Pairs Trading: A Copula Approach

Cecilia Augustine

A dissertation submitted to the Department of Actuarial Science, Faculty of Commerce, at the University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

May 27, 2014

Master of Philosophy specializing in Mathematical Finance,
University of Cape Town,
Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy in the University of Cape Town. It has not been submitted before for any degree or examination in any other University.

May 27, 2014

Abstract

Pairs trading is an arbitrage strategy that involves identifying a pair of stocks known to move together historically and trading on them when relative mispricing occurs. The strategy involves shorting the overvalued stock and simultaneously going long on the undervalued stock and closing the positions once the prices have returned to fair values. The cointegration method and the distance method are the most common techniques used in pairs trading strategy. However under these methods, the measure of divergence between the stocks or the spread is assumed to be symmetrically distributed about the mean zero. In addition, the spread is assumed to be a stationary time series (cointegration method) or mean-reverting (distance method). These assumptions are the main drawbacks of these methods and may lead to missed and/or inaccurate trading signals. The purpose of this dissertation is to explore an alternative approach to pairs trading by use of copulas. This dissertation aims to investigate if copulas can improve the profitability of pairs trading. To achieve this aim, results of pairs trading by use of copulas are compared against those of cointegration and distance methods.

Acknowledgements

I would like to give special thanks to my supervisor Mr Petrus Bosman and his colleague Mr Dan Golding of Prescient Securities for their invaluable guidance and assistance during this research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Setting

Pairs trading is a long/short investment strategy popular with hedge funds and institutional investors (Vidyamurthy, 2011, Ch. 5). Its origins date back to the 1980s when it was established by Wall Street quant Nunzio Tartaglia and his team at Morgan Stanley with reported profits in excess of $50 million in 1987 (Gatev, Goetzmann and Rouwenhorst, 2006). Recent studies show that pairs trading continues to be a profitable investment strategy albeit with a declining trend in profitability (Do and Faff, 2010).

Pairs trading involves taking advantage of the temporary mispricing of two stock prices that are known to move together historically (Xie and Wu, 2013). Short-term market disturbances may cause the stocks to respond differently resulting in a temporary shift from the underlying long-term co-movement tendency. If this happens with continued consistency, then profits can be made by executing the pairs trading strategy (Mashele, Terblanche and Venter, 2013). Pairs trading involves shorting the overvalued stock and simultaneously going long on the undervalued stock and closing the positions once the prices have returned to fair values (Xie and Wu, 2013). Profits can be made if on closing the positions the previously overvalued stock is priced lower and the previously undervalued stock is priced higher (Mashele *et al.*, 2013). There are a number of ways of identifying suitable pairs for the strategy. In their seminal paper, Gatev *et al.* (2006) identified suitable pairs by using a minimum distance criterion where they matched stocks by finding two stocks that had the minimum sum of squared deviations of normalised prices. Liew and Wu (2013) note that this is equivalent to the maximum correlation criterion. Alternatively, one could use cointegration analysis which is a formal way of testing whether two time series have a long-term relationship (Utkulu, n.d.).

The two most common pairs trading strategies are the distance method and the cointegration method. Both of these methods rely on the idea of spread as an indication of mispricing. The spread is assumed to be symmetrically distributed about the mean of zero (Liew and Wu, 2013). In these strategies, a trading signal occurs when the spread deviates by more than a pre-specified standard deviation from the mean of zero. Typically, the pre-specified standard deviation would be two standard deviations calculated during the formation period. In a case where the pre-specified standard deviation is more than the actual variation of the spread experienced during the trading period, trading signals may be missed. Conversely, there will be too many trading signals and some of them likely to be inaccurate if the pre-specified standard deviation is less than the actual standard deviation (Xie and Wu, 2013). Refer to Chapter 2 below for more on the spread.

The premise of these approaches is the assumption that there is only a linear relationship between the stocks and hence the use of correlation and cointegration as a measure of dependency. This would give accurate results if the financial data under investigation was normally distributed. Ignoring non-linear dependency turns out to be detrimental in these methods since empirical evidence has shown that financial data is hardly normally distributed in practice (Ling, 2006). Kat (2003) documented that assets tend to exhibit negative skewness (or the high probability of extreme losses) and/or excess kurtosis (high probability of large profits or large loses).

If a method exists that does not impose any assumptions on the data but rather develops a strategy based on the behaviour of the data, then surely this should have a better chance at capturing the dependency structure and indicate optimal trading signals than one that makes assumptions. Admittedly, if the normal assumption of the distribution of the financial data turns out to be correct then the distance and cointegration method should perform well. The problem arises then when the assumptions imposed on the data are not borne out in practice. A model built on incorrect assumptions would not, except by chance, perform better than a model that does not make any assumptions prior any analysis on the data.

With particular reference to the distance method, Xie, Liew, Wu and Zou, 2014 argue that this raises two concerns. Firstly, it is unlikely that the marginal distribution of a stock return is normal and secondly, the spread as a single distance measurement, will not capture the non-linear dependency that the joint dependency structure may have. Xie *et al.* (2014) argue that essentially the distribution of spread is therefore unknown. It is unlikely that the spread is symmetric about zero and that its distribution conditional on normalised prices is no longer valid.

In a nutshell, correlation and cointegration can pose significant limitations in correctly describing the association between the stocks and possible future movements (Kat, 2003) and may lead to inaccurate or no trading signals at all (Liew and

Wu, 2013).

To this end, a proposed solution is the application of copulas in implementing the pairs trading strategy (Liew and Wu, 2013; Xie and Wu, 2013). A copula represents the relationship between two or more variables in terms of the individual marginal cumulative distributions (Sweeting, 2011). Copula methods are said to capture correctly the dependence relation between variables as they allow for a measure of tail dependency, something which can not be fully appreciated by assuming normality of data (Liew and Wu, 2013). Copulas have a relatively young history in statistics, but are increasingly used in finance and economics to analyse dependence structures (Nelsen, 2006; Patton, 2008a). Further details of copulas are provided in Appendix A.

The purpose of this dissertation is to investigate an alternative approach to pairs trading by use of copulas. Because of the attractive characteristics of copulas of doing away with rigid assumptions made by the conventional methods, it is expected that a copula strategy will produce more trading opportunities and may potentially be more profitable than the cointegration and distance approaches. The inspiration and structure of this dissertation follows closely the research paper titled: "Pairs trading: A copula approach" by Liew and Wu (2013).

This dissertation is organised as follows: section 1.2 provides an overview of the relevant research on the pairs trading strategy, its potential profitability and the use of copulas in finance. A brief description of the research methods and aims is given in section 1.3. Chapter 2 provides a detailed account of the methodology used in this dissertation and chapter 3 discusses the empirical results. In Chapter 4, possible limitations and assumptions of the dissertation are discussed. Finally, Chapter 5 concludes with a summary of the main results and recommendations are made for potential future research.

1.2 Literature Review

Recent research by Liew and Wu (2013), Xie and Wu (2013) and Ferreira (2008) gives a strong indication that a copula approach applied in pairs trading is more robust, leads to more trading strategies and is potentially more profitable. A number of authors laud the use of copulas in bivariate and multivariate studies. A summary of the literature reviewed in this dissertation is given below.

- Liew and Wu (2013) compared the results of trading three pairs of stocks known to be highly correlated and cointegrated, using the distance, cointegration and copula methods. Their empirical results showed that a copula approach produced the highest number of trading signals and led to more profits than the conventional methods. The focus of their study was on the profitability of trading pairs already chosen, the authors recommended further research be done in the use of copulas in selecting the pairs of stocks.

- Xie and Wu (2013) noted that despite its good track record, recent research by Do and Faff (2010) indicated a decreasing profitability of the pairs trading strategy. Xie and Wu (2013) proposed a copula based approach and provide a step-by-step algorithm on how to execute the strategy and set out the distance and cointegration approaches as generalised special cases of copula under certain conditions. Xie and Wu (2013) recommended further research to test the profitability of the method and similar to Liew and Wu (2013), the authors also recommended further research on how to incorporate copulas in pairs selection.

- Ferreira (2008) and Stander (2011) provided an algorithm to follow in implementing the copula based pairs trading strategy. Firstly, the process entails determining the individual marginal distributions of the stocks, secondly, these are used to determine an optimal copula function and finally, the conditional probabilities are derived. Using the conditional probabilities, the relative undervalued and overvalued positions of the stock are then identified to trigger trading signals. Ferreira (2008) stated that one of the main advantages of a copula method over the conventional methods is not requiring the rigid assumptions required by conventional methods. A suitable joint distribution for the two stocks can be found regardless of the individual marginal distributions of the stock returns. This in turn provides more information including the shape and nature of the dependency of the stock pairs, while a correlation coefficient is the only resulting information provided when using the distance or cointegration methods. Further literature on choosing the optimal copula is detailed in Nelsen (2006). Ferreira (2008) recommended further research in understanding the relationship between the results of copula method and fundamental analysis.

- Trivedi and Zimmer (2007) document the superiority of copula methods in joint modelling, and in particular when the distributions are non-normal. In the bivariate context, copulas can be used to define the dependence of pairs of random variables. Other listed applications of copulas are in portfolio management, in understanding "broken heart" syndrome in joint annuities, in constructing joint distributions for variables discrete in nature and in generating a flexible joint distribution in dealing with a larger modelling problem.

- Patton (2008a) discussed the application of copulas in economics and finance. He noted that the rising popularity of copulas in finance was as a direct result of the mounting empirical evidence against normality of asset returns and dependencies between assets. Early use of copulas was in risk management where the ability of being able to correctly factor in lower tail dependency is crucial to risk managers. Patton (2008a) discussed "multivariate option pricing, portfolio decisions, credit risk and studies of contagion between financial markets" as other uses of copulas in finance.

