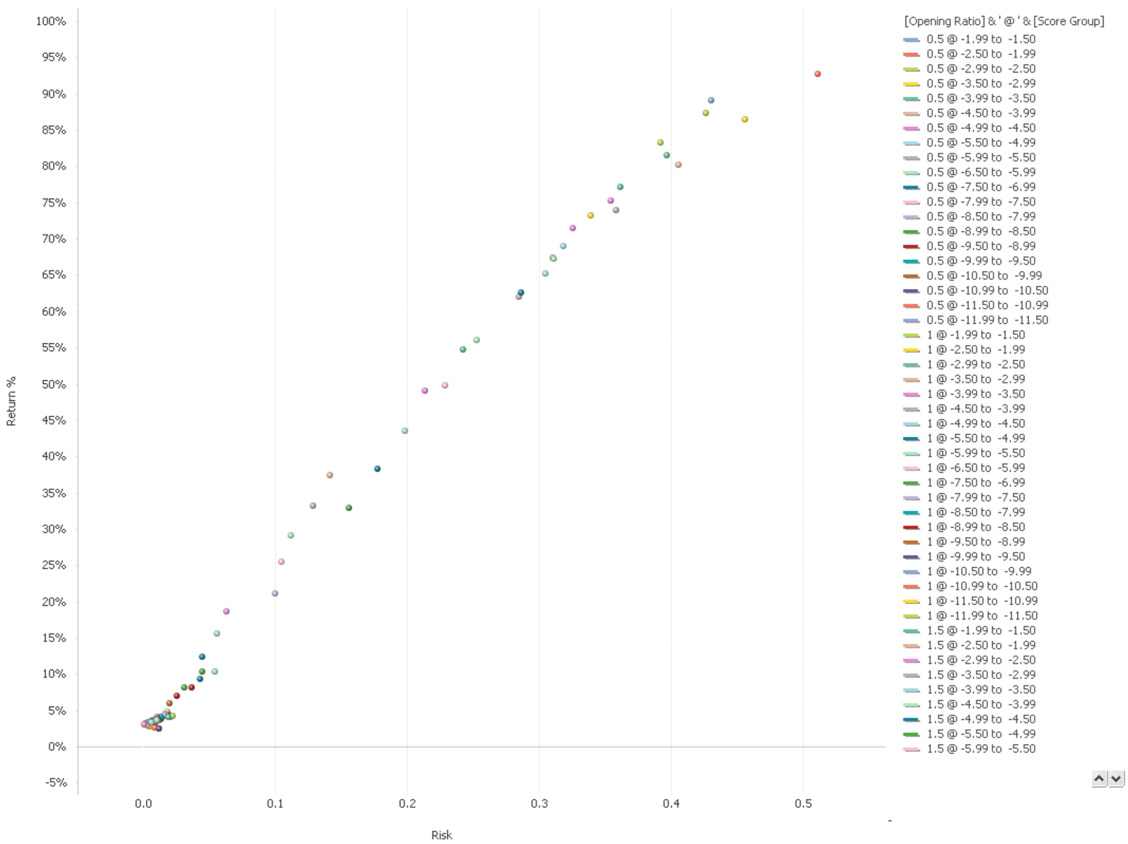

Figure 17 -- Scatter Plot of Top 25 Returns by Sharpe Ratio

The above figure plots the various risk-and-return profiles generated by the various portfolios. There were exactly 100 possible theoretical portfolios in the study. Examination of a graphical representation of the data is difficult to understand because of the strong linear relationship exhibited across the data. The table below is easier to interpret.

Opening Standard	Cointegration Score Group	Annualised Return	Risk (Standard	Sharpe Ratio
2.5	-10.50 to -9.99	2.997%	0.00100	13.25
2.5	-8.99 to -8.50	3.056%	0.00177	7.83
2.5	-9.50 to -8.99	3.079%	0.00183	7.73
2	-10.50 to -9.99	3.069%	0.00221	6.34
2	-9.50 to -8.99	3.112%	0.00236	6.14
1	-10.50 to -9.99	3.116%	0.00310	4.67
2.5	-7.99 to -7.50	3.164%	0.00329	4.55
2	-8.99 to -8.50	3.189%	0.00343	4.44
1.5	-10.50 to -9.99	3.164%	0.00338	4.43

Opening Standard	Cointegration Score Group	Annualised Return	Risk (Standard	Sharpe Ratio
2.5	-9.99 to -9.50	2.814%	0.00297	3.87
1	-8.50 to -7.99	3.296%	0.00421	3.87
2.5	-7.50 to -6.99	3.007%	0.00350	3.83
1.5	-8.50 to -7.99	3.332%	0.00481	3.46
1.5	-8.99 to -8.50	3.293%	0.00481	3.38
2.5	-8.50 to -7.99	3.297%	0.00498	3.28
0.5	-10.50 to -9.99	3.232%	0.00479	3.27
2	-8.50 to -7.99	3.337%	0.00537	3.11
2.5	-10.99 to -10.50	3.273%	0.00526	3.06
2	-10.99 to -10.50	3.300%	0.00566	2.89
2	-7.99 to -7.50	3.470%	0.00674	2.67
0.5	-4.99 to -4.50	18.541%	0.06323	2.67
0.5	-8.50 to -7.99	3.579%	0.00726	2.64
2.5	-11.99 to -11.50	3.273%	0.00612	2.62
2	-5.99 to -5.50	3.448%	0.00688	2.59
2.5	-5.99 to -5.50	3.249%	0.00619	2.56
1.5	-7.99 to -7.50	3.496%	0.00724	2.53
0.5	-4.50 to -3.99	37.338%	0.14198	2.51
1	-4.99 to -4.50	15.491%	0.05607	2.47
1	-7.99 to -7.50	3.499%	0.00747	2.45
1	-4.50 to -3.99	33.126%	0.12858	2.45
1.5	-4.50 to -3.99	29.051%	0.11213	2.44
0.5	-7.50 to -6.99	3.422%	0.00736	2.39
1.5	-4.99 to -4.50	12.330%	0.04522	2.36
1.5	-10.99 to -10.50	3.403%	0.00747	2.32
0.5	-7.99 to -7.50	3.556%	0.00828	2.28
2	-4.50 to -3.99	25.404%	0.10458	2.27
0.5	-10.99 to -10.50	3.473%	0.00817	2.21
1	-3.99 to -3.50	48.969%	0.21384	2.21

Table 5 - Sharpe Ratios by Opening Standard Deviation and Cointegration Score Group

The above table shows the Sharpe Ratio for the top 50 theoretical portfolios ranked by Sharpe Ratio in descending numeric order. It shows the two factors that are used to calculate the Sharpe Ratio, both Risk and Return. The full table is included in the Appendices.

Opening Standard Deviation	**Grouping**	**Annualised Return**	**Risk (Standard Deviation)**	**Sharpe Ratio**
0.5	Cointegrated	67.068%	0.31	2.1
0.5	Not Cointegrated	132.37%	1.15	1.1

Table 6 – Consolidation of Sharpe Ratios by Pair Type

The table above presents a scenario not proposed in the original research. It shows the results of combining the various portfolios presented in the previous tables and charts into only two possible combinations. Based on the fact that the opening standard deviation of 0.5 produced the highest returns, this was selected as the opening standard deviation for these two new proposed portfolios.

The cointegration score groups were then consolidated into two groups, “Cointegrated” and “Not Cointegrated”. The “Cointegrated” group consisted of all pair trades with a cointegration score of -3.9 and below. The “Not Cointegrated” group consisted of all other pairs tested in the simulation. Due to the relatively large number of trades performed in the “Not Cointegrated” group, the absolute profit was much higher than for the “Cointegrated” group, but this was at the cost of higher volatility. When both risk and return are taken into account, the “Cointegrated” portfolio produced a much more favourable result.

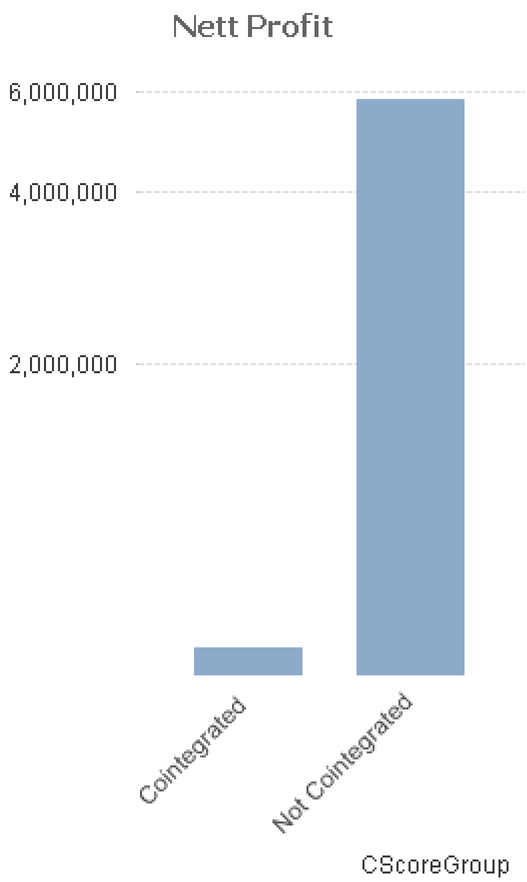

Figure 18 - Number of Trades by Pair Type

The above table illustrates the difference in number of trades executed over the lifetime of the simulation for the two portfolio groups. The "Cointegrated" group required only 10% of the number of trades compared to the "Not Cointegrated" group.

4.3.5 *Conclusion*

The research proposition is that there exists a set of key independent variables which can be controlled to produce the most profitable, least volatile returns using pairs trading as a strategy. These key independent variables are the squared distance at which the trade is opened (the opening standard deviation) and the t-test cointegration score.

Based on the presentation of results in this chapter, the research proposition was proven to be true. The results of the simulation show that using different combinations of opening standard deviations and cointegration scores for the 100 theoretical portfolios resulted in a variety of returns with varying risk profiles.

