Multi-Dimensional Pairs Trading Using Copulas

Apr 19, 2016

EFM Classification Codes & Research Area: 370 – Portfolio Management and Asset Allocation

Chzee An Lau[a,b], Wenjun Xie[c], Yuan Wu[d]

[a] Corresponding author who will attend and present.

[b] Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798. Tel: +65 9351 1877. Email: CLAU008@e.ntu.edu.sg.

[c] Division of Banking and Finance, Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798. Tel: +65 8228 0370. Email: XIEW0008@e.ntu.edu.sg.

[d] Division of Banking and Finance, Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798. Tel: +65 9740 8898. Email: AYWU@ntu.edu.sg.

Abstract

Pairs trading is a well-established statistical arbitrage strategy that is most often conducted in 2 dimensions with two stocks as a pair using the distance method. However, the profitability of the distance method has proven to be on a decline in the recent years. Hence, the copula approach is proposed to eliminate the need of implying normal distribution of stock returns, which is inevitable in the distance method, as well as to capture the dependency structure precisely by providing an explicit joint distribution function. However, using a single stock as a benchmark for another could be misleading as only a portion of the full dependency information is being utilized. Therefore, this paper extends the application of copula to propose a multi-dimensional pairs trading framework that involves the trading of three or more stocks as a group to increase dependency information and measure relative pricing more comprehensively. With more dependency information being utilized, practitioners could potentially benefit from more trading opportunities, reliable trading signals and diversification effects. The overall results show that the multi-dimensional copula strategy is able to identify more trading opportunities and generate a higher return than those of the 2-dimensional copula and distance strategies.

Keywords: Pairs Trading; Multi-Dimensional; Copula; Dependency Structure; Correlation

Acknowledgement

We wish to acknowledge the funding for this project from Nanyang Technological University under the Undergraduate Research Experience on CAmpus (URECA) programme.

Multi-Dimensional Pairs Trading Using Copulas

1. Introduction

Traditionally, pairs trading is conducted in 2 dimensions with a pair of stocks that have demonstrated strong historical co-movements. It is generally defined as a statistical arbitrage strategy that capitalizes on the temporary relative mispricing between two stocks whose prices are expected to converge due to strong historical co-movements. To generate excess return, one would simultaneously long the relatively undervalued stock and short the relatively overvalued stock whenever temporary relative mispricing occurs. As opposite positions are taken for two stocks that move together, this strategy is said to be market risk-free. Then, the positions are reversed after the stock prices converged. Thus, we can infer that the profitability of a pairs trading strategy is dependent on two factors: the identification of high-quality pairs whose prices have a high tendency to converge and the modeling of dependency structure between the two stocks to measure the degree of relative mispricing. This research paper aims to introduce multi-dimensional pairs trading, which is a new concept that involves the trading of three or more stocks as a group. Furthermore, it is expected to be a generalization of the 2-dimensional pairs trading techniques that offers richer information about the dependency structure and utilizes the information to measure the degree of relative mispricing so that trading positions could be entered in a more informed and less risky manner.

Among all conventional methods of 2-dimensional pairs trading, the distance method stands out as the most popular method. This popularity is attributed to the use of simple linear correlation coefficient as a measure of dependence which makes implementation convenient. Researchers have extended the studies of this method to larger samples and documented significant excess returns (Andrada et al., 2005; Gatev et al., 2006; Perlin, 2009; Pizzutilo, 2013). However, Do and Faff (2010) found that the profitability of the distance method has been declining over a long sample period. The mean excess return per month drops from 1.24% for the period of year 1962-1988 to 0.33% for the period of year 2003-2009. This significant drop in profitability was attributed to an increase in the proportion of pairs which diverge (Do and Faff, 2010). This risk of divergence is also known as arbitrage risk because the strategy is no longer market risk-free when pairs diverge, which is undesirable as the first factor that determines the profitability of a pairs trading strategy is convergence of stock prices (De Long et al., 1999).

Besides, the distance method implies that the financial data are normally distributed when capturing the dependency structures, ignoring non-linear associations (Liew and Wu, 2013). In reality, financial data are rarely normally distributed (Cont, 2001). Furthermore, negative skewness and excess kurtosis are being repeatedly observed in most financial assets (Kat, 2003; Crook and Moreira, 2011). Therefore, implying a symmetric distribution of spread around the mean of 0 may result in false trading opportunities that seem to be profitable initially but eventually turn out to be otherwise (Bock and Mestel, 2008).