- The journal article by Mashele *et al.* (2013) is relevant in that it provides a good reference point for pairs trading in the South African context. Their paper follows the method of Do and Faff (2010) by selecting suitable pairs by considering specific economic groups (e.g. banking or mining) and does not consider the notion of a formation period and trading period as in Gatev *et al.* (2006), Xie and Wu (2013), Ferreira (2008) and Stander (2011). Mashele *et al.* (2013) also provide a detailed method for allowing for trading costs in the strategy. Further areas discussed in their paper are out of the scope of this dissertation.

1.3 Research Method and Aims

The main aim of this dissertation will be to test the hypothesis that a copula based pairs trading strategy leads to more trading opportunities and potentially more profitable returns compared to the distance and cointegration methods. Thus each technique will be applied to the selected pairs and the results compared. The main interest is to investigate highly cointegrated pairs from the top one hundred and three (103) listed stocks on the Johannesburg Stock Exchange (JSE). Using these top listed stocks ensures that we have a set of relatively liquid shares and this will facilitate the process of pair formation and provide sufficient data for the trading exercise (Gatev *et al.*, 2006).

Suitable pairs of stocks that have exhibited a long-term relationship are identified using cointegration techniques.

During the formation period, data is analysed to determine the appropriate marginal distributions and parameters required for each of the distance, cointegration and copula approaches.

The paramaters are then applied in the trading period for each strategy to determine the trading signals. Positions are opened and/or closed in response to trading signals indicated by each strategy.

Matlab was used to carry out the statistical tests and fitting of the copula.

Microsoft Excel was used in determining the trading signals and calculation of the payoff. In addition, the author made use of matlab code titled "Copula toolbox for Matlab, version 1.07, 5apr08" by Patton (2008b). The code is available at the following url: $http://public.econ.duke.edu/\ ap172/code.html$.

With some modification, the code facilitated the calculation of the Log Likelihood function and in the selection of the optimal copulas used in this dissertation. This copula toolbox was considered an extremely useful resource in providing the author with an initial understanding of the relevant coding required.

Chapter 2

Methodology

2.1 Selection of Pairs

Cointegration tests were used to identify pairs of stocks that appear to 'move together' in the long-term. Engle-Granger and Johansen tests were performed on the log closing price series of the top 103 JSE shares available from 06/01/1998 to 13/08/2013 on the Bloomberg database. The result was a grid of all the stocks indicating pairs that were cointegrated. Out of the cointegrated pairs, the pairs were analysed further to consider only the most liquid stocks. Of the most liquid stocks five pairs with the lowest test statistics, i.e. highly cointegrated, were selected for the study. See section 3.1 below for a list of the pairs studied in this dissertation.

Given the short-term nature of the pairs trading strategy, liquidity of stocks is emphasised so that trade execution does not result in significant movement in the market prices and distort the whole purpose of taking advantage of price anomalies. In practice, illiquid stocks may also have higher bid-offer spread.

2.2 Copula Approach

This method follows closely the method implemented by Stander (2011) and Liew and Wu (2013).

During the formation period, the following steps are carried out:

Firstly, the cumulative marginal distributions of the two stocks are estimated. To do this, the closing prices were first normalised to log prices to make them suitable for this step (Liew and Wu, 2013). The marginal distributions can either be determined using a parametric or a non-parametric approach. In a parametric approach, analytical software such as Matlab or Excel is used to fit a known statistical distribution and estimate its parameters. In a non-parametric approach, an empirical marginal distribution function is also estimated by use of statistical software and the result is vector of values uniformly distributed on $[0, 1]$ (Ferreira, 2008).

The second approach is attractive in its simplicity and is the approach used in this dissertation. This step was implemented by the function "*empricialcdf*" written by Patton (2008b).

Secondly, copula parameters are estimated using the function "$copulafit$" in matlab.

For the purpose of this dissertation, the choice of copulas is limited to the Gaussian and Student-t copulas and three of the one-parameter Archimedean Copulas: Gumbel, Frank and Clayton.

The Archimedean Copulas form an important family of copulas as they can be easily constructed (by using a generator function), constitute a large variety of copulas and they possess desirable mathematical characteristics as they can be expressed in closed form (Nelsen, 2006; Ferreira, 2008; Sweeting, 2011).

Each of the five copulas capture upper and lower tail dependency to a different extent and were thus considered to be sufficient to adequately cover all dependency structures that would have ensued from the stock pairs. See section A.3 for further details of tail dependencies of the copulas.

The next step involves determining an optimal copula that best describes the dependency structure of the stocks. Here the negative log-likelihood for each copula was determined by using functions written by Patton (2008b). The functions took in the copula parameter(s) determined above and the cumulative marginal distribution of the stocks and gave an output of the negative log-likelihood. The optimal copula was the one that gave the most negative log-likelihood.

Finally, the conditional marginal distributions, are determined. $MI_{U|V}$ or $Pr(U <= u|V = v)$, is by definition derivative of the copula with respect to v and $MI_{V|U}$ or $Pr(V <= v|U = u)$ is derivative of the copula with respect to u, as shown in the equations below:

$$Pr(U <= u|V = v) = \frac{\partial C(u, v)}{\partial v} \tag{2.1}$$

$$Pr(V <= v|U = u) = \frac{\partial C(u, v)}{\partial u} \tag{2.2}$$

Where u and v are the empirical marginal cumulative distributions for stocks in the selected pair determined above.

If U is fairly priced with respect to V then $MI_{U|V} = 0.5$. If $MI_{U|V} < 0.5$, then U is undervalued relative to V and if $MI_{U|V} > 0.5$, U is overvalued relative to V, where U and V refer to the stocks in the pair selected (Liew and Wu, 2013; Xie and Wu, 2013).

In addition, the higher the value of the conditional distribution, the more confidence there is on the extent of the relative pricing position (Ferreira, 2008). There are closed form formulae for all the conditional marginal distributions for all the copulas considered in this study. A summary of the formulae, drawn from Liew and Wu (2013), for the conditional probability functions is provided in Appendix section A.4 (Table A.1). The conditional formulae for the Gaussian and Student-t copulas are also provided by Cherubini, Luciano and Vecchiato, 2004.

During the trading period the following steps are carried out:

Firstly, the empirical marginal distributions of the stocks during the trading periods are obtained, denoted here $u1$ and $v1$.

Thereafter, using the optimal copula paramater(s) determined during the formation period above together with the marginals $u1$ and $v1$, the conditional distributions $MI_{U|V}$ and $MI_{U|V}$ are calculated.

In this dissertation, an upper bound of 0.95 and a lower bound of 0.05 are used as signals for trading. Therefore, if a marginal conditional distribution crosses the upper/lower bound, a sell/buy signal is indicated. The positions are closed when the marginal conditional distribution crosses the 0.50 mark from above/below (Ferreira, 2008).

Any positions still open at the end of the trading period are closed even if the mispricing has yet to be corrected.

2.3 Distance Method

This section closely follows the method used by Gatev *et al.* (2006).

Under the distance method, the measure of mispricing is defined as the "spread" or the difference between the standardised prices of the two stocks in the pair.

Standardised prices during any period were assumed to be cumulative returns during that period. Therefore, to obtain the standardised price during the formation period, the closing prices were divided by the closing price on the first day of the formation period. Thereafter, the spread, as defined above, and its standard deviation are calculated. This standard deviation is the parameter used in the trading period as described below.

The standardised prices and the spread during the trading period are obtained in a similar fashion as above. When the spread during the trading period deviates by more than two historical standard deviations as estimated during the formation period above, a position is created by buying the relatively cheap stock and selling the relatively overpriced stock.

The position is maintained until the spread reverts to zero, i.e. when the standardised prices cross over. The process is repeated throughout the trading period. Any positions still open at the end of the trading period are closed even if the mispricing has yet to be corrected (Gatev *et al.*, 2006).

2.4 Cointegration Method

Under the cointegration method, the spread is defined as the cointegration error term estimated from the regression analysis of log prices of the two stocks during the formation period (Liew and Wu, 2013). The error term was taken to be the output $reg1.res$ of the function *egcitest* in Matlab where a vector of the log prices of the pair was taken as the input.

Similar to the distance method, the standard deviation of the error terms in the formation period is determined. Thereafter, positions are opened when the spread during the trading period deviates by more than two historical standard deviations as estimated during the formation period.

The positions are then maintained until the spread reverts to zero. The process is repeated throughout the trading period. Any positions still open at the end of the trading period are closed even if the mispricing has yet to be corrected (Gatev *et al.*, 2006).

2.5 Formation and Trading period

The initial formation period was taken to be the first five years of the data (assumed to be the first 1,245 closing prices). For stock pairs where one or both stocks were not listed on the JSE from the initial date of 06/01/1998, and were listed say on $date1$ and $date2$ respectively where $date1 < date2$, then the first 1,245 days were taken from $date2$.

The trading period was then taken to be the remainder of the series up to 13/08/2013. Therefore for stocks listed on the JSE from the start date of the dataset, the trading period was 2,655 days and less than this for stocks listed later.

It was decided that taking a formation period of five years was a sufficiently long period to gather the most accurate dependency structure that will be little influenced by short term deviations.

Similarly, the choice of trading period was considered to be long enough to be able to make meaningful conclusions from the results.

2.6 Updating of Parameters

Given the length of the chosen trading period, it was considered necessary to update the formation period parameters every quarter (assumed to be after every 62 trading days) so that trading signals were developed from parameters that were reliable and not out of date.

The formation period data was augmented by each quarter's data as each quarter came to a close. The update was carried out as follows for each of the trading strategies:

- Copula approach: A new copula parameter(s) is(are) fitted, optimal copulas determined, then the optimal copula parameter is used together with the trading data cumulative marginals to determine the marginal conditional distribution. See section 2.2 above for a detailed version of the method.