The most profitable and stable portfolios were those using pairs with an opening standard deviation of 0.5 and a cointegration t-test value of less than -3.9. When compared to previous studies done on pairs trading, this presentation of results suggests that previous researchers may have produced better results by using a t-test critical value of at least -3.9, instead of the standard value of -2.5 proposed by McKinnon (MacKinnon, Haug, & Michelis, 1995).

Researchers	**Signal to Open**
(Do, Faff, & Hamzal, 2006)	Two Standard Deviations
(Gatev, Goetzman, & Rouwenhorst, 2006)	Two Standard Deviations
(Herlemont, 2004)	Two Standard Deviations
(Nath, 2003)	Variable percentile of distance
(Vidyamurthy, 2004)	0.7 standard Deviations

Table 7 - Opening Standard Deviation Used in Previous Research

The opening standard deviation used by previous researchers also seems to be higher than the optimal value presented by the simulation used in this research. As illustrated in the table below, most previous studies used an opening standard deviation of two, whereas the simulation shows that 0.5 standard deviations are more profitable and more stable.

4.4 Summary of the results

In this chapter, data from the simulation was presented. The four key areas of data presentation were the cointegration t-test critical values, the opening standard deviation and its effect on profitability, and the cointegration t-testing of pairs and its effect on profitability. Finally, the chapter presented how all of these factors and variables could be represented using an analysis of each portfolio using the Sharpe Ratio.

The results of running the Engle-Granger two-step test against all possible pairs showed that the critical score was -3.9 at the 90th percentile. The data set used for this calculation had an average of 80 records per pair tested.

The opening standard deviation affected profitability and risk. Five different values for the opening standard deviation value were tested against 20 portfolios differentiated by cointegration score. This gave a total of 100 different theoretical portfolios. An opening standard deviation of 0.5 produced the most profits. An opening standard deviation of 1.0 resulted in the set of portfolios that had the lowest risk.

The cointegration score was also shown to affect profitability and risk. Lower cointegration scores tended to produce higher average profits and were less volatile than higher cointegration scores. However, relative to the total number of potential trades, there were not many trades available at the lowest end of the cointegration score spectrum (-7 to -12). Therefore, a balance needed to be found between risk and profitability.

The Sharpe Ratio analysis was used to find this balance. By analysing returns using the Sharpe Ratio, it was possible to identify that a portfolio that had a cointegration score of less than -3.9 and an opening standard deviation of 0.5 produced the best profits relative to risk.

CHAPTER 5: DISCUSSION OF THE RESULTS

5.1 Introduction

This chapter discusses the results of the simulation. It looks into the meaning behind the various results that were recorded from the simulation, and how these relate to the research proposition. The research proposition is that there exists a set of key independent variables which can be controlled to produce the most profitable, least volatile returns using pairs trading as a trading strategy. These key independent variables are:

1) The squared distance at which the trade is opened. (The opening standard deviation)
2) The ADF t-test cointegration score

5.2 Profile of the Data

Although initially the full set of all stocks from the inception of Single Stock Futures on the SAFEX through to the present was going to be used for the study, one the data had been obtained, it was clear that the number of shares available for trade up until 2004 was simply too small to be considered beneficial to the study. All share combinations traded before 2004 were therefore eliminated from the simulation.

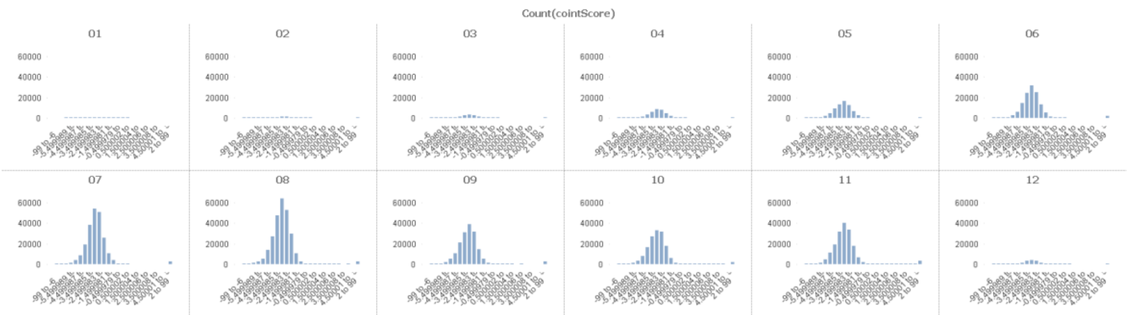

Figure 19 – Distribution of Pairs by Cointegration Score by Year

The figure above shows the total number of pairs available to the study, and their distributions across the cointegration t-score spectrum by year.

The data sample was further reduced due to questionable data received from the data service provider. For example, there were a number of shares where the share price fluctuated by more than 50% in the space of a single day. To avoid incorrect data contaminating the simulation, all share prices that fluctuated by more than 50% between two days, and had a share price of more than 100 cents were eliminated from the dataset. Random observations of the data showed that shares with a price of below 100 cents tend to fluctuate by more than 50% day-to-day, and therefore this fluctuation was considered expected in real-life trading conditions, and these shares were not eliminated.

The data also contained a large number of rows for single stock futures past their known close-out date. According to the rules of the SAFEX, single stock futures are closed out on the last Thursday of each quarter. In order to eliminate any fictitious trading after the SSF close-out, all share price movements after the 15th day of the last month of each quarter were eliminated from the dataset.

Finally, in order to ensure that the simulation was as closely modelled on the real world as possible, all share trade simulations where the simulated trading volume exceeded the maximum actual volume traded were removed from the results of the simulation. For example, if the historical data showed that the maximum volume traded for a particular single stock future was 1500, then if the simulation at any point called for a trade of larger than 1500, then the pair was eliminated from the sample population.

5.3 **The relationship between the cointegration score, the opening trade variable and trading success**

As shown in Chapter 4, a critical value of -3.9 was calculated as being the most appropriate to determine cointegration. Based on the research proposition, the questions then become, “Does it make any difference if a share pair is cointegrated when using pairs trading?”, and “Does the opening standard deviation affect the success of the trading strategy?" Sub-questions to this include: "Does the extent to which a pair of shares show cointegration impact on their profitability, or does it make no difference to profitability, as long as the shares are indeed cointegrated?" These questions are explored and discussed in this section of the chapter.

5.3.1 ***The Opening Standard Deviation***

Based on the results from the various charts, the greatest profit on an absolute basis was from the share trades where the trade was opened at 0.5 standard deviations. Over time, the profitability of a 0.5 standard deviation made a substantial difference, with a net overall profit of approximately 50% more than a standard deviation of 2.5. It is beyond the scope of this research to understand why this caused such a difference.

Although the charts show the net profits by opening standard deviation, they do not reveal much about the resulting volatility of returns. The volatility of the results will be discussed in the Sharpe Ratio section further down.

Based purely on an analysis of the profits, it would appear that previous researchers may have achieved better results if they had used a lower opening standard deviation. (Do, Faff, & Hamzal, 2006), (Gatev, Goetzman, & Rouwenhorst, 2006), and (Herlemont, 2004) all used an opening standard deviation of 2.0. This study seems to indicate that a lower opening standard deviation of either 0.5 or 1.0 could have resulted in larger profits if they had used it instead. It is interesting to note that (Vidyamurthy, 2004) used an opening standard deviation of 0.7, which is in the range where the simulation produced the greatest profits

.

5.3.2 *The Cointegration Score*

Table 8- Critical Values for ADF t-statistic

Percentile	Mean value of *n*	Critical Score
98%	80	-5.11
95%	80	-4.56
90%	80	-3.99

Compared to (MacKinnon, Haug, & Michelis, 1995), the boot-strapped cointegration scores produced by the simulations run in this research are lower. This suggests that perhaps for time series data of this nature, the critical values determined by this research may be more appropriate for future researchers.

5.3.3 *Combining the Opening Standard Deviation and the Cointegration t-statistic*

It is clear from the charts presented in the previous chapter that trades where the shares were the most co-integrated (having the lowest critical t-statistic from the Engle-Granger two-step test) produced the highest profits. This fits the theory of cointegration, that over time, two sets of time series data will revert relative to each other if they are cointegrated (Alexander, 2001). The very theory of pairs trading is based on this assumption, and the simulations run for this research back this theory up.

Although the average net profit is substantially higher for pairs displaying the most cointegration, they tend to appear relatively less often in practice. This means that although pairs trading with highly-cointegrated pairs makes sense in practice, if the trader were to wait for the perfect conditions with the perfect pairs to occur, there would not be much trading going on at all.

For this reason, it is important to examine the strategy in light of an absolute return amount. Although the strategy employed in the simulation was for all intents and purposes cash neutral, due to the nature of margin calls, the fund needed to run with a cash balance of sorts. As each opening trade is speculative, and the potential for loss is real, a cash balance needs to be maintained by the fund to cater for margin calls when required.

Assigning an opening cash balance to each fund allowed for a means to measure relative growth of each trading strategy against an absolute amount. The amount selected for the purposes of the analysis was R1 000 000.

As this is a theoretical construct, this amount was selected on the basis of going back through the simulation results, and looking for the lowest value any of the funds ever attained, and then building in a fairly large buffer. The lowest amount any fund ever attained from inception through to the final trading was -R112 354. Based on this figure, a buffer of roughly eight times this amount was selected, giving a starting fund value for each theoretical portfolio of R1 million.

Given this starting base for each portfolio, it was possible to calculate the Sharpe ratio for each. The following section discusses the results of the Sharpe Ratio analysis.

An alternative strategy is to create two portfolios, one for pairs that are cointegrated, and another for ones that are not, using the bootstrapped value obtained from the initial analysis.

If this approach is used, instead of having to analyse 20 different theoretical portfolios based on the cointegration score, we are left with two portfolios. The results then are a lot clearer and possibly more practical in the real world.

5.3.4 *Sharpe Ratio*

The previous two sections have dealt with the returns from the various possible portfolios which could be modelled from the simulation. This last section discusses how both risk and return interact across the various portfolios.