In view of the aforementioned shortcomings of conventional methods of 2-dimensional pairs trading, the copula method is proposed. The copula method possesses two major advantages. Firstly, it separates the estimations of marginal distributions and joint distribution into two different procedures. This eliminates the need to imply normal marginal distributions of stock returns, which is inevitable in the conventional distance method. In fact, copula is a powerful and flexible tool that is able to capture the dependency structure accurately regardless of the forms of marginal distribution (Liew and Wu, 2013; Xie et al., 2014). This means that the copula method is a generalization of conventional pairs trading methods. For instance, the distance method is a special case of the copula method because the copula method is equivalent to the distance method when financial data are normally distributed. Nonetheless, the copula method remains applicable with any other forms of marginal distribution. The second advantage is that the copula connects the individual marginal distributions to the joint distribution of stock returns by providing an explicit function to describe dependence, which gives a more precise understanding of the dependency structure.

However, in all 2-dimensional pairs trading methods, only one stock is benchmarked against another and the relative mispricing is measured solely based on the common portion of their full dependency information while the uncommon portion, which also plays a role in the determination of the intrinsic value of a stock, is not being taken into account. Thus, there could be an improvement if the uncommon portion of their full dependency information is utilized in benchmarking. When an additional stock is being considered together with the existing pair of stocks, some of the aforementioned uncommon portion could be revealed as the uncommon portion of the existing pair of stocks may be dependent on the additional stock. This could provide new information about the behaviors of the pair of stocks. Through this means, multi-dimensional pairs trading is able to increase dependency information to be used for benchmarking and measure the degree of relative mispricing more reliably.

Furthermore, in 2-dimensional pairs trading, the relative mispricing is measured locally as the price movement of a stock could be due to its idiosyncratic risk. In this case, when it is being benchmarked against another stock of the same pair, it could be deemed as a potential arbitrage opportunity. On the other hand, the degree of relative mispricing is being measured more comprehensively in the case of multi-dimensional pairs trading, where a stock is being benchmarked against a group of stocks. In other words, a stock may be locally overvalued when benchmarked against another stock of the same pair but it is intrinsically fairly-valued when benchmarked against a group of stocks that have strong co-movements. Therefore, this kind of false trading signal could potentially be avoided in multi-dimensional pairs trading.

Last but not least, multi-dimensional pairs trading could reduce risk through diversification. Although pairs trading has always been claimed as a market risk-free strategy, it is not completely so because two or more stocks that are perfectly correlated do not exist. Despite pairs trading strategies only consider stocks that demonstrate strong co-movements or are highly and positively correlated, there is still room for diversification effects as the number of stocks increases, especially when the existing number of stocks is significantly low (which is two in 2-dimensional pairs trading strategies).

Since the 2-dimensional copula method generalizes the conventional methods of 2-dimensional pairs trading, the copula method should be applied to multi-dimensional pairs trading to ensure that there is no loss of generality. In fact, the multi-dimensional copula method is actually a generalization of pairs trading methods because the 2-dimensional copula method is just a special case of the multi-dimensional copula method.

Motivated by these inspiring ideas, the contributions of this paper to pairs trading are two-fold. Firstly, it proposes an innovative concept – multi-dimensional pairs trading, which is expected to be more profitable than the 2-dimensional pairs trading strategies. Secondly, as there are scant literatures about the use of copula in pairs trading, this paper promotes the use of copula by demonstrating and extending the application of copula to multi-dimensional pairs trading.

In the subsequent parts of this paper, we will discuss the framework of our methods and respective trading strategies under Section 2, followed by discussion and interpretation of results under Section 3. Lastly, we conclude the paper and provide directions for future studies under Section 4.

2. Methodology

The approaches we cover in this paper consist of two common stages: formation period and trading period. In the formation period, the historical data of the underlying stocks are studied to produce useful information about the behaviors of the stocks as a pair or as a group. For example, in the case of 2-dimensional copula method, the best-fitted marginal cumulative distributions are first estimated and then the optimal copula that best describes the dependency structure is determined. Subsequently