- Distance method: A new standard deviation is estimated from the augmented formation period data and used in determining trading signals in the ensuing quarter. Refer to section 2.3 above for more details on the method.

- Cointegration method: Similar to the distance method above, a new standard deviation was estimated from the augmented formation period data and used to determine the trading signals in the ensuing quarter. See section 2.4 above for more details.

2.7 Calculation of Excess Returns

The payoffs are calculated each time a position is opened and closed during the trading period. Given the length of the trading period, positions were often opened more than once during the trading period leading to a series of randomly distributed payoffs during the trading period (Gatev *et al.*, 2006).

In practice, traders have to calculate a hedge ratio to determine the amount of each stock to hold during trading period (Ferreira, 2008). When a position is entered into at the outset, the intention is to hold stocks with equal exposure such that the portfolio has a net zero market value. Therefore, by way of simplification, for the purpose of this dissertation it was assumed that a capital of ZAR 1,000,000 was invested in each stock when a position is opened during the trading period (Gatev *et al.*, 2006). The return in each trade was calculated as accumulated return of the long and the short positions from the start to the close of each trade. See equation 2.3 below:

$$R_t = \frac{P_L(t) - P_S(t)}{Capital} \tag{2.3}$$

where
$R_t =$ *Return of each trade*
$P_L(t) =$ *Value of the long position at the close of the trade*
$P_S(t) =$ *Value of the short position at the close of each trade*
$Capital =$ *The gross exposure taken in each leg of the trade*

In addition, stop loss and stop profit rules were set up whereby if accumulated return of the net position fell by more than 10%, the trade would be closed even if stocks had not converged at that stage. Similarly, if a particular position made more than 15%, the position would be closed. These rules are meant to preserve the gains and prevent loss of said gains should stocks continue diverging.

Furthermore, an additional rule was employed whereby no single trade was allowed to remain open for more than one hundred (100) days. A trade was closed after 100 days whether or not there was a loss or the trading signals had indicated convergence.

Trading costs, short-selling costs,the notion of bid offer spread and effect of dividends were ignored throughout this dissertation. The main focus was on comparing the performance of the methods.

2.8 Robustness checks

The results and conclusions based on the central assumptions discussed in section 3 below are only valid under those central assumptions. Limitations of these assumptions are discussed in 4. There are many variations of the assumptions that can be considered to test the validity of the conclusions. However, for the purpose of this dissertation, only the bounds for the copula method were adjusted for each pair so as to represent, approximately, the conditional probability's 2.5th and 97.5th percentiles during the initial formation period. This represents a 95% confidence interval that roughly reflects the two standard deviation band considered in the distance and cointegration method. The function "*Percentile*" in Excel was used to obtain the required boundaries.

Comparison was then made between the results of the copula under the new bounds and the central results of the distance and copula methods. See section 3.4 below for details.

Other variations of assumptions such as different formation period and frequency of updating parameters could provide an avenue of further research.

Chapter 3

Empirical Results

3.1 Pairs Selected

This section gives a brief description of the selected pairs based on the cointegration analysis and the approach described in section 2.1.

1. FSR-RMH: FirstRand Ltd (FSR) and RMB Holdings Ltd (RMH). Data were available for both stocks from 06/01/1998.

2. INL-INP: Investec Ltd (INL) and Investec PLC (INP). This is a dual listing and it is understandable why the two stocks would be highly cointegrated. Data for INL were available from 06/01/1998 but INP only became listed on 22/07/2002. The 1,245 initial formation days were thus taken from 22/07/2002 to 12/07/2007 and there was a the total of 1,522 trading days for the pair up to 13/08/2013.

   Pairs 1 and 2 both fall under the Financials Sector, Banking Industry and 'Banks' Sub-Industry according to Bloomberg (2013).

3. GFI-HAR: Gold Fields Ltd (GFI) and Harmony Gold Mining Co Ltd (HAR) are both classified under the Materials Sector, Metals & Mining Industry and Precious Metal Mining Sub-Industry according to Bloomberg (2013). Data were available for both stocks from 06/01/1998.

4. REM-SHP: Remgro Ltd (REM) and Shoprite Holdings Ltd (SHP) are both listed under the Consumer Staples Sector. REM is classified under Consumer Products Industry and the Food Manufacturing Sub-Industry. SHP falls under the Retail Staples Industry and Food Retailers Sub-Industry (Bloomberg, 2013). REM only became listed on 26/09/2000 whereas RMH data was available from the start of the data series. Consequently, there was a total of 1,976

trading days from 19/09/2005. These companies operate in the same industry and their fortunes likely to be influenced by similar factors.

5. MMI-RMH: MMI Holdings Ltd/South Africa (MMI) and RMB Holdings Ltd (RMH). Per Bloomberg (2013) MMI falls under the Financials Sector, Insurance Industry and Life Insurance Sub-Industry. Classification of RMH was given under 1 above. MMI only became listed on 21/09/2001 whereas SHP data was available from the start of the data series. Consequently, there was a total of 1,728 trading days from 15/09/2006.

3.2 Trading Results - FSR-RMH Pair

This section will give a detailed analysis of the empirical results of pair 1. The remainder of the results will be summarised in section 3.3 below. It should be borne in mind that the returns referred to in this dissertation are calculated according to the methodology described in section 2.7 above.

Figure 3.1 below demonstrates the close relationship between the FSR and RMH stocks. Only the last 500 days of the trading period are shown for reasons of space.

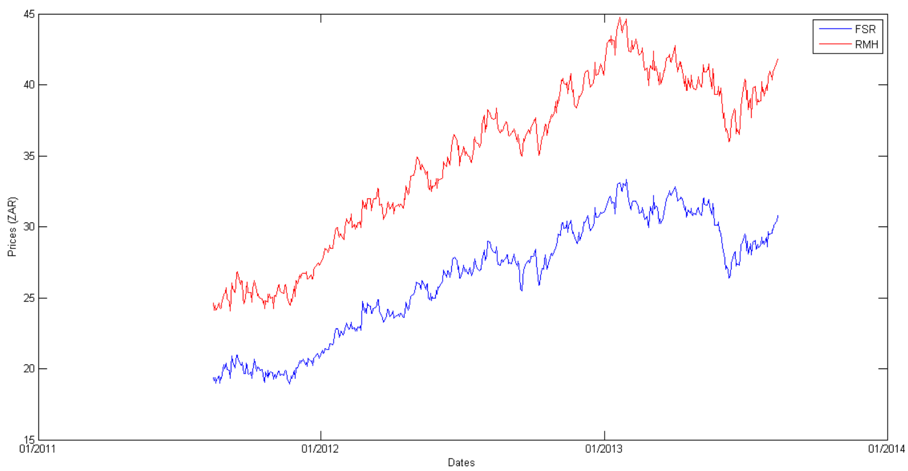

Fig. 3.1: FSR-RMH: Closing daily prices for the last 500 trading days up to 13/08/2013.

For the whole period of investigation, the Engle-Granger cointegration test rejected the null hypothesis with a very low p-value 0.001, a test statistic of -9.5269 and the correlation coefficient of 99.54%. During the formation period the p-value was also very low at 0.001, a test statistic of -5.8710 and the correlation coefficient was 91.43%. For the trading period, the p-value was 0.001, the test statistic was -7.5436, and the correlation coefficient was 99.51%. In a nutshell, the FSR-RMH pair was highly cointegrated and highly correlated.

3.2.1 Copula Approach

Figure 3.2 below illustrates the dependency structure of the pair during the initial five-year formation period. The scatter plot of the cumulative marginals u and v of FSR and RMH respectively is shown in the top left corner. The other five copula scatter plots were obtained by plotting the output from the Matlab function "*copularnd*" which takes in the following inputs: the copula parameter(s) and the u and v vectors. It is clear from the figure below that Clayton was not a good fit. In fact, Clayton was the least optimal copula as shown in table 3.1 below. Clayton has lower tail dependency whereas the data appear to have symmetric tail dependency. The symmetric tail dependency is fully captured by the Student-t copula shown in the bottom right corner. See Appendix A.3 for more on copula tail dependency.

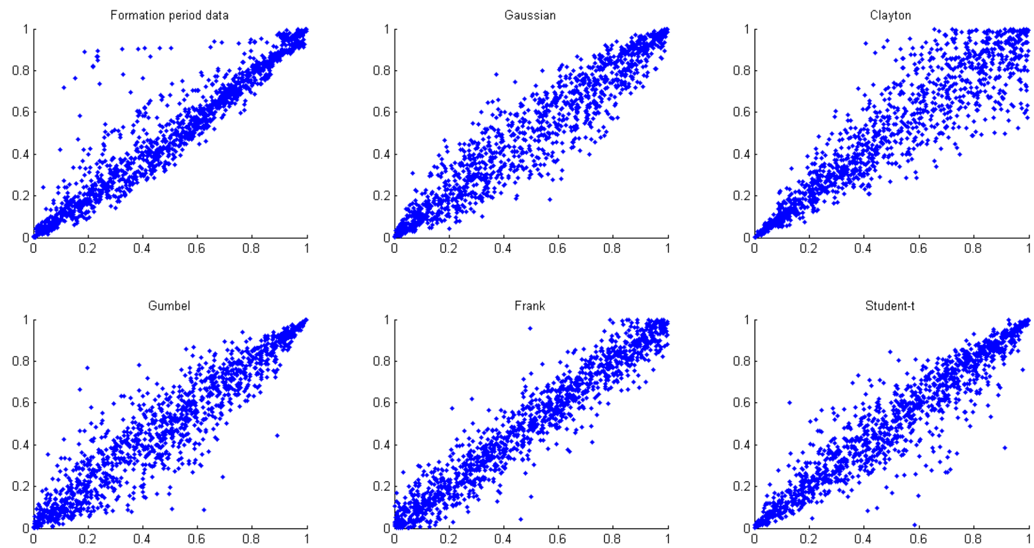

Fig. 3.2: FSR-RMH: Formation period data - u and v plots and fitted copulas.