The Sharpe Ratio is used to quantify the interplay between risk and return. Stable and high-performing funds have the best Sharpe Ratios (Sharpe, 1994).

The data presented in the previous chapter indicated that the higher the degree of cointegration, the higher the average profit. In addition, the lower the opening standard deviation, the higher the average profit returned by the portfolio. An analysis of the Sharpe Ratios produced by the 100 portfolios revealed that the best portfolios were those that had a cointegration score of less than -3.9 and an opening standard deviation of 0.5.

When the portfolios were combined to produce a consolidated portfolio based on the findings represented above, a consolidated portfolio consisting of all cointegrated share pairs with an opening standard deviation of 0.5 was found to be highly favourable.

5.3.5 *Distribution of Returns*

The distribution of net profit was only analysed for an opening standard deviation of 0.5, and was grouped by Cointegration Type. The distribution of returns indicates a negatively-skewed distribution with a positive mean return. This is consistent with the finding that the proposed strategy is profitable.

Statistic	**Cointegrated**	**Non-cointegrated**	**Comment**
Average	R731.93	R684.05	Indicates a higher return per trade for cointegrated pairs.

Statistic	Cointegrated	Non-cointegrated	Comment
Skewness	10.5	10.2	Indicates a higher return per trade for cointegrated pairs.
Kurtosis	371.0	486.5	Indicates a lower volatility in returns for cointegrated pairs.

Table 9 - Summary of Statistical Analysis of Returns

5.3.6 *Other observations*

As evidenced by the chart "Average Net Profit by Year" in Chapter 4, net profit by year displays a downward trend when looking at both cointegrated and non-cointegrated share pairs. However, if the analysis is narrowed to only include cointegrated share pairs, the results are quite different. Pairs trading with shares pairs that have been identified ahead of time as cointegrated do not show a downward trend, although the distribution of returns still seems to vary considerably by year.

It is also interesting to note that the bear market of 2008 did not seem to have an effect on the profitability of the trading strategy. This is one of the benefits of a strategy that minimises the risk associated with market beta (Alexander, 2001).

5.3.7 *Conclusion*

The research proposition is that there exists a set of key independent variables which can be controlled to produce the most profitable, least volatile returns using pairs trading as a trading strategy. These key independent variables are the squared distance at which the trade is opened (the opening standard deviation) and the t-test cointegration score.

Based on the discussion of results in this chapter, the research proposition was proven to be true. The results of the simulation show that using different combinations of opening standard deviations and cointegration scores for the 100 theoretical portfolios resulted in a variety of returns with varying risk profiles.

The most profitable and stable portfolios were those using pairs with an opening standard deviation of 0.5 and a cointegration t-test value of less than -3.9. When compared to previous studies done on pairs trading, this presentation of results suggests that previous researchers may have produced better results by using a t-test critical value of at least -3.9, instead of the standard value of -2.5 proposed by McKinnon (MacKinnon, Haug, & Michelis, 1995).

5.4 Conclusion

5.4.1 *Summary*

In this chapter, data from the simulation was discussed. The four key areas of discussion were the cointegration t-test critical values, the opening standard deviation and its effect on profitability, and the cointegration t-testing of pairs and its effect on profitability. Finally, the chapter presented how all of these factors and variables combined created a platform for analysis of each portfolio using the Sharpe Ratio.

5.4.2 *Conclusion*

The research proposition is that there exists a set of key independent variables which can be controlled to produce the most profitable, least volatile returns using pairs trading as a trading strategy. These key independent variables are the squared distance at which the trade is opened (the opening standard deviation) and the t-test cointegration score.

Based on the discussion of results in this chapter, the research proposition was proven to be true. The results of the simulation show that using different combinations of opening standard deviations and cointegration scores for the 100 theoretical portfolios resulted in a variety of returns with varying risk profiles.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

This chapter is a conclusion of the study. The conclusion provides input into recommendations to the target audience, and finally provides a list of suggested further areas for research.

6.2 Conclusions of the study

The objective of the research was to prove whether there exists a set of key independent variables which can be controlled to produce the most profitable, least volatile returns using pairs trading as a trading strategy. These key independent variables are the squared distance at which the trade is opened (the opening standard deviation) and the t-test cointegration score.

The research was successful in its objective, and it indicated a relationship between these key independent variables and the success of the portfolio. In addition, it showed that the more cointegrated the pairs, the more profitable and less volatile the portfolio. However, more cointegrated pairs are rarer than less cointegrated pairs, so there are not as many opportunities to exercise pairs trading using this strategy.

In terms of absolute returns, a cointegrated-only approach may not yield a higher return than an approach where all viable pairs are traded, but in terms of average return, and volatility of return, a cointegrated-only approach reduces risk, while at the same time increases the average net profit per trade.

6.3 Recommendations

Academic researchers are recommended to continue investigating the viability of pairs trading from an academic perspective. As discussed in the literature review, pairs traders tend to limit pairs to being from the same market sector.

Although this research does not outright prove that this is an unwise limitation, it shows that cross-sector pairs trading can be profitable.

Hedge fund traders are recommended to use the Engle-Granger two-step test for cointegration as a measure of suitability of pairs for trading. Yet, they are cautioned to find the balance between selecting only highly cointegrated pairs, which are rare, and those which occur more frequently, but tend to result in lower profits and more volatility.

6.4 **Suggestions for further research**

The research raised a number of questions which future researchers may use as the basis for future studies.

The research looked at two key independent variables which affect pairs trading, namely the cointegration t-statistic and the opening standard deviation. Future studies could look for additional independent variables, and attempt to quantify the effect they have on the profitability and risk of pairs trading.

The research also limited the study to pairs with a cointegration t-statistic of less than -1.5. Future research could include a study of all pairs, regardless of cointegration t-statistic.

Another idea for future research would be to compare the success of pairs trading against current market conditions at the time. These conditions include whether the market is currently trending upward or downward (a bull or bear market), market volatility, liquidity and efficiency.

The research conducted used a time interval of daily trades. Future research could investigate the success of pairs trading when conducted in a high-frequency trading environment, where trades are conducted at intervals of up to sub-seconds.

Pairs trading is not limited to single stock futures. A future study could attempt the same simulation, but across all derivatives, or a different class of derivatives, such as commodities futures.

A final suggested research topic would be to try different strategies for pairs trading. This study used a cash-neutral trading strategy, but there are also studies that have been done using beta-neutral trades, as well as multivariate pairs trading. Multivariate pairs trading involves using more than two derivatives to make up the "pair", and would make a good topic for future research.

REFERENCES

Ackworth, P., Broadie, M., & Glasserman, P. (1997). *A comparison of some Monte Carlo techniques for option pricing* (1st ed.). (H. Hellekalek, P. Larcher, & G. Zinterhof, Eds.) New York: Springer.

Agarwal, V., & Narayan, Y. (2004). Risks and Portfolio Decisions Involving Hedge Funds. *The Review of Financial Studies, 17*(1), 63-98.

Alexander, C. (2001). *Market Models: A Guide to Financial Data Analysis.* Sussex: Wiley.

Anderson, G. (2009). *Unit Trust Fund Holdings and Stock Returns.* MBA Thesis, University of Pretoria.

Andrade, S., di Pietro, V., & Seasholes, M. (2005). *Understanding the Profitability of Pairs Trading.* Berkley: University of California.

Bennett, P. (1995). *Market Timing on the Sydney Stock Exchange Using Derivative Instruments.* MBA Thesis, University of the Witwatersrand.

Bodie, Z., Kane, A., & Marcus, A. (2009). *Investments* (8th ed.). New York: McGraw-Hill.

Chng, M. (2007). Understanding the Risks in and Rewards for Pairs Trading. International Conference of Business and Management Education.

Do, B., Faff, R., & Hamzal, K. (2006). A New Approach to Modelling and Estimation for Pairs Trading. *Proceedings of 2006 Financial Management Association European Conference.* Stockholm.

Elliot, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 813-836.

Elliot, R., van der Hoek, J., & Malcolm, W. (2005). 2005. *Quantitative Finance, 5*(3), 271-276.

Engle, R., & Granger, C. (1987). Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica, 55*(2), 251-276.

Engle, R., & Granger, C. (2003). *Time Series Econometrics: Cointegration and Autoregressive Conditional Heteroskedasticity.* Swedish Prize in Economic Sciences in Memory of Alfred Nobel.

Finance Glossary. (2012, 02 24). *Gearing*. Retrieved from Finance Glossary: http://www.finance-glossary.com/define/gearing/614

Gatev, E., Goetzman, W., & Rouwenhorst, K. (2006). Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies, 19*(3), 797-827.

Granger, C. (2004). *Time Series Analysis, Cointegration, and Applications.* Nobel Lecture, University of California, San Diego.

Groenewald, J. P. (1989). *Social Research: Design and Analysis.* Stellenbosch: University Publishers & Booksellers (Pty) Limited.

Herlemont, D. (2004). *Pairs Trading, Convergence Trading, Cointegration.* YATS Finance & Technologies.

Huberman, G., & Zhenyu, W. (2005). Arbitrage Pricing Theory. Retrieved October 22, 2008, from Arbitrage Pricing Theory: http://www.ny.frb.org/research/staff%5Freports/sr216.html.

Huck, N. (2008). Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research*, 819-825.

Johannesburg Securities Exchange. (2010). *Single Stock Futures Overview*. Retrieved June 22, 2010, from Johannesburg Stock Exchange: http://www.jse.co.za/Products/All-Products/Single-Stock-Futures-Overview.aspx

Johannesburg Stock Exchange. (2010). *History of the JSE*, 1. Retrieved October 15, 2010, from http://www.jse.co.za/About-Us/History-Of-The-JSE.aspx

Johannesburg Stock Exchange. (2012, 02 29). *Single Stock Futures Overview*. Retrieved from Johannesburg Stock Exchange: http://www.jse.co.za/Products/All-Products/Single-Stock-Futures-Overview.aspx

Joy, C., Boyle, P., P., & Tang, K. S. (1996). Quasi-Monte Carlo methods in numerical finance. *Management Science, 42*(6), 926 - 938.