Copula	Negative Log-likelihood
Student-t	-1608.06
Frank	-1469.99
Gumbel	-1440.87
Gaussian	-1418.57
Clayton	-1330.57

Tab. 3.1: FSR-RMH: Negative log-likelihood values using initial formation period data.

The results show that a total of 104 trades were identified. Of these, 95 of the trades resulted in positive returns. Table 3.2 below gives a summary of the trades opened for this pair.

Trades	Buy/Sell FSR	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	17/01/2003	6	0.1%
2	1	-1	29/01/2003	5	0.2%
3	-1	1	20/02/2003	5	2.3%
4	-1	1	05/03/2003	5	0.6%
5	-1	1	02/04/2003	12	2.8%
6	1	-1	15/04/2003	1	3.2%
7	-1	1	23/04/2003	3	2.6%
8	1	-1	17/06/2003	18	2.9%
9	-1	1	04/07/2003	2	2.1%
10	1	-1	04/12/2003	35	1.0%
11	1	-1	23/12/2003	3	6.6%
12	1	-1	13/01/2004	1	0.04%
13	1	-1	06/02/2004	3	-0.5%
14	1	-1	05/03/2004	14	0.7%
15	-1	1	12/03/2004	3	1.1%
16	-1	1	25/03/2004	3	0.7%
17	1	-1	23/06/2004	13	2.0%
18	1	-1	16/07/2004	15	1.1%
19	1	-1	12/08/2004	11	3.9%
20	1	-1	23/09/2004	22	-0.3%
21	-1	1	29/10/2004	1	1.6%
22	1	-1	29/11/2004	2	-0.6%
23	-1	1	02/12/2004	1	2.8%
24	1	-1	10/12/2004	4	1.3%
25	1	-1	04/01/2005	10	2.6%
26	-1	1	02/02/2005	3	3.8%
27	-1	1	07/03/2005	20	5.8%
28	1	-1	23/03/2005	6	0.7%
29	1	-1	20/05/2005	8	2.9%
30	1	-1	31/05/2005	5	3.1%
31	1	-1	08/06/2005	2	1.2%
32	-1	1	14/06/2005	1	2.4%
33	-1	1	24/06/2005	4	2.4%

Trades	Buy/Sell FSR	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
34	-1	1	11/07/2005	1	4.5%
35	-1	1	03/08/2005	6	0.2%
36	1	-1	15/09/2005	1	2.3%
37	1	-1	19/09/2005	1	2.6%
38	-1	1	20/10/2005	13	3.6%
39	-1	1	30/12/2005	40	0.5%
40	1	-1	10/01/2006	1	1.1%
41	-1	1	18/01/2006	5	-1.0%
42	1	-1	27/01/2006	4	1.7%
43	-1	1	27/01/2006	4	1.7%
44	1	-1	06/03/2006	9	0.5%
45	-1	1	31/05/2006	48	0.4%
46	-1	1	15/06/2006	4	2.2%
47	1	-1	27/11/2006	13	3.6%
48	1	-1	08/01/2007	25	6.2%
49	-1	1	19/01/2007	7	3.7%
50	-1	1	25/01/2007	1	0.9 %
51	-1	1	05/02/2007	2	1.9%
52	-1	1	08/03/2007	21	1.7%
53	-1	1	23/04/2007	18	15.3%
54	-1	1	07/05/2007	5	5.5%
55	1	-1	28/05/2007	1	2.0%
56	-1	1	31/05/2007	1	1.8%
57	1	-1	07/06/2007	4	3.0%
58	1	-1	19/06/2007	1	2.2%
59	-1	1	26/07/2007	6	4.2%
60	-1	1	13/08/2007	1	2.6%
61	-1	1	15/08/2007	1	2.2%
62	1	-1	12/09/2007	6	2.4%
63	-1	1	26/09/2007	5	3.7%
64	1	-1	15/10/2007	6	2.3%
65	-1	1	07/11/2007	10	6.3%
66	1	-1	16/11/2007	5	4.4%

Trades	Buy/Sell FSR	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
67	-1	1	30/11/2007	7	5.2%
68	-1	1	07/12/2007	1	4.6%
69	-1	1	30/01/2008	30	7.7%
70	-1	1	19/03/2008	22	8.0%
71	-1	1	03/10/2008	50	4.4%
72	-1	1	04/12/2008	2	7.9%
73	-1	1	06/02/2009	31	6.4%
74	1	-1	10/11/2009	82	5.9%
75	1	-1	14/09/2010	100	-5.0%
76	1	-1	07/03/2011	8	-13.5%
77	-1	1	23/09/2011	11	0.4%
78	-1	1	29/09/2011	1	2.2%
79	1	-1	06/10/2011	3	0.9%
80	-1	1	11/10/2011	1	1.4%
81	-1	1	25/10/2011	3	1.5%
82	-1	1	12/12/2011	25	1.9%
83	-1	1	21/12/2011	2	2.3%
84	1	-1	10/02/2012	32	4.2%
85	-1	1	23/02/2012	1	-0.2%
86	-1	1	19/03/2012	16	-1.1%
87	-1	1	28/05/2012	26	2.7%
88	-1	1	13/06/2012	6	1.2%
89	-1	1	29/06/2012	9	1.8%
90	-1	1	17/07/2012	4	1.7%
91	1	-1	17/07/2012	6	-0.4%
92	-1	1	13/08/2012	5	0.6%
93	1	-1	28/08/2012	9	0.7%
94	-1	1	10/09/2012	1	1.0%
95	1	-1	28/09/2012	8	3.4%
96	1	-1	17/10/2012	7	1.3%
97	-1	1	27/11/2012	27	7.1%
98	1	-1	11/12/2012	2	0.7%
99	-1	1	10/01/2013	2	0.7%

Trades	Buy/Sell FSR	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
100	1	-1	17/01/2013	1	0.6%
101	-1	1	04/02/2013	4	1.9%
102	-1	1	14/03/2013	6	1.5%
103	-1	1	31/05/2013	46	3.2%
104	1	-1	13/08/2013	29	0.7%

Tab. 3.2: FSR-RMH: Summary of trade opportunities identified using the copula.

3.2.2 Distance Method

Figure 3.3 below illustrates the progression of the pre-specified standard deviation and the spread during the trading period. The pre-specified standard deviation is observed to experience a significant increase between mid-2006 and mid-2008. The spread does not appear to be symmetrically distributed about zero during this period. With the exception of a period in 2003 and a few occasions in 2007 and 2008, it seems consistently below zero throughout the trade period.

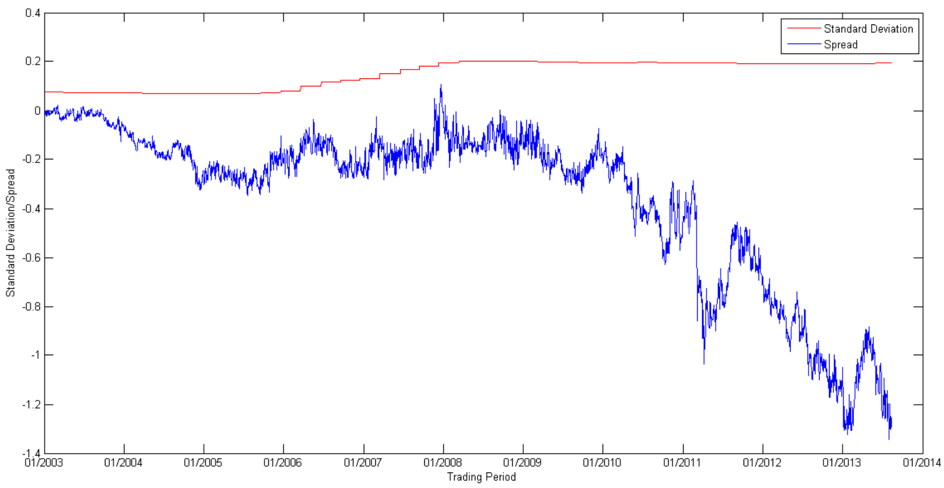

Fig. 3.3: FSR-RMH: Spread and pre-specified standard deviation during trading period using the distance method.

The Augmented Dickey-Fuller (ADF) test for unit root was conducted on the spread during the trading period. The ADF test was conducted in Matlab using the function "*adftest*". The "*adtest*" failed to reject the unit root null hypothesis which means that it is unlikely that the spread is mean reverting. This confirms the observed behaviour of the spread in figure 3.3 above.

The two trading opportunities summarised in table 3.3 below were both closed as a result of reaching the 100-day trade limit. This is explained by the widening of the spread shown in figure 3.3 above. Had the first trade not been closed after 100 days, figure 3.3 shows that the trade would have only been closed in early 2006 when the spread would have converged to zero. Similarly, for the second trade the spread is observed to continue widening until the end of the trading period towards the end of 2013.

Trades	Buy/Sell FSR	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	19/07/2004	100	0.7%
2	-1	1	06/09/2010	100	5.2%

Tab. 3.3: FSR-RMH: Summary of trade opportunities identified using the distance method.

3.2.3 Cointegration Method

Figure 3.4 below illustrates the progression of the pre-specified standard deviation and the spread during the trading period. The pre-specified standard deviation is observed to remain fairly level during the trading period. Figure 3.4 shows the spread to be alternatively biased below and above zero during different parts of the trading period. For example, between 2004 and 2006, the spread is biased below zero and it appears to be biased positively during 2007.

There were only two times when the spread breached the two standard deviation boundary (from the left). These were in February 2007 and April 2011, and these resulted in the trading opportunities summarised below in table 3.4.

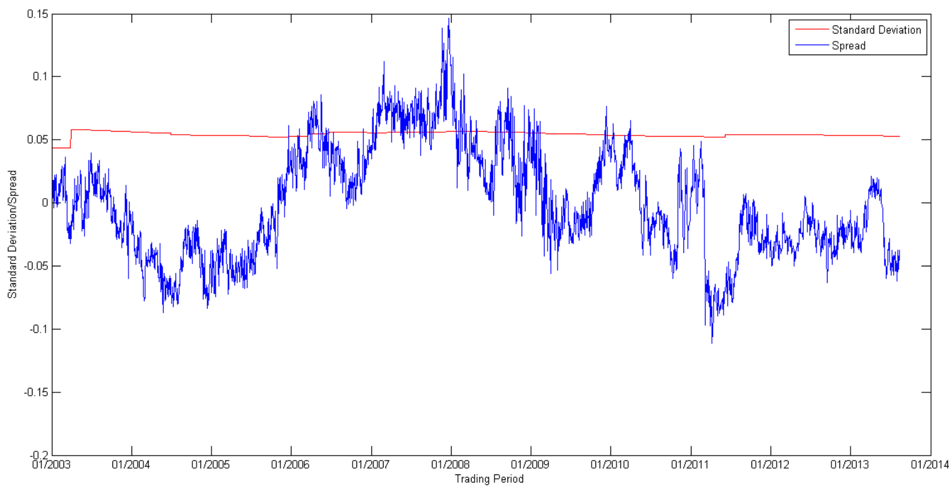

Fig. 3.4: FSR-RMH: Spread and pre-specified standard deviation during trading period using the cointegration method.

Liew and Wu (2013) note that when a pair of stocks has been verified to be cointegrated then the spread, as defined in section 2.4 above, should, in theory, be a stationary time series with a random distribution about a long-term mean of zero.

The Kwiatkwski, Phillips, Schmidt and Shin test (KPSS) test for stationarity was conducted on the spread during the trading period. The KPSS test assesses whether a univariate time series is stationary (*KPSS test for stationarity*, 2013). The Matlab function "*kpsstest*"' was performed on the spread during the trading period. The test indicated that the spread was not stationary.

Trades	Buy/Sell FSR	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	25/07/2007	100	2.9%
2	-1	1	10/08/2011	82	-10.1%

Tab. 3.4: FSR-RMH: Summary of trade opportunities identified using cointegration method.

3.2.4 Summary of FSR-RMH Results

Table 3.5 below shows a summary of the comparison of the trading results discussed in sections 3.2.1 to 3.2.3 above. It is observed that the cointegration method was the worst performer as the average return for the method was negative. The average return of the distance method was higher than that of the copula method. However, if one considers that the total number of trading opportunities for the copula method (104) and distance method (2), then the total return from the copula method during the trading period is significantly higher than that obtained from the distance method.

Regarded in this way, the copula method was more profitable than both the distance and the cointegration methods.

It can be argued that the performance of the distance and cointegration method is limited by the assumptions that the spread has a symmetric distribution about the mean, and/or that it is stationary and exhibits mean reverting behaviour. This was discussed in sections 3.2.2 and 3.2.3 above.

It was also observed that the length of trades in the copula method was shorter than for both the distance and the cointegration methods. This would be good news to investors looking for short holding periods.

These results are in line with the hypothesis of this dissertation that the copula method is more robust in accurately identifying trading signals with a potential for higher profits. See table 3.5 below for a summary of the comparison of the three methods.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	11.1	2.3%	100.0	3.0%	91.0	-3.6%
Std	15.8	2.9%	0.0	3.2%	12.7	9.2%
Min	1.0	-13.5%	100.0	0.7%	82.0	-10.1%
Max	100.0	15.3%	100.0	5.2%	100.0	2.9%

Tab. 3.5: FSR-RMH: Summary of trade results for all methods.

3.3 Overview of Results of Pairs 2 to 5

This section gives an overview of trading results of pairs 2 to 5. The trading results for these pairs are discussed below.

3.3.1 INL-INP

- Copula Approach:
  The copula approach resulted in a total of 46 trading opportunities. Table 3.6 below gives a summary of the trading results using the copula approach.

Trades	Buy/Sell INL	Buy/Sell INP	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	05/09/2007	33	7.1%
2	-1	1	05/11/2007	39	5.4%
3	1	-1	20/12/2007	27	15.5%
4	1	-1	01/02/2008	3	1.7%
5	1	-1	08/02/2008	3	3.1%
6	1	-1	27/02/2008	9	1.2%
7	-1	1	04/04/2008	23	6.9%
8	1	-1	19/05/2008	26	2.6%
9	1	-1	29/07/2008	48	10.5%
10	-1	1	05/08/2008	1	3.0%
11	1	-1	15/08/2008	6	2.4%
12	1	-1	30/09/2008	19	18.7%
13	1	-1	27/11/2008	14	9.8%
14	-1	1	12/12/2008	9	5.8%
15	1	-1	31/12/2008	8	8.4%
16	1	-1	06/01/2009	1	0.8%
17	-1	1	06/02/2009	19	15.1%
18	1	-1	03/04/2009	12	1.7%
19	1	-1	15/04/2009	12	1.3%
20	-1	1	04/05/2009	7	18.7%
21	1	-1	19/05/2009	3	0.6%
22	1	-1	08/06/2009	12	1.3%
23	1	-1	01/07/2009	9	12.8%
24	-1	1	13/08/2009	5	6.6%
25	-1	1	21/10/2009	40	4.2%
26	1	-1	10/11/2009	2	2.6%
27	-1	1	06/01/2010	33	1.3%
28	1	-1	18/01/2010	4	1.8%
29	1	-1	25/01/2010	1	1.4%
30	-1	1	19/02/2010	8	2.1%
31	-1	1	09/03/2010	7	1.5%
32	1	-1	17/06/2010	40	11.6%
33	-1	1	13/07/2010	9	1.8%

Trades	Buy/Sell INL	Buy/Sell INP	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
34	1	-1	12/08/2010	1	0.9%
35	1	-1	29/09/2010	17	1.3%
36	-1	1	18/10/2010	7	1.5%
37	-1	1	28/10/2010	6	1.6%
38	1	-1	05/11/2010	2	1.0%
39	1	-1	07/04/2009	100	-2.4%
40	-1	1	21/02/2012	35	2.6%
41	1	-1	12/06/2012	33	0.9%
42	1	-1	29/10/2012	35	2.6%
43	1	-1	15/03/2013	79	-0.5%
44	1	-1	30/04/2013	9	1.7%
45	1	-1	01/07/2013	5	3.2%
46	1	-1	23/07/2013	11	0.9%

Tab. 3.6: INL-INP: Summary of trade opportunities identified using the copula approach.

- Distance Method:
  This method did not lead to any trading opportunities. Figure 3.5 below explains this phenomenon. It was observed that the pre-specified standard deviation was higher than the spread during the trading period and that the spread did not deviate by more than two standard deviations during the trading period. The spread appears to be positively skewed in the first half of the trading period and negatively skewed in the second half of the trading period.

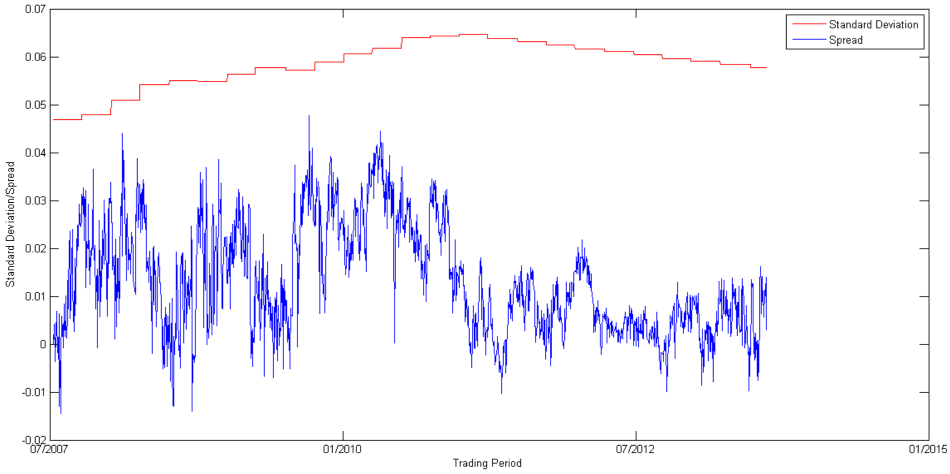

Fig. 3.5: INL-INP: Spread during trading period and pre-specified standard deviation using distance method.

- Cointegration Method:
  The cointegration method performed poorly and led to losses in all the trading positions. Figure 3.6 below shows that the spread was not symmetrically distributed around the mean of zero during the second half of the trading period as it was negatively skewed. This is evidenced by the lack of trading signals beyond October 2010. See table 3.7 below.