Leedy, P. D., & Ormonde, J. E. (2001). *Practical Research: Planning and Design, Seventh Edition.* Upper Saddle River, NJ: Merrill Prentice Hall.

Mabhunu, M. (2004). *The Market Efficiency Hypothesis and the Behaviour of Stock Returns on the JSE Securities Exchange.* Rhodes University. Rhodes University.

MacKinnon, J., Haug, A., & Michelis, L. (1995). Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration . *Journal of Applied Econometrics, 14*(5), 563-578.

Madhavan, A. N. (2010). Volume-Weighted Average Price (VWAP). Encyclopedia of Quantitative Finance. San Francisco: Barclays Global Investors.

Mkhize, H., & Msweli Mbange, P. (2006, October). A critical review of the restructuring of the South African capital market. *International Review of Business Research Papers, 2*(2), pp. 80-91.

Nath, P. (2003). *High Frequency Pairs Trading with U.S. Treasury Securities: Risks and Rewards for Hedge Funds.* Working Paper, London Business School.

Noriega, A., & Ventosa-Santaularia, D. (2006). *Spurious Cointegration: The Engle-Granger Test in the Presence of Structural Breaks.* Mexico City: Bank of Mexico.

Noriega, A., & Ventosa-Santaularia, D. (2006). *Spurious Cointegration: The Engle-Granger Test in the Presence of Structural Breaks.* Mexico City: Bank of Mexico.

Perlin, M. (2007). *Evaluation of a Pairs Trading Strategy at the Brazilian Financial Market.* Working Paper, University Library of Munich, Germany.

Perlin, M. (2007). *M of a Kind: A Multivariate Approach at Pairs Trading.* Working Paper, University Library of Munich, Germany.

Ross, S. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory, 13*, 341-360.

Sharpe, W. F. (1994). The Sharpe Ratio. *The Journal of Portfolio Management*, 1-49.

Simons, D., & Laryea, S. (2005). *Testing the Efficiency of African Markets .* Social Sciences Research Network.

Standard Bank. (2010, June 22). *Futures Product Guides - Basic Guide.* Retrieved June 22, 2010, from Online Share Trading: https://securities.standardbank.co.za/ost/sp/FrontOfficeSecure/ContentManagement/TransferFile.aspx?resource=\ssfeducationalbrochure.PDF

Vidyamurthy, G. (2004). *Pairs Trading Quantitative Methods and Analysis.* New Jersey: Wiley.

Yule, G. U. (1926). Why do we sometimes get nonsense correlations between time series - a study in sampling and the nature of time series. *Journal of the Royal Statistical Society, 89*, 1-64.

APPENDIX A

Sample Data - Daily Single Stock Future

DATE	DN_NAME	DNSECTOR	CLOSE	HIGH	LOW	VOLUME
05 October 2006	ASXQMAR07	800	98333	98333	98333	0
18 September 2007	ASXQDEC07	800	96231	96231	96231	0
11 January 2007	ASXQJUN07	800	96137	96137	96137	0
26 September 2011	RMCQDEC11	800	95506	95506	95506	0
16 May 2007	ASXQSEP07	800	94411	94411	94411	0
12 December 2005	IMPQDEC05	800	92050	92226	85000	109
16 January 2009	ASXQMAR09	800	91740	91740	91740	0
04 February 2008	ASXQMAR08	800	91231	91231	91231	0
17 April 2009	ASXQJUN09	800	90259	90259	90259	0
06 November 2006	AMSQDEC06	800	88927	89114	83094	538
25 September 2008	ASXQDEC08	800	88198	88198	88198	0
02 June 2008	ASRQSEP08	800	87823	87823	87823	0
07 August 2008	ASXQSEP08	800	87214	87214	87214	11
21 November 2011	RMCQMAR12	800	87202	87202	87202	0
23 April 2008	ASXQJUN08	800	87049	87049	87049	0
09 September 2011	RMCQSEP11	800	85953	85953	85953	0
27 May 2009	ASXQSEP09	800	85667	85667	85667	0
28 August 2006	AMSQSEP06	800	84937	84677	83380	293
15 April 2010	AMSQJUN10	800	83535	83615	82968	1511
15 April 2010	AMXQJUN10	800	83535	83535	83535	1
15 December 2009	AMSQMAR10	800	82110	82110	82110	223
15 April 2010	AMXQSEP10	800	81480	81480	81480	0
15 April 2010	AMSQSEP10	800	81480	81480	81480	0
20 April 2010	ASRQJUN10	800	80831	80831	80831	0
11 December 2009	AMSQDEC09	800	80649	80877	77473	742
10 August 2010	ASRQSEP10	800	78792	78792	78792	0
14 January 2011	AMSQJUN11	800	77656	77656	77656	0
08 March 2010	ASRQMAR10	800	77645	77645	77645	0
15 September 2010	ASRQDEC10	800	76941	76941	76941	6
01 September 2008	ASRQDEC08	800	76611	76611	76611	0
18 January 2011	AMXQMAR11	800	76610	76610	76610	0
14 January 2011	AMSQMAR11	800	76610	76610	76610	460
23 November 2011	RMOQMAR12	800	76441	76441	76441	0
16 November 2011	RMOQDEC11	800	75714	75714	75714	0
13 July 2010	AMSQDEC10	800	74986	74986	74986	1
13 July 2010	AMXQDEC10	800	74986	74986	74986	0
12 September 2005	IMPQSEP05	800	73272	72920	70935	28
24 November 2009	ASRQDEC09	800	72811	72811	72811	0
15 September 2009	AMSQSEP09	800	72026	72026	72026	523
10 May 2006	AMSQJUN06	800	70816	72312	69777	146
08 March 2010	AMXQMAR10	800	70582	70582	70582	12
29 April 2002	IMPQJUN02	800	70330	70330	70330	0
29 April 2002	IMPQSEP02	800	70330	70330	70330	0
14 October 2008	AMSQMAR09	800	68685	68685	68685	40
22 May 2002	ANGQSEP02	800	68374	68374	68374	0

DATE	DN_NAME	DNSECTOR	CLOSE	HIGH	LOW	VOLUME
14 March 2002	IMPQMAR02	800	67420	67420	67420	0
05 April 2011	AMSQSEP11	800	66727	66727	66727	0
22 May 2002	ANGQJUN02	800	66700	66700	66700	0
14 January 2004	IMPQJUN04	800	65095	65095	65095	0
09 June 2011	RMOQSEP11	800	64654	64654	64654	0
31 October 2003	IMPQDEC03	800	64050	63927	63175	10
31 October 2003	IMPQMAR04	800	64050	64050	64050	0
06 November 2002	IMPQDEC02	800	63600	63186	63186	2
06 November 2002	IMPQMAR03	800	63600	63600	63600	0
06 July 2011	AMSQDEC11	800	63309	63309	63309	0
15 September 2009	ASRQSEP09	800	63023	63023	63023	3
05 June 2009	AMSQJUN09	800	62660	64087	62445	792
12 July 2007	LONQSEP07	800	62364	62029	62029	49
09 September 2003	IMPQSEP03	800	61739	61122	57856	105
22 January 2003	IMPQJUN03	800	60800	60800	60800	0
23 September 2002	ANGQDEC02	800	60691	60691	60691	0
17 December 2002	ANGQMAR03	800	60390	61950	60209	79
28 May 2008	AGQQSEP08	800	60134	60134	60134	0
15 June 2005	IMPQJUN05	800	59051	59051	59051	0
28 October 2011	AMQQDEC11	800	58658	58658	58658	0
28 October 2011	AMSQMAR12	800	58622	58622	58622	1
15 June 2007	LONQJUN07	800	58588	58588	58588	0
15 February 2002	ANGQMAR02	800	57800	57800	57800	0
19 June 2008	AGLQSEP08	800	56992	56741	43000	4731
12 March 2008	LONQJUN08	800	56666	56666	56666	10
23 August 2004	IMPQDEC04	800	56232	56232	56232	0

APPENDIX B

Define and analyse a hedge fund trading strategy based on the concept of pairs trading using cointegration.					
Sub-problem	**Source of theory (the literature used in your literature review)**	**Hypotheses or Propositions or Research questions**	**Source of data**	**Type of data**	**Analysis**
Define and analyse a derivative-trading strategy created using the mathematical principles of cointegration.	Alexander, Vidyamurthy, Engel, Granger	The opening deviation and the cointegration t-statistic affect the profitability and volatility of a pairs trading strategy on the JSE.	Commercially-available subscription from Sharenet.	Secondary Data (Closing Futures Prices on the JSE)	Time Series Analysis using a stratified quasi-Monte Carlo simulation

Table 10 - Consistency Matrix

APPENDIX C

Complete Table of Sharpe Values

Opening Ratio	Score Group	Return	Risk	Sharpe Ratio
2.5	-10.499987 to -9.999987	2.997%	0.001004568	13.24559504
2.5	-8.99999 to -8.49999	3.056%	0.001773335	7.8342928
2.5	-9.499985 to -8.999985	3.079%	0.001826407	7.731074549
2	-10.499987 to -9.999987	3.069%	0.002211199	6.33975212
2	-9.499985 to -8.999985	3.112%	0.002355369	6.138324604
1	-10.499987 to -9.999987	3.116%	0.003104538	4.668419323
2.5	-7.999988 to -7.499988	3.164%	0.003291944	4.547696344
2	-8.99999 to -8.49999	3.189%	0.003429008	4.438250907
1.5	-10.499987 to -9.999987	3.164%	0.003377915	4.43248747
2.5	-9.999986 to -9.499986	2.814%	0.002966137	3.867654068
1	-8.499989 to -7.999989	3.296%	0.004212745	3.867238435
2.5	-7.499987 to -6.999987	3.007%	0.003503826	3.825193345
1.5	-8.499989 to -7.999989	3.332%	0.004813627	3.459784999