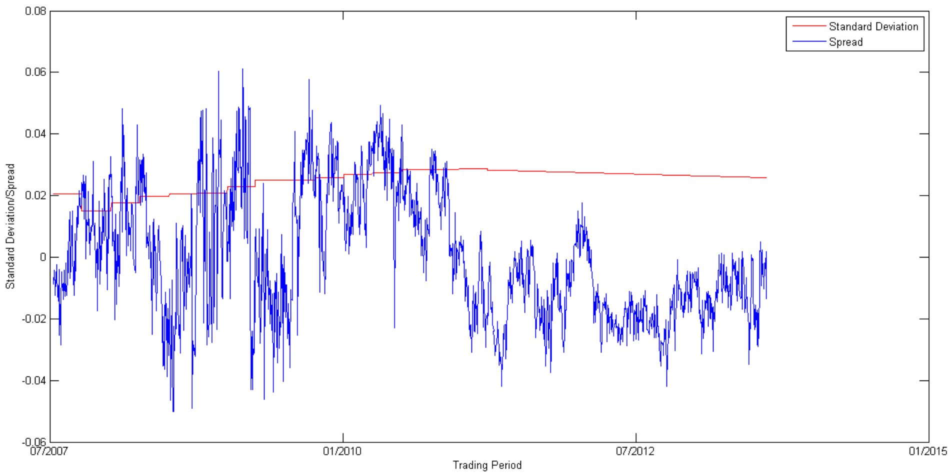

Fig. 3.6: INL-INP: Spread during trading period and pre-specified standard deviation using the cointegration method.

Trades	Buy/Sell INL	Buy/Sell INP	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	28/08/2007	14	-6.0%
2	1	-1	22/11/2007	4	-3.7%
3	1	-1	22/01/2008	8	-5.2%
4	1	-1	20/03/2008	23	-4.9%
5	1	-1	06/05/2008	22	-4.5%
6	-1	1	28/07/2008	9	-10.3%
7	-1	1	30/09/2008	5	-6.0%
8	1	-1	24/10/2008	7	-7.1%
9	1	-1	04/11/2008	1	-7.6%
10	1	-1	19/12/2008	5	-7.4%
11	1	-1	19/03/2009	29	-5.9%
12	1	-1	21/10/2009	21	-6.9%

Tab. 3.7: INL-INP: Summary of trade opportunities identified using cointegration method.

- Summary of INL-INP Results:
  Table 3.8 below gives a summary of the trading results for all the methods. The summary favours the copula method in all respects. The standard deviation of the returns was higher than in the cointegration method. However, this is irrelevant since all the cointegration method trades resulted in negative returns. It was interesting to observe that the cointegration method failed to pick up accurate trading opportunities during the period of Financial crisis whereas the copula identified quite a number of profitable trades as shown in table 3.6 above.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	12.4	3.3%	N/A	N/A	12.3	-6.3%
Std	15.9	3.1%	N/A	N/A	9.2	1.7%
Min	1.0	0.2%	N/A	N/A	1.0	-10.3%
Max	78.0	16.6%	N/A	N/A	29.0	-3.7%

Tab. 3.8: INL-INP: Summary of trade results for all methods.

3.3.2 GFI-HAR

- Copula Approach:
  The copula approach resulted in three trades. Details of the trades are presented in tables 3.9 and 3.10 below.

Trades	Buy/Sell GFI	Buy/Sell HAR	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	18/02/2003	20	1.4%
2	1	-1	07/04/2004	75	8.7%
3	-1	1	15/10/2004	4	-10.0%

Tab. 3.9: GFI-HAR: Summary of trade opportunities identified using the copula approach.

- Distance and Cointegration methods:
  No trading signals were indicated using these methods. In both cases the pre-specified standard deviation was very high and the spread during the trading period stayed well within pre-specified standard deviation.

- Summary of GFI-HAR results:
  Table 3.10 below gives a summary of the trading results of the method. The copula method did not perform exceptionally well since the negative return in the third trade nearly offsets all the gains made in the first two trades (average return nearly was nearly zero). The distance and cointegration methods did not produce any trading opportunities.

  The dismal performance of this pair prompted further investigation into the relationship between the stocks. It was found that whilst the pair was highly cointegrated during the whole period of investigation (Engle-Granger cointegration test rejected the null hypothesis with a p-value of 0.0198, a test statistic of -3.6814 and a correlation coefficient of 88.88%), it failed the cointegration test during the trading period with a p-value of 0.1010, test statistic of -3.0442 and the correlation coefficient was only 69.26%). During the formation period the p-value was 0.0134, the test statistic was -3.8173 and correlation coefficient of 90.46%. So one could attribute the reduced trading opportunities to the fact the stocks were not cointegrated and/or low level of correlation during the trading period i.e. lack of evidence showing strong linear association. It is possible that any non-linear association expected to be captured by the copula method was not strong enough to result in consistent optimal trades. Under the copula method, the first two trades were profitable whilst the last trade resulted in a loss.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	33.0	0.02%	N/A	N/A	N/A	N/A
Std	37.2	9.4%	N/A	N/A	N/A	N/A
Min	4.0	-10.0%	N/A	N/A	N/A	N/A
Max	75.0	8.7%	N/A	N/A	N/A	N/A

Tab. 3.10: GFI-HAR: Summary of trade results for all methods.

3.3.3 REM-SHP

- Copula Approach:
  The copula approach resulted in three trades. Details of the trades are presented in tables 3.11 and 3.14 below.

- Distance Method:
  This method resulted in two (2) trading opportunities. A summary is presented below in table 3.12.

- Cointegration Method:
  The cointegration method also resulted in two (2) trading opportunities for this pair. A summary of the trading results is presented in table 3.13 below.

- Summary of REM-SHP results:
  Table 3.14 shows a summary of the results for all methods for this pair. Even with a negative return in one of the trades, the copula method still has the highest mean return and the lowest standard deviation. The copula method also has the highest number of trading opportunities. This, combined with the higher average return would mean highest total overall return for the copula method. However, the summary of the length of trades indicates that an investor or trader would have held pair positions for longer than with the distance and cointegration methods.

Trades	Buy/Sell REM	Buy/Sell SHP	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	11/10/2005	5	1.2%
2	-1	1	10/11/2005	11	2.9%
3	-1	1	16/11/2005	1	3.6%
4	1	-1	23/11/2005	1	1.5%
5	1	-1	30/11/2005	3	0.9%
6	-1	1	19/12/2005	1	4.9%
7	1	-1	12/01/2006	13	5.9%
8	1	-1	31/01/2006	10	2.1%
9	1	-1	20/02/2006	6	7.2%
10	1	-1	15/03/2006	2	7.4%
11	1	-1	12/04/2006	12	-10.7%
12	-1	1	09/06/2006	4	20.6%
13	-1	1	21/09/2006	54	-11.2%
14	1	-1	21/11/2006	17	1.2%
15	1	-1	27/11/2006	2	7.3%
16	-1	1	02/02/2007	42	-4.0%
17	-1	1	09/02/2007	3	2.5%
18	-1	1	19/02/2007	1	3.7%
19	1	-1	02/05/2007	44	-13.7%
20	-1	1	04/07/2007	12	3.1%
21	1	-1	11/07/2007	1	2.9%
22	-1	1	27/07/2007	2	0.8%
23	1	-1	23/08/2007	2	0.8%
24	1	-1	14/09/2007	6	10.6%
25	-1	1	05/08/2013	20	-11.0%

Tab. 3.11: REM-SHP: Summary of trade opportunities identified using the copula approach.

3.3.4 MMI-RMH

- Copula Approach:
  Results are summarised under table 3.15 below.
- Distance Method:
  There were no trading opportunities under this method.

Trades	Buy/Sell REM	Buy/Sell SHP	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	02/06/2006	31	-10.4%
2	-1	1	09/11/2007	45	18.4%

Tab. 3.12: REM-SHP: Summary of trade opportunities identified using the distance method.

Trades	Buy/Sell REM	Buy/Sell SHP	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	24/11/2008	30	16.9%
2	1	-1	16/04/2013	22	-10.5%

Tab. 3.13: REM-SHP: Summary of trade opportunities identified using cointegration method.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	11.3	1.1%	38.0	4%	26.0	3.2%
Std	14.5	7.9%	9.9	20.4%	5.7	19.4%
Min	1.0	-13.7%	31.0	-10.4%	22.0	-10.5%
Max	54.0	20.6%	45.0	18.4%	30.0	16.9%

Tab. 3.14: REM-SHP: Summary of trade results for all methods.

- Cointegration Method:
  Results are summarised under table 3.16 below.

Trades	Buy/Sell MMI	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	22/09/2006	2	0.05%
2	1	-1	28/09/2006	1	3.1%
3	1	-1	03/11/2006	20	11.9%
4	-1	1	14/11/2006	1	1.7%
5	-1	1	22/11/2006	5	1.5%
6	1	-1	27/11/2006	1	1.9%
7	1	-1	04/12/2006	2	2.8%
8	-1	1	18/12/2006	7	3.5%
9	-1	1	08/01/2007	5	0.01%
10	1	-1	19/01/2007	3	0.9%
11	-1	1	30/01/2007	2	2.7%
12	1	-1	08/02/2007	1	4.6%
13	1	-1	20/02/2007	1	4.6%
14	-1	1	23/02/2007	1	3.6%
15	1	-1	28/02/2007	1	8.9%
16	1	-1	26/04/2007	28	16.1%
17	1	-1	18/09/2007	49	3.0%
18	-1	1	09/07/2007	48	16.3%
19	-1	1	14/05/2013	44	-10.4%

Tab. 3.15: MMI-RMH: Summary of trade opportunities identified using the copula method.

Trades	Buy/Sell MMI	Buy/Sell RMH	Date of End of Trade	Length of Trade(Days)	Total return of each Trade
1	-1	1	11/11/2008	3	-10.1%
2	-1	1	14/06/2013	12	16.1%

Tab. 3.16: MMI-RMH: Summary of trade opportunities identified using the cointegration method.

- Summary of MMI-RMH results:
  Table 3.17 shows a summary of the trade results for this pair. There were no trading signals using the distance method. A loss on trade number 19 was noted for the copula method but overall this method achieved the highest

average return and lowest standard deviation of returns. The high average return and higher number of trades executed would mean that overall the copula method resulted in higher total return during the trading period.

The average length of trading period was higher than the cointegration method. Therefore, an investor or a trader would have held pair positions for longer than with the cointegration method.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	11.7	4.0%	N/A	N/A	7.5	3.0%
Std	17.3	6.0%	N/A	N/A	6.4	18.5%
Min	1.0	-10.4%	N/A	N/A	3.0	-10.1%
Max	49.0	16.3%	N/A	N/A	12.0	16.1%

Tab. 3.17: MMI-RMH: Summary of trade results for all methods.

3.4 Results: Robustness Checks

Tables 3.19 to 3.23 below show how the copula results change as a result of adjusting in the trigger boundaries described in section 2.8 above. Instead of 0.05 and 0.95 lower and upper bounds respectively, table 3.18 shows the bounds applied under this check:

Pair	Lower Bound	Upper Bound
FSR-RMH	0.030	0.974
INL-INP	0.014	0.987
GFI-HAR	0.019	0.938
REM-SHP	0.018	0.964
MMI-RMH	0.036	0.964

Tab. 3.18: Adjusted lower and upper bounds for the copula method.

3.4.1 Pair 1: FSR-RMH

The total number trading opportunities decreased from 104 to 93 under the new bounds for the copula method. The average return remained unchanged at 2.3%. Table 3.19 below summarises the results of copula method and shows the comparison to the distance and cointegration methods under the central assumptions.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	10.5	2.3%	100.0	3.0%	91.0	-3.6%
Std	15.7	2.6%	0.0	3.2%	12.7	9.2%
Min	1.0	-11.7%	100.0	0.7%	82.0	-10.1%
Max	100.0	8.6%	100.0	5.2%	100.0	2.9%

Tab. 3.19: FSR-RMH: Summary of trade results for all methods under new copula boundaries.

3.4.2 Pair 2: INL-INP

The total number trading opportunities decreased from 46 to 43 under the new bounds for the copula method. The average return increased from 3.3% to 4.0%. Table 3.20 below summarises the results of copula method and shows the comparison to the distance and cointegration methods under the central assumptions.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	14.60	4.0%	N/A	N/A	12.3	-6.3%
Std	19.63	4.2%	N/A	N/A	9.2	1.7%
Min	1.0	-4.6%	N/A	N/A	1.0	-10.3%
Max	100.0	18.7%	N/A	N/A	29.0	-3.7%

Tab. 3.20: INL-INP: Summary of trade results for all methods under new copula boundaries.

3.4.3 Pair 3: GFI-HAR

The total number trading opportunities remained unchanged from the central assumptions. The average return increased significantly from 0.02% to 2.3%. Table 3.21 below summarises the results of copula method and shows the comparison to the distance and cointegration methods under the central assumptions.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	22.7	2.3%	N/A	N/A	N/A	N/A
Std	11.9	12.4%	N/A	N/A	N/A	N/A
Min	13.0	-11.1%	N/A	N/A	N/A	N/A
Max	36.0	13.3%	N/A	N/A	N/A	N/A

Tab. 3.21: GFI-HAR: Summary of trade results for all under new copula boundaries.

3.4.4 Pair 4: REM-SHP

The total number trading opportunities decreased from 25 to 21 under the new bounds for the copula method. The average return increased from 1.1% to 1.7%. Table 3.22 below summarises the results of copula method and shows the comparison to the distance and cointegration methods under the central assumptions.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	10.3	1.7%	38.0	4.0%	26.0	3.2%
Std	13.8	6.0%	9.9	20.4%	5.7	19.4%
Min	1.0	-11.2%	31.0	-10.4%	22.0	-10.5%
Max	54.0	10.6%	45.0	18.4%	30.0	16.9%

Tab. 3.22: REM-SHP: Summary of trade results for all methods.

3.4.5 Pair 5: MMI-RMH

The total number trading opportunities decreased from 19 to 18 under the new bounds for the copula method. The average return decreased slightly to 3.9% from 4.0%. Table 3.23 below summarises the results of copula method and shows the comparison to the distance and cointegration methods under the central assumptions.

	Copula		Distance		Cointegration	
	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns	Length of Trade(Days)	Returns
Mean	12.2	3.9%	N/A	N/A	7.5	3.0%
Std	17.6	6.1%	N/A	N/A	6.4	18.5%
Min	1.0	-10.4%	N/A	N/A	3.0	-10.1%
Max	49.0	16.3%	N/A	N/A	12.0	16.1%

Tab. 3.23: MMI-RMH: Summary of trade results for all methods.

Chapter 4

Limitations of the Study

4.1 Costs

This study ignored the issue of all costs involved in executing the pairs trading strategy. Such costs include transaction costs, short-selling costs and the issue of a bid-offer spread. Stander (2011), who also tested the pair FSR-RMH using a copula approach, showed that it is possible for profits after costs to be depleted. However, Gatev *et al.* (2006) showed pairs trading can be profitable even with costs involved in the context of US equity market.

Mashele *et al.* (2013) were of the view that in the JSE context, pairs trading can be profitable provided that trading costs are lower than 20 basis points for the average skilled trader. Their study showed that for the highly skilled trader profits are possible even in a high trading cost environment.

4.2 Use of Closing Prices

The results of this dissertation are based on daily closing prices. Thus it is assumed that the returns can be realised by using prices that are no longer available unless these trades are executed at the very start of each following day before significant changes in prices have occurred. In addition, number of trading opportunities is likely to be higher with the use of real time data (Stander, 2011).

4.3 Assumptions

The results of this dissertation are limited to the assumptions used in the implementation of the trades and the methodology. The assumptions include: choice of initial formation period, choices of stop loss and profit limit, maximum trading day limits, pre-specified standard deviations and upper and lower bounds for the copula conditional probabilities. Different assumptions could lead to different results and hence different conclusions.

4.4 Dividends

The dissertation did not consider the effect of dividends on the performance of the pairs. When a stock pays a dividend, this typically leads to a decrease of the stock price (Stander, 2011). Dividends may affect the performance in one of two ways. Firstly, the dividend amounts for a pair (e.g. MMI-RMH) may be different even if they are paid at the same time or they may be paid out at different times, thus affecting only one of the stocks in the pair (e.g. REM in April and November, SHP in March and September).

The effect of amount and timing of dividends on the performance of pairs is proposed as a topic for further research.

Chapter 5

Conclusion

This dissertation explored an alternative approach to pairs trading by use of copulas. Pairs trading strategy under the copula, distance, and cointegration methods was implemented on five different pairs of stocks listed on the JSE. The results showed some evidence of the robustness of copulas in identifying trading signals compared to the distance and the cointegration methods. The copula method was carried out in two different conditions, the unadjusted version referred to the conventional lower and upper boundaries of 0.05 and 0.95. The adjusted version considered the actual distribution of the conditional probability during the initial formation period for each pair and identifying values that correspond to the 2.5th and 97.5th percentile. This latter version aimed to approximately cover a confidence interval similar to that covered by the two standard deviations in the conventional methods.

This study concludes that, based on the data used and assumptions made, the copula approach has the potential to capture the co-dependencies of the stocks more accurately than the distance or cointegration methods. The copula method gave the most trading opportunities as summarised in table 5.1 below. A higher number of trading opportunities is desired when the average return is sufficiently positive.

	Copula Approach (unadjusted)	Copula Approach (adjusted)	Distance Method	Cointegration Method
FSR-RMH	104	93	2	2
INL-INP	46	43	0	12
GFI-HAR	3	3	0	0
REM-SHP	25	21	2	2
MMI-RMH	19	18	0	2

Tab. 5.1: Summary of number of trades.

For pairs 2 and 5, the average return under the copula method was higher compared to the conventional methods. For pair 1, the average higher than under the cointegration method but lower than in the distance method. The average return for pair 4 was lower than both distance and cointegration methods. The average return was positive for pair 3 under the copula method but this could not be compared to the conventional methods as no trading opportunities were identified in that regard.

	Copula Approach (unadjusted)	Copula Approach (adjusted)	Distance Method	Cointegration Method
FSR-RMH	2.3%	2.3%	3.0%	-3.6%
INL-INP	3.3%	4.0%	N/A	-6.3%
GFI-HAR	0.02%	2.3%	N/A	N/A
REM-SHP	1.1%	1.7%	4.0%	3.2%
MMI-RMH	4.0%	3.9%	N/A	3.0%

Tab. 5.2: Summary of average return per trade.

Overall, the total return under the copula method exceeded that of the other methods by virtue of the combination positive average returns and high number of trades. This can be shown in table 5.3 below. The Rand amounts represent the total amount that would have accumulated by the end of each trading period (ignoring interest and inflation) for each pair, where it is assumed ZAR 1,000,000 was invested in each leg of the transaction. This values in this table were estimated by multiplying ZAR 1,000,000 and the results in table 5.1 by the results in table 5.2.

	Copula Approach (unadjusted)	Copula Approach (adjusted)	Distance Method	Cointegration Method
FSR-RMH	2,392,000	2,139,000	60,000	-72,000
INL-INP	1,518,000	1,720,000	N/A	-756,000
GFI-HAR	600	69,000	N/A	N/A
REM-SHP	275,000	357,000	80,000	64,000
MMI-RMH	760,000	702,000	N/A	60,000

Tab. 5.3: Overall return in Rand terms during the whole trading period.

Pair 3 total return was very low and this is attributed to the low level of correlation during the trading period as discussed in section3.3.2 above. It is therefore essential to ensure a pair is highly cointegration such as pairs 1, 2, 4 and 5 to achieve sufficient profitability.