Opening Ratio	Score Group	Return	Risk	Sharpe Ratio
1.5	-8.99999 to -8.49999	3.293%	0.004811981	3.37999045
2.5	-8.499989 to -7.999989	3.297%	0.004976575	3.275610152
0.5	-10.499987 to -9.999987	3.232%	0.004794005	3.265070133
2	-8.499989 to -7.999989	3.337%	0.005373242	3.10808272
2.5	-10.999988 to -10.499988	3.273%	0.005256488	3.056239943
2	-10.999988 to -10.499988	3.300%	0.005657516	2.887573777
2	-7.999988 to -7.499988	3.470%	0.006744427	2.673167137
0.5	-4.999988 to -4.499988	18.541%	0.063231303	2.668741272
0.5	-8.499989 to -7.999989	3.579%	0.007256292	2.635919967
2.5	-11.99999 to -11.49999	3.273%	0.00612192	2.623873044
2	-5.99999 to -5.49999	3.448%	0.006883822	2.588141337
2.5	-5.99999 to -5.49999	3.249%	0.006187127	2.557940463
1.5	-7.999988 to -7.499988	3.496%	0.00724177	2.525925258
0.5	-4.499987 to -3.999987	37.338%	0.141982539	2.512379939
1	-4.999988 to -4.499988	15.491%	0.056074287	2.465289672
1	-7.999988 to -7.499988	3.499%	0.007473455	2.452176873
1	-4.499987 to -3.999987	33.126%	0.128581138	2.446660118
1.5	-4.499987 to -3.999987	29.051%	0.112134759	2.442119114
0.5	-7.499987 to -6.999987	3.422%	0.007358719	2.386002744
1.5	-4.999988 to -4.499988	12.330%	0.045219452	2.358136793
1.5	-10.999988 to -10.499988	3.403%	0.007469402	2.324541378
0.5	-7.999988 to -7.499988	3.556%	0.008280282	2.28192989
2	-4.499987 to -3.999987	25.404%	0.104576024	2.26981956
0.5	-10.999988 to -10.499988	3.473%	0.008165811	2.212469422
1	-3.999986 to -3.499986	48.969%	0.213843144	2.212012087
1	-11.499989 to -10.999989	3.444%	0.008038214	2.211300063
0.5	-3.999986 to -3.499986	54.612%	0.242203299	2.185981062
1.5	-5.99999 to -5.49999	4.009%	0.010800554	2.169064007
2	-3.499985 to -2.999985	55.946%	0.252913659	2.146165997
2	-1.999982 to -1.499982	71.343%	0.325738607	2.139023715
1	-10.999988 to -10.499988	3.474%	0.008497063	2.126688756
2.5	-2.999984 to -2.499984	62.514%	0.286888734	2.120937839
1	-7.499987 to -6.999987	3.389%	0.008132921	2.117385873
1	-3.499985 to -2.999985	67.284%	0.310193255	2.115363507
1.5	-3.499985 to -2.999985	61.897%	0.284832121	2.114596594
2	-2.999984 to -2.499984	68.876%	0.318366367	2.111059492
1.5	-3.999986 to -3.499986	43.477%	0.198418512	2.107183231
0.5	-3.499985 to -2.999985	73.174%	0.339468475	2.106454909
2.5	-2.499983 to -1.999983	67.148%	0.311229988	2.103947757
2.5	-3.499985 to -2.999985	49.771%	0.22892891	2.101270285
1	-5.99999 to -5.49999	4.022%	0.011220878	2.099296677
1.5	-1.999982 to -1.499982	77.133%	0.36173019	2.086260065
2.5	-1.999982 to -1.499982	65.191%	0.305255448	2.081009412
1	-1.999982 to -1.499982	83.170%	0.392132559	2.078465705
2	-9.999986 to -9.499986	2.708%	0.00501243	2.078240781
1.5	-2.999984 to -2.499984	75.136%	0.354600387	2.071882816
2.5	-5.499989 to -4.999989	5.849%	0.02021515	2.069022398
1.5	-5.499989 to -4.999989	8.089%	0.031201997	2.058310803
2	-3.999986 to -3.499986	38.220%	0.177718809	2.056818689
1.5	-7.499987 to -6.999987	3.264%	0.007855326	2.032901206

Opening Ratio	Score Group	Return	Risk	Sharpe Ratio
2	-5.499989 to -4.999989	6.875%	0.025655175	2.030081252
0.5	-1.999982 to -1.499982	89.057%	0.430878608	2.028194574
2	-11.499989 to -10.999989	3.457%	0.008894107	2.012446691
2	-2.499983 to -1.999983	73.821%	0.358916155	2.010346763
1	-2.999984 to -2.499984	81.434%	0.397218564	2.008154385
0.5	-2.999984 to -2.499984	87.207%	0.426837317	2.004060554
1.5	-9.499985 to -8.999985	3.659%	0.009959949	1.999951851
2.5	-3.999986 to -3.499986	32.774%	0.155819639	1.99635117
2	-4.999988 to -4.499988	10.322%	0.0446589	1.938179711
2.5	-4.499987 to -3.999987	21.000%	0.099916695	1.934952131
1.5	-2.499983 to -1.999983	80.061%	0.405574905	1.932906463
2.5	-6.499985 to -5.999985	3.810%	0.011278275	1.900709103
1	-2.499983 to -1.999983	86.375%	0.456153228	1.857006033
1.5	-11.499989 to -10.999989	3.490%	0.009884544	1.844715265
0.5	-2.499983 to -1.999983	92.700%	0.511311818	1.780393667
2	-7.499987 to -6.999987	3.308%	0.009230283	1.778134659
1	-5.499989 to -4.999989	9.307%	0.042999272	1.77691134
2.5	-11.499989 to -10.999989	3.512%	0.010526896	1.752894247
2.5	-4.999988 to -4.499988	8.112%	0.037209752	1.73208276
2	-6.499985 to -5.999985	4.030%	0.013649229	1.731829673
2	-11.99999 to -11.49999	3.537%	0.011067326	1.68989025
1.5	-6.499985 to -5.999985	4.344%	0.015990453	1.674543705
1	-8.99999 to -8.49999	3.649%	0.01186972	1.670301115
0.5	-6.499985 to -5.999985	4.677%	0.018082101	1.664810171
0.5	-5.99999 to -5.49999	4.638%	0.018192893	1.633139788
1	-6.499985 to -5.999985	4.437%	0.017010772	1.628478805
0.5	-11.499989 to -10.999989	3.833%	0.013583987	1.594808787
0.5	-5.499989 to -4.999989	10.328%	0.054382304	1.592594746
0.5	-8.99999 to -8.49999	3.751%	0.01323786	1.574760247
1	-9.499985 to -8.999985	4.149%	0.016316285	1.521426825
0.5	-9.499985 to -8.999985	4.361%	0.019316216	1.395089411
1.5	-11.99999 to -11.49999	3.958%	0.01918396	1.194567884
0.5	-11.99999 to -11.49999	4.054%	0.020665499	1.15537916
1	-11.99999 to -11.49999	4.112%	0.022172373	1.102664398
1.5	-9.999986 to -9.499986	2.497%	0.008887159	0.934254248
0.5	-9.999986 to -9.499986	2.416%	0.011837713	0.633196064
1	-9.999986 to -9.499986	2.359%	0.011976821	0.578178105

APPENDIX D

Source Code for Simulation

```
Sub AddFormulas()
    'Add Summary Formulas
    Application.Calculation = xlManual
    Range("SolverMin").Formula = "=SUMXMY2(DynamicAdjustedShares,DynamicComparisonShares)"
    Range("AdditiveFactor").Value = 1
    'Range("MultiplicativeFactor").Value = 1
    Range("StdDev") = "=STDEVP(ShareDeviation)"
    Range("DynamicSolverMin") = "=SUMXMY2(DynamicAdjustedShares,DynamicComparisonShares)"
    Range("DynamicStdDev") = "=STDEVP(DynamicComparisonShares,DynamicAdjustedShares)"
    'Add Summary Labels
    Range("AdditiveFactor").Offset(0, -1).Value = "AdditiveFactor"
    Range("SolverMin").Offset(0, -1).Value = "SolverMin"
    Range("MultiplicativeFactor").Offset(0, -1).Value = "MultiplicativeFactor"
    Range("StdDev").Offset(0, -1).Value = "StdDev"
    Range("DynamicSolverMin").Offset(0, -1).Value = "Dynamic Solver Min"
    Range("DynamicStdDev").Offset(0, -1).Value = "Dynamic Std Dev"
    'Add Column Formulas
    Range("AdjustedShares").Formula = "=SUM(CurrentShares*MultiplicativeFactor)"
    'Range("AdjustedShares").Formula = "=SUM(CurrentShares*MultiplicativeFactor+AdditiveFactor)"
    Range("ShareDeviation").Formula = "=(DynamicAdjustedShares-DynamicComparisonShares)^2"
    Application.Calculation = xlAutomatic
End Sub

Sub runPairsSimulation(objRecordSet As ADODB.Recordset)
    Dim myCounter As Integer
    Dim theCounter As Integer
    Dim inTheDeal As Boolean
    Dim currentShare As String
    Dim comparisonShare As String
    Dim currentSharePrice As Double
    Dim comparisonSharePrice As Double
```

```
Dim stdDeviation As Double
Dim currentDeviation As Double
Dim multiplicativeFactor As Double
Dim stdDeviationFactor As Double
Dim maxDeviationFactor As Double
Dim longPurchasePrice As Double
Dim shortSellingPrice As Double
Dim longCurrentPrice As Double
Dim shortCurrentPrice As Double
Dim longShare As String
Dim shortShare As String
Dim pctCostOfBorrowing As Double
Dim pctCostOfTransacting As Double
Dim shortBorrowPrice As Double
Dim shortPrice As Double
Dim longPrice As Double
Dim resultsCounter As Integer
Dim results(0 To 21) As String
Dim stopTradingDays As Integer
Dim closeDeviationFactor As Double
Dim netResult As Double
Dim numberBorrowed As Double
Dim numberSold As Double
Dim shareDistances() As Double
Dim startPrice(0 To 1) As Double
Dim tradingDate As String
Dim sharePair(1 To 2) As Variant
Dim lastValidDate As Date
Dim FactRecordSet As ADODB.Recordset
Dim ResultsRecordSet As ADODB.Recordset
Dim SimulationDetailsRecordSet As ADODB.Recordset
Dim TranSummaryRecordSet As ADODB.Recordset
Dim sqlCriteria(1 To 3) As String
Dim sqlString As String
Dim loopStdDeviation As Double
Dim loopMaxDeviationFactor As Double
Dim loopDeviationClose As Double
Dim startStdDeviation As Double
Dim startMaxDeviationFactor As Double
Dim startDeviationClose As Double
```

```
Dim endStdDeviation As Double
Dim endMaxDeviationFactor As Double
Dim endDeviationClose As Double
Dim stepStdDeviation As Double
Dim stepMaxDeviationFactor As Double
Dim stepDeviationClose As Double
Dim investMentAmount As Double
Dim scenarioResults(0 To 6) As String
Dim startTime As Date
Dim totalRecords As Double
Dim writeTransaction As Boolean
Dim simulationID As Double
Dim runningTotal As Double
Dim shareRatio As Double
Dim theStdDev As Double
Dim avgDeviation As Double
Dim adjustedClosing As Double
Dim closing As Double
Dim comboClosing As Double
Dim shareName As String
Dim comboShareName As String
Dim numberShares As Integer
Dim comboNumberShares As Integer
Dim longNumberShares As Integer
Dim shortNumberShares As Integer
Dim amountBought As Double
Dim amountSold As Double
Dim costs As Double
Dim transactionDate As Date
Dim tradeAction As Integer
Dim shortRevPrice As Double
Dim longRevPrice As Double
Dim shortRevNumberShares As Integer
Dim shortMarginRequired As Double
Dim marginRequired As Double
Dim marginRevRequired As Double
Dim shortRevExposure As Double
Dim shortExposure As Double
Dim longExposure As Double
Dim longRevExposure As Double
```

```
    Dim transCosts As Double
    Dim shortTransCosts As Double
    Dim shortRevMarginRequired As Double
    Dim revMarginRequired As Double
    Dim numberOfTrades As Integer

    'Initialise For loop variables NOTE - throws an "expression too complex error" if you
do this directly in the forloop
    startStdDeviation = Range("StdDeviationFactor").Offset(0, 1)
    startMaxDeviationFactor = Range("maxDeviationFactor").Offset(0, 1)
    startDeviationClose = Range("deviationClose").Offset(0, 1)

    endStdDeviation = Range("StdDeviationFactor").Offset(0, 2)
    endMaxDeviationFactor = Range("maxDeviationFactor").Offset(0, 2)
    endDeviationClose = Range("deviationClose").Offset(0, 2)

    stepStdDeviation = Range("StdDeviationFactor").Offset(0, 3)
    stepMaxDeviationFactor = Range("maxDeviationFactor").Offset(0, 3)
    stepDeviationClose = Range("deviationClose").Offset(0, 3)

    investMentAmount = Range("investmentAmount") ' Convert to cents

    'Transaction Costs
    pctCostOfBorrowing = Range("pctCostOfBorrowing")
    pctCostOfTransacting = Range("pctCostOfSelling")
    stopTradingDays = Range("stopTradingDays")

    'Initialise Variables
    resultsCounter = 0
    totalRecords = 0
    simulationID = 0

    sharePair(1) = objRecordSet!dn_name
    sharePair(2) = objRecordSet!dn_name_combo
    lastValidDate = objRecordSet!lastValidDate
    shareRatio = objRecordSet!shareRatio
    theStdDev = objRecordSet!stdDeviation
    avgDeviation = objRecordSet!avgDeviation

    'Retrieve Share Prices
```

```
  sqlString = _
        "SELECT A.DN_NAME as ShareName, " + _
        "A.[DATE] AS effDate, " + _
        "A.[DATE] AS OTHERDATE, " + _
        "A.[CLOSE] AS Closing, " + _
        "A.[CLOSE] *  " & shareRatio & " AS AdjustedClosing, " + _
        "B.DN_NAME as ComboShareName," + _
        "B.[DATE] AS ComboEffDate," + _
        "B.[DATE] AS ComboOtherDate," + _
        "B.[CLOSE] AS ComboClosing " + _
"FROM FactDailyShare A INNER JOIN " + _
"FactDailyShare B ON A.DATE = B.DATE INNER JOIN " + _
"DimShareCombo C ON A.DN_NAME = C.dn_name AND " + _
"B.DN_NAME = C.dn_name_combo "

'Range("PairsRunningStatus") = sqlString
'SQL Criteria
sqlCriteria(1) = " WHERE A.DN_NAME = " & "'" & Trim(sharePair(1)) & "'" & " AND
B.DN_NAME = " & "'" & Trim(sharePair(2)) & "'"
sqlCriteria(2) = " AND A.[DATE] > " & lastValidDate & " ORDER BY 2 "

'Get all share prices for the time series.
Set FactRecordSet = getRecordSetFromDB(sqlString, sqlCriteria, 3)

'Loop through all days for that share pair -- Reset back to first day

'Apparently this is necessary to get an accurate count... OK then
On Error Resume Next
FactRecordSet.MoveLast
If FactRecordSet.RecordCount <= 0 Then
  Exit Sub
End If

totalRecords = FactRecordSet.RecordCount

'Get the recordSet we are going to use to write the results back to
sqlCriteria(1) = ""
sqlCriteria(2) = ""
sqlString = "SELECT TOP 1 * FROM FACTTRANSACTIONNEW "
Set ResultsRecordSet = getRecordSetFromDB(sqlString, sqlCriteria, 2)
```

```
  'Get the recordSet we are going to write the details of the parameters used in the simulation back to.
  sqlString = "SELECT TOP 1 * FROM FACTSIMULATION"
  Set SimulationDetailsRecordSet = getRecordSetFromDB(sqlString, sqlCriteria, 2)

  'Get the recordSet we are going to write the summary of the trades for each pair back to.
  sqlString = "SELECT TOP 1 * FROM FACTTRANSACTIONSUMMARY"
  Set TranSummaryRecordSet = getRecordSetFromDB(sqlString, sqlCriteria, 2)

    'Loop through open position variable
    For loopStdDeviation = startStdDeviation To endStdDeviation Step stepStdDeviation
      'Loop throug max deviations before closing
      For loopMaxDeviationFactor = startMaxDeviationFactor To endMaxDeviationFactor Step stepMaxDeviationFactor
        'Loop through close position variable
        For loopDeviationClose = startDeviationClose To endDeviationClose Step stepDeviationClose
          FactRecordSet.MoveFirst
          inTheDeal = False
          runningTotal = 0
          numberOfTrades = 0
          stdDeviationFactor = loopStdDeviation
          maxDeviationFactor = loopMaxDeviationFactor
          closeDeviationFactor = loopDeviationClose
          'Write the simulation details to the simulation table
          SimulationDetailsRecordSet.AddNew
          SimulationDetailsRecordSet!openingStdDev = loopStdDeviation
          SimulationDetailsRecordSet!closingStdDev = loopDeviationClose
          SimulationDetailsRecordSet!costOfBorrowing = pctCostOfBorrowing
          SimulationDetailsRecordSet!costOfSelling = pctCostOfTransacting
          SimulationDetailsRecordSet!stopTradingDaysBefore = stopTradingDays
          SimulationDetailsRecordSet!amountToInvest = investMentAmount
          SimulationDetailsRecordSet!maximumDeviation = loopMaxDeviationFactor
          SimulationDetailsRecordSet.Update
          SimulationDetailsRecordSet.MoveNext
          simulationID = SimulationDetailsRecordSet!ID

          Do While Not FactRecordSet.EOF
```

```
'Initialise Variables
adjustedClosing = FactRecordSet!adjustedClosing
closing = FactRecordSet!closing
comboClosing = FactRecordSet!comboClosing
shareName = FactRecordSet!shareName
comboShareName = FactRecordSet!comboShareName
transactionDate = FactRecordSet!effDate
'If not in a trade
If Not inTheDeal Then
  'If shares have diverged, open the trade
  'If entrance criteria match
  If ((adjustedClosing - comboClosing) ^ 2 > avgDeviation + stdDeviationFactor * theStdDev) And _