There was no clear-cut effect of adjusting the boundaries for the copula method. In some cases, it led to an improvement of the total return (pairs 2, 3 and 4) in other cases it resulted in a reduced total return (pairs 1 and 5).

All the returns presented above do not take into account trading costs. The incorporation of trading costs with use of copula was not investigated in this dissertation and could be a topic for further research.

The poor performance of the distance and co-integration methods can potentially be attributed to asymmetric distribution of the spread, the absence of mean reverting behaviour and/or non-stationarity during trading period. For example, under the discussion of results for pair 1 in section 3.2, it was noted that the spread was not stationary under the cointegration method and for the distance method, the spread was not mean-reverting during the trading period. Under the discussion for pair 2 in section 3.3.1, the spread was observed to be asymmetrically distributed about the mean of zero. On the other hand, copula method did not require one to make any assumptions on the data. This resulted in the copula finding the accurate association between stocks and correctly identifying the optimal trading opportunities.

	Copula Approach (unadjusted)	Copula Approach (adjusted)	Distance Method	Cointegration Method
FSR-RMH	2.6%	2.9%	3.2%	9.2%
INL-INP	4.0%	3.1%	N/A	1.7%
GFI-HAR	12.4%	9.4%	N/A	N/A
REM-SHP	6.0%	7.9%	20.4%	19.4%
MMI-RMH	6.1%	6.0%	N/A	18.5%

Tab. 5.4: Summary of standard deviation of the return.

Table 5.5 below gives a summary of the average length of trade for each period. For each method the average length of trade varied and there was no straightforward relationship on how this affected the returns for the trades. The relationship between the length of trades and robustness of returns could also be considered as a topic for further research.

	Copula Approach (unadjusted)	Copula Approach (adjusted)	Distance Method	Cointegration Method
FSR-RMH	11.1	10.5	100.00	91.00
INL-INP	12.4	14.6	N/A	12.33
GFI-HAR	33.0	22.7	N/A	N/A
REM-SHP	11.3	10.3	38.00	26.00
MMI-RMH	11.7	12.2	N/A	7.50

Tab. 5.5: Summary of average length of trade per method.

Bibliography

Bloomberg (2013). Companies Listed in South Africa.
URL: *Available at http://www.bloomberg.com/markets/companies/country/south-africa/2/*

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*, The Wiley Finance Series, Wiley.

Do, B. and Faff, R. (2010). Does Simple Pairs Trading Still Work?, *Financial Analysts Journal* **66**(4): 1 – 13.

Ferreira, L. (2008). New Tools for Spread Trading, *Futures: News, Analysis & Strategies for Futures, Options & Derivatives Traders (Summit Business Media)* **37**(12): 38.

Gatev, E., Goetzmann, W. N. and Rouwenhorst, K. G. (2006). Pairs Trading: Performance of a Relative-value Arbitrage Rule, *Review of Financial Studies* **19**(3): 797–827.

Jarque-Bera Test (2013). *Technical report.*
URL: *Available at http://www.mathworks.com/help/stats/jbtest.html*

Kat, H. M. (2003). The Dangers of using Correlation to Measure Dependence, *Journal of Alternative Investments* **6**(2): 54 – 58.

KPSS test for stationarity (2013). *Technical report.*
URL: *Available at http://www.mathworks.com/help/econ/kpsstest.html*

Liew, R. Q. and Wu, Y. (2013). Pairs Trading: A Copula Approach, *Journal of Derivatives & Hedge Funds* **19**(1): 12–30.

Ling, H. (2006). Dependence Patterns across Financial Markets: a Mixed Copula Approach, *Applied Financial Economics* **16**(10): 717 – 729.

Mashele, H., Terblanche, S. and Venter, J. (2013). Pairs Trading on the Johannesburg Stock Exchange, *Investment Analysts Journal* **78**: 13–23.

Nelsen, R. B. (2006). *An Introduction to Copulas (Springer Series in Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Patton, A. (2008a). Copula-based Models for Financial Time Series, *OFRC working papers series*, Oxford Financial Research Centre.
URL: *Available at http://EconPapers.repec.org/RePEc:sbs:wpsefe:2008fe21*

Patton, A. (2008b). Copula Toolbox for Matlab, version 1.07, 5apr08.
URL: *Available at http://public.econ.duke.edu/ ap172/code.html*

Stander, Y. (2011). *Multivariate Copulas in Financial Market Risk with Particular Focus on Trading Strategies and Asset Allocation*, PhD thesis, University of Johannesburg.

Sweeting, P. (2011). *Financial Enterprise Risk Management*, Cambrige University Press.

Trivedi, P. and Zimmer, D. (2007). *Copula Modeling: An Introduction for Practitioners*, Foundations and trends and econometrics, Now Publishers.

Utkulu, U. (n.d.). How to Estimate Long-Run Relationships in Economics: An Overview of Recent Developments.

Vidyamurthy, G. (2011). *Pairs Trading: Quantitative Methods and Analysis*, Wiley Finance, Wiley.

Xie, W., Liew, Q. R., Wu, Y. and Zou (2014). Pairs Trading with Copulas.
URL: *Available at http://ssrn.com/abstract=2383185*

Xie, W. and Wu, Y. (2013). Copula-based Pairs Trading Strategy.
URL: *Available at http://ssrn.com/abstract=2209209*

Appendix A

Copula Theory

A.1 Definition

Sweeting (2011) provides an easy to understand introduction to copulas. A copula is essentially a joint cumulative distribution function whose inputs are also cumulative distribution functions as opposed to raw values. Hence it is not the shape of the marginal distributions but the order of the data that is relevant. This means that copulas are invariant under certain transformations such as the logarithm function which is commonly used in applying economic data (Ling, 2006). That is to say, the same copula linking a set of data to other data would be obtained as long as the order of former data does not change in relation to the latter. This is an attractive and useful property and can be applied in statistical and financial modelling (Sweeting, 2011, p. 195).

A.2 Sklar's Theorem

Consider two variables U and V with marginal cumulative distribution functions $F(u)$ and $F(v)$ respectively i.e. $F(u) = Pr(U <= u)$ and $F(v) = Pr(V <= v)$. These variables can also be defined by way of a joint cumulative distribution function $F(u, v) = Pr(U <= u|V <= v)$. According to Sklar, a copula $C(F(u), F(v))$ defines the link between the function $F(u, v)$ and the marginal cumulative distributions, where

$$F(u, v) = C(F(u), F(v)) \tag{A.1}$$

Sklar's theorem can be extended to more than two variables to generate multivariate copulas. Nelsen (2006) provides in-depth details on copula theory and Cherubini *et al.* (2004) describe how copulas can be used in Finance.

A.3 Tail Dependence

Liew and Wu (2013), Sweeting (2011), and Patton (2008b) describe the tail dependencies exhibited by copulas. Below is a summary of the extent of tail dependence measured by the copulas considered in this dissertation:

- Gumbel copula has upper tail dependence and no lower tail dependence.

- Frank copula has neither upper or lower tail dependence.
- Clayton copula only has lower tail dependence when its parameter is positive, else it has no upper or lower tail dependence.
- Student-t copula has symmetric tail dependence.
- Gaussian copula has zero tail dependence.

A.4 Closed Form Conditional Probability Formulas

The table below was drawn from Liew and Wu (2013).

Copula	Θ range	$Pr(U <= u \| V = v) = \frac{\partial C}{\partial v}$	$Pr(V <= v \| U = u) = \frac{\partial C}{\partial u}$
Student-t	$\theta_1 \in (0,\infty)$ $\theta_2 \in (-1,1)$	$t_{(\theta_1+1)}\left(\sqrt{\frac{\theta_1+1}{\theta_1+{t_{\theta_1}}^{-1}(v)^2}} * \frac{{t_{\theta_1}}^{-1}(u)-\theta_2 {t_{\theta_1}}^{-1}(v)}{\sqrt{1-{\theta_2}^2}}\right)$	$t_{(\theta_1+1)}\left(\sqrt{\frac{\theta_1+1}{\theta_1+{t_{\theta_1}}^{-1}(u)^2}} * \frac{{t_{\theta_1}}^{-1}(v)-\theta_2 {t_{\theta_1}}^{-1}(u)}{\sqrt{1-{\theta_2}^2}}\right)$
Gaussian	(-1,1)	$\Phi\left(\frac{\Phi^{-1}(u)-\theta\Phi^{-1}(v)}{\sqrt{(1-\theta^2)}}\right)$	$\Phi\left(\frac{\Phi^{-1}(v)-\theta\Phi^{-1}(u)}{\sqrt{(1-\theta^2)}}\right)$
Gumbel	[1,∞]	$C_\theta(u,v)*[(-\ln u)^\theta+(-\ln v)^\theta]^{\frac{1-\theta}{\theta}}*(-\ln v)^{\theta-1}*\frac{1}{v}$	$C_\theta(u,v)*[(-\ln u)^\theta+(-\ln v)^\theta]^{\frac{1-\theta}{\theta}}*(-\ln u)^{\theta-1}*\frac{1}{u}$
Frank	$(-\infty,\infty)\setminus\{0\}$	$\frac{e^{-\theta v}\left(e^{-\theta u}-1\right)}{\left(e^{-\theta}-1\right)+\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}$	$\frac{e^{-\theta u}\left(e^{-\theta v}-1\right)}{\left(e^{-\theta}-1\right)+\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}$
Clayton	$(-1,\infty)\setminus\{0\}$	$v^{-(\theta+1)}\left(u^-\theta+v^{-\theta}-1\right)^{-\frac{1}{\theta}-1}$	$u^{-(\theta+1)}\left(u^-\theta+v^{-\theta}-1\right)^{-\frac{1}{\theta}-1}$

Tab. A.1: Conditional probability formulae of copulas used in this dissertation