    (FactRecordSet.AbsolutePosition + stopTradingDays < totalRecords) And _
    ((adjustedClosing - comboClosing) ^ 2 < avgDeviation + maxDeviationFactor * theStdDev) Then

    'enter trade
    inTheDeal = True
    'Get Number of Shares Right.  You can't buy half a share.
    If shareRatio < 1 Then
        comboNumberShares = Round(1 / shareRatio, 0)
        numberShares = 1
    Else
        comboNumberShares = 1
        numberShares = Round(shareRatio, 0)
    End If
    If adjustedClosing > comboClosing Then
      shortShare = shareName
      shortPrice = closing
      shortNumberShares = numberShares * -1
      longShare = comboShareName
      longPrice = comboClosing
      longNumberShares = comboNumberShares
    Else
      shortShare = comboShareName
      shortPrice = comboClosing
      shortNumberShares = comboNumberShares * -1
      longShare = shareName
```

```
    longPrice = closing
    longNumberShares = numberShares
End If

'Handle the Long Trade
amountBought = longNumberShares * longPrice
'Convert to ZAR
amountBought = amountBought / 100
longExposure = amountBought * 100
marginRequired = amountBought * 10 * -1
'Calculate Costs
transCosts = pctCostOfTransacting * amountBought * -1
'Add to CashBalance
runningTotal = runningTotal + marginRequired + transCosts

'Write to the DB
ResultsRecordSet.AddNew
ResultsRecordSet!theDate = transactionDate
ResultsRecordSet!dn_name = longShare
ResultsRecordSet!quantity = longNumberShares
ResultsRecordSet!lineValue = amountBought
ResultsRecordSet!closing = longPrice
ResultsRecordSet!Total = runningTotal
ResultsRecordSet!simulationReference = simulationID
ResultsRecordSet!stampDate = Date
ResultsRecordSet!comboShare = shortShare
ResultsRecordSet!exposure = longExposure
ResultsRecordSet!marginRequired = marginRequired
ResultsRecordSet!stampTime = Time
ResultsRecordSet!costs = transCosts
ResultsRecordSet!Action = 1
ResultsRecordSet.Update

'Handle the Short Trade
amountSold = shortNumberShares * shortPrice
'Convert to ZAR
amountSold = amountSold / 100
shortExposure = amountSold * 100
shortMarginRequired = amountSold * 10
'Calculate Costs
```

```
shortTransCosts = pctCostOfTransacting * amountSold
'Add to CashBalance
runningTotal = runningTotal + shortMarginRequired +
shortTransCosts
'Write the Transaction to the database

'Write to the DB
ResultsRecordSet.AddNew
ResultsRecordSet!theDate = transactionDate
ResultsRecordSet!dn_name = shortShare
ResultsRecordSet!quantity = shortNumberShares
ResultsRecordSet!lineValue = amountSold
ResultsRecordSet!closing = shortPrice
ResultsRecordSet!Total = runningTotal
ResultsRecordSet!simulationReference = simulationID
ResultsRecordSet!stampDate = Date
ResultsRecordSet!comboShare = longShare
ResultsRecordSet!exposure = shortExposure
ResultsRecordSet!marginRequired = shortMarginRequired
ResultsRecordSet!stampTime = Time
ResultsRecordSet!costs = shortTransCosts
ResultsRecordSet!Action = 2
ResultsRecordSet.Update



numberOfTrades = numberOfTrades + 2
End If
Else
'If exit criteria match
If ((adjustedClosing - comboClosing) ^ 2 <= avgDeviation +
closeDeviationFactor * theStdDev) Or _
(FactRecordSet.AbsolutePosition >= totalRecords) Or _
((adjustedClosing - comboClosing) ^ 2 >= avgDeviation +
maxDeviationFactor * theStdDev) Then
inTheDeal = False
'Reverse the Trade, just using the prices of the day.
If shortShare = shareName Then
shortRevPrice = closing
longRevPrice = comboClosing
Else
```

```
    shortRevPrice = comboClosing
    longRevPrice = closing
End If
'Handle the Reversal of the Short Trade
'i.e. Go long on it by the same qty you shorted it earlier.
shortRevNumberShares = shortNumberShares * -1
amountBought = shortRevNumberShares * shortRevPrice
'Convert to ZAR
amountBought = amountBought / 100
'Calculate Exposure
shortRevExposure = amountBought * 100
'This is the margin call or profit at the end of the trade
shortRevMarginRequired = (shortExposure + shortRevExposure +_
shortMarginRequired) * -1
'Calculate Costs
transCosts = pctCostOfTransacting * amountBought * -1
'Add to CashBalance
runningTotal = runningTotal + shortRevMarginRequired
runningTotal = runningTotal + transCosts


'Write to the DB
ResultsRecordSet.AddNew
ResultsRecordSet!theDate = transactionDate
ResultsRecordSet!dn_name = shortShare
ResultsRecordSet!quantity = shortRevNumberShares
ResultsRecordSet!lineValue = amountBought
ResultsRecordSet!closing = shortRevPrice
ResultsRecordSet!Total = runningTotal
ResultsRecordSet!simulationReference = simulationID
ResultsRecordSet!stampDate = Date
ResultsRecordSet!comboShare = longShare
ResultsRecordSet!exposure = shortRevExposure
ResultsRecordSet!marginRequired = shortRevMarginRequired
ResultsRecordSet!stampTime = Time
ResultsRecordSet!costs = transCosts
ResultsRecordSet!Action = 3
ResultsRecordSet.Update

'Handle the Reversal of the Long Trade
```

```
                    '(i.e. Short it by the same qty you longed it earlier)
                    amountSold = longNumberShares * longRevPrice * -1
                    'Convert to ZAR
                    amountSold = amountSold / 100
                    longRevExposure = amountSold * 100
                    'This is the margin call or profit at the end of the trade
                    revMarginRequired = (longExposure + longRevExposure +
marginRequired) * -1
                    'Calculate Costs
                    transCosts = pctCostOfTransacting * amountSold
                    'Add to CashBalance
                    runningTotal = runningTotal + revMarginRequired + transCosts


                    'Write to the DB
                    ResultsRecordSet.AddNew
                    ResultsRecordSet!theDate = transactionDate
                    ResultsRecordSet!dn_name = longShare
                    ResultsRecordSet!quantity = longNumberShares * -1
                    ResultsRecordSet!lineValue = amountSold
                    ResultsRecordSet!closing = longRevPrice
                    ResultsRecordSet!Total = runningTotal
                    ResultsRecordSet!simulationReference = simulationID
                    ResultsRecordSet!stampDate = Date
                    ResultsRecordSet!comboShare = shortShare
                    ResultsRecordSet!exposure = longRevExposure
                    ResultsRecordSet!marginRequired = revMarginRequired
                    ResultsRecordSet!stampTime = Time
                    ResultsRecordSet!costs = transCosts
                    ResultsRecordSet!Action = 4
                    ResultsRecordSet.Update

                    numberOfTrades = numberOfTrades + 2
                  End If
                End If
                'End Loop through the day
                FactRecordSet.MoveNext
            Loop
            'Write Summary back to the database
            TranSummaryRecordSet.AddNew
```

```
            TranSummaryRecordSet!dn_name = shareName
            TranSummaryRecordSet!dn_name_combo = comboShareName
            TranSummaryRecordSet!numberOfTrades = numberOfTrades
            TranSummaryRecordSet!nettProfit = runningTotal
            TranSummaryRecordSet!simulationID = simulationID
            TranSummaryRecordSet!dateRun = Date
            TranSummaryRecordSet.Update
        Next loopDeviationClose
    Next loopMaxDeviationFactor
  Next loopStdDeviation
  FactRecordSet.Close
  ResultsRecordSet.Close
  SimulationDetailsRecordSet.Close
  TranSummaryRecordSet.Close
End Sub

Sub RunSimulationController()
  'Look at Excel field to work out whether to run for all Shares, or a specific combination
  'Initialise Variables
  Dim objRecordSet As ADODB.Recordset
  Dim getStatsForAllShares As String
  Dim getStatsForThisShare As String
  Dim sqlCriteria(1 To 3) As String
  Dim sqlString As String
  Dim sharePair(1 To 2) As Variant
  Range("pairsRunningStatus") = "Started at " & Now()

  Application.ScreenUpdating = False
  'Set Variables from Sheet
  getStatsForAllShares = Range("getStatsForAllShares").Text
  getStatsForThisShare = Range("getStatsForThisShare").Text
  sqlString = "SELECT " + _
    "A.dn_name, " + _
    "A.dn_name_combo, " + _
    "d.dn_name AS prev_dn_name, " + _
    "d.dn_name_combo AS prev_dn_name_combo, " + _
    "d.cointScore, " + _
    "d.matchingRows, " + _
    "d.shareRatio, " + _
```

```
        "d.stdDeviation, " + _
        "d.avgDeviation " + _
    "FROM DimShareCombo A " + _
    "INNER JOIN DimShareLookup B ON " + _
        "A.dn_name = B.dn_name " + _
    "INNER JOIN DimShareLookup C ON " + _
        "A.dn_name_combo = C.dn_name " + _
    "INNER JOIN DimShareCombo D ON " + _
        "B.dn_name_predecessor = D.dn_name and " + _
        "C.dn_name_predecessor = D.dn_name_combo "

    sqlString = "SELECT * FROM factSimulationDriver"
    If getStatsForAllShares = "TRUE" Then
        sqlCriteria(1) = ""
    Else
        sqlCriteria(1) = " WHERE DN_NAME = " & "'" & Trim(getStatsForThisShare) & "'
AND COINTSCORE < -1.5 AND DN_NAME_COMBO = 'SBKQJUN05'"
    End If
    sqlCriteria(2) = "  ORDER BY 1 , 2 "
    'Grab all Shares Combos that are valid for the simulation
    Set objRecordSet = getRecordSetFromDB(sqlString, sqlCriteria, 3)
    'Loop through all shares selected
    Do While Not objRecordSet.EOF
        'Update the screen
        Application.ScreenUpdating = True
        Range("PairsRunningStatus") = "Still Running at " & Now() & " on " &
objRecordSet!dn_name & objRecordSet!dn_name_combo
        Application.ScreenUpdating = False
        'Start for loop through recordset
        Call runPairsSimulation(objRecordSet)
        objRecordSet.MoveNext
    'End loop through recordset of shares
    Loop
    'Close RecordSet
    objRecordSet.Close
    Range("PairsRunningStatus") = "Completed at " & Now()
End Sub

Sub StorePairsStatsController()
```

```
    'Look at Excel field to work out whether to run for all Shares, or a specific combination
    'Initialise Variables
    Dim objRecordSet As ADODB.Recordset
    Dim getStatsForAllShares As String
    Dim getStatsForThisShare As String
    Dim sqlCriteria(1 To 3) As String
    Dim sqlString As String
    Dim sharePair(1 To 2) As Variant
    Range("pairsRunningStatus") = "Started at " & Now()

    Application.ScreenUpdating = False
    'Set Variables from Sheet
    getStatsForAllShares = Range("getStatsForAllShares").Text
    getStatsForThisShare = Range("getStatsForThisShare").Text
    sqlString = "SELECT DN_NAME, DN_NAME_COMBO, BETA, COINTSCORE, NOROWS, ALPHA FROM DIMSHARECOMBO "

    If getStatsForAllShares = "TRUE" Then
        sqlCriteria(1) = ""
    Else
        sqlCriteria(1) = " WHERE DN_NAME = " & "'" & Trim(getStatsForThisShare) & "'"
    End If
    sqlCriteria(2) = "  ORDER BY 1"

    'This section Picks up the list of shares from DB
    'Get RecordSet
    'Connect to DB
    'Grab all Shares Combos that are valid
    Set objRecordSet = getRecordSetFromDB(sqlString, sqlCriteria, 2)
    'Loop through all shares selected
    Do While Not objRecordSet.EOF
        'Update the screen
        Application.ScreenUpdating = True
        Range("PairsRunningStatus") = "Still Running at " & Now() & " on " & objRecordSet!dn_name
        Application.ScreenUpdating = False
        'Start for loop through recordset
        sharePair(1) = objRecordSet!dn_name
        sharePair(2) = objRecordSet!dn_name_combo
```

```
        Call storePairsStats(sharePair, objRecordSet)
        'Update Back to the database -- Arguably, this could be done at the end.
        objRecordSet.Update
        objRecordSet.MoveNext
    'End loop through recordset of shares
    Loop

    'Close RecordSet
    objRecordSet.Close
    Range("PairsRunningStatus") = "Completed at " & Now()
End Sub


Sub storePairsStats(sharePair() As Variant, rs As ADODB.Recordset)
    'This function writes the statistical variables (e.g. Cointegration, etc,) back to the
recordSet
    'After calculating them
    'Initialise Variables
    Dim sqlCriteria(1 To 3) As String
    Dim sqlString As String
    Dim rowsAffected As Integer
    'Fill Sheet with share prices for both shares
    'SQL Statement
    sqlString = _
        "SELECT A.DN_NAME, " + _
        "A.[DATE] AS EFFDATE, " + _
        "A.[DATE] AS OTHERDATE, " + _
        "A.HIGH, " + _
        "A.LOW, " + _
        "A.[CLOSE] AS CLOSING, " + _
        "A.VOLUME," + _
        "NULL AS FORMULASPOT, " + _
        "B.DN_NAME," + _
        "B.[DATE] AS EFFDATE," + _
        "B.[DATE] AS OTHERDATE," + _
        "B.HIGH," + _
        "B.LOW," + _
        "B.[CLOSE] AS CLOSING," + _
        "B.VOLUME " + _
  "FROM FactDailyShare A INNER JOIN " + _
```

```
  "FactDailyShare B ON A.DATE = B.DATE INNER JOIN " + _
  "DimShareCombo C ON A.DN_NAME = C.dn_name AND " + _
  "B.DN_NAME = C.dn_name_combo "

  'Range("PairsRunningStatus") = sqlString
  'SQL Criteria
  sqlCriteria(1) = " WHERE A.DN_NAME = " & "'" & Trim(sharePair(1)) & "'" & " AND
B.DN_NAME = " & "'" & Trim(sharePair(2)) & "'"
  sqlCriteria(2) = " AND A.[DATE] > DATEADD(day,-90,c.lastValidDate) ORDER BY 2 "

  'Connect to DB and Write to Sheet
  rowsAffected = getSharesFromDB(sqlString, sqlCriteria, 0, "TempAnalysisStart")
  Range("totalRows") = rowsAffected

  If rowsAffected > 0 Then
    'Fill Formulas, etc.
    Call AddFormulas
    'Call Solver here? Any other calls that need to be made?
    'Get Variables back from Excel into Array
    rs!cointScore = CDbl(Range("ADFTStat"))
  End If
End Sub

Sub performInitialTransaction(shortPrice As Double, longPrice As Double, runningTotal
As Double)
    'Borrow the shortShare
    runningTotal = runningTotal - AddTransactionCosts(100, shortPrice)
    'Sell the shortShare
    runningTotal = runningTotal + shortPrice - AddTransactionCosts(200, shortPrice)
    'Buy the long share
    runningTotal = runningTotal - longPrice - AddTransactionCosts(300, longPrice)
End Sub

Sub closeTransaction(shortPrice As Double, longPrice As Double, runningTotal As
Double, shortBorrowPrice As Double)
    'Sell the long share
    runningTotal = runningTotal + longPrice - AddTransactionCosts(200, shortPrice)
    'Buy the short share
    runningTotal = runningTotal - shortPrice - AddTransactionCosts(300, shortPrice)
    'Return it to the owner
```

```
    runningTotal = runningTotal - AddTransactionCosts(400, shortPrice)
End Sub

Function AddTransactionCosts(transactionType As Integer, amount As Double) As Double
   Select Case transactionType
    Case 100 'Borrow
        AddTransactionCosts = Range("pctCostOfBorrowing") * amount
    Case 200 'Sell
        AddTransactionCosts = Range("pctCostOfSelling") * amount + fixedCostOfSelling
    Case 300 'Buy
        AddTransactionCosts = Range("pctCostOfBuying") * amount + fixedCostOfBuying
    Case 400 'Return to owner
        AddTransactionCosts = Range("pctCostOfReturning") * amount
    Case Else
   End Select
End Function

Sub writeSimulationResults(results() As String, rowCount As Integer)
    Dim resultsSheet As String
    resultsSheet = "Results"
    'Find first empty row if necessary
    If (rowCount = 0) Then
        'Sheets(resultsSheet).Activate
        'Find First Empty Row
        'Columns("A:A").Select
        'rowCount = ActiveSheet.UsedRange.Rows.Count
        rowCount = Sheets(resultsSheet).UsedRange.Rows.Count
    End If
    'Application.ScreenUpdating = True
    'Write Array of results to Sheet
    Range("ResultsLine").Offset(rowCount, 0) = results
    'Application.ScreenUpdating = False
End Sub

Sub writeScenarioResults(scenarioSummary() As String, rowCount As Integer)
    Dim resultsSheet As String
    resultsSheet = "ScenarioResults"
    'Find first empty row if necessary
    If (rowCount = 0) Then
```

```
        Sheets(resultsSheet).Activate
        'Find First Empty Row
        Columns("A:A").Select
        rowCount = ActiveSheet.UsedRange.Rows.Count
    End If
    'Write Array of results to Sheet
    Range("SummaryResults").Offset(rowCount, 0) = scenarioSummary
End Sub

Sub writeTransactionSteps(shortPrice As Double, numberShort As Double, longPrice As Double, numberLong As Double, runningTotal As Double, _
                          shortShare As String, longShare As String, currentShare As String, comparisonShare As String, tradingDate As String)
    Dim resultsSheet As String
    Dim tradingSummary(1 To 10) As Variant
    tradingSummary(1) = currentShare
    tradingSummary(2) = comparisonShare
    tradingSummary(3) = tradingDate
    tradingSummary(4) = shortShare
    tradingSummary(5) = shortPrice
    tradingSummary(6) = numberShort
    tradingSummary(7) = longShare
    tradingSummary(8) = longPrice
    tradingSummary(9) = numberLong
    tradingSummary(10) = runningTotal
    resultsSheet = "TradingSteps"
    'Find first empty row if necessary
    If (rowCount = 0) Then
        rowCount = Sheets(resultsSheet).UsedRange.Rows.Count
        'Find First Empty Row
        'Columns("A:A").Select
        'rowCount = ActiveSheet.UsedRange.Rows.Count
    End If
    'Write Array of results to Sheet
    'Range("TradingSteps").Offset(rowCount, 0) = tradingSummary
    Range("TradingStepsLine").Offset(rowCount, 0) = tradingSummary
End Sub

Sub Analyse()
' Analyse Macro
```

```
'   ActiveCell.Offset(0, -3).Columns("A:A").EntireColumn.Select
    Selection.Insert Shift:=xlToRight, CopyOrigin:=xlFormatFromLeftOrAbove
    ActiveCell.Offset(1, 0).Range("A1").Select
    ActiveCell.FormulaR1C1 = "=RC[1]&RC[2]"
    ActiveCell.Select
    Selection.AutoFill Destination:=ActiveCell.Range("A1:A553")
    ActiveCell.Range("A1:A553").Select
    Sheets("ShareComparison").Select
    ActiveCell.Columns("A:A").EntireColumn.Select
    Selection.Insert Shift:=xlToRight, CopyOrigin:=xlFormatFromLeftOrAbove
    ActiveCell.Offset(1, 0).Range("A1").Select
    ActiveCell.FormulaR1C1 = "=RC[1]&RC[2]"
    ActiveCell.Select
    Selection.AutoFill Destination:=ActiveCell.Range("A1:A552")
    ActiveCell.Range("A1:A552").Select
    ActiveCell.Offset(0, 8).Columns("A:A").EntireColumn.Select
    Selection.Insert Shift:=xlToRight, CopyOrigin:=xlFormatFromLeftOrAbove
    ActiveCell.Offset(1, 0).Range("A1").Select
    ActiveCell.FormulaR1C1 = "=VLOOKUP(RC[-8],Results!R2C1:R554C4,4,FALSE)"
    ActiveCell.Select
    ActiveCell.Offset(0, -8).Columns("A:A").EntireColumn.ColumnWidth = 14.43
    Selection.Copy
    ActiveCell.Offset(1, 0).Range("A1").Select
    ActiveSheet.Paste
End Sub

Function getRecordSetFromDB(sqlString As String, sqlCriteria() As String, openMode As Integer) As ADODB.Recordset
    Const adOpenStatic = 3
    Const adOpenDynamic = 2
    Const adLockReadOnly = 1
    Const adLockOptimistic = 3
    Const ForReading = 1
    Set cnn = CreateObject("ADODB.connection")
    cnn.Open "Provider=SQLOLEDB;Data Source=localhost;User ID=mba;Password=*;"
    Set objRecordSet = CreateObject("ADODB.recordset")
    objRecordSet.CursorType = openMode
    For counter = LBound(sqlCriteria) To UBound(sqlCriteria)
        sqlString = sqlString & " " & sqlCriteria(counter)
    Next counter
```

```
    'Enable the line below to debug the SQL string.
    'Range("pairsRunningStatus") = sqlString
    objRecordSet.Open sqlString, cnn, openMode, adLockOptimistic
    Set getRecordSetFromDB = objRecordSet
End Function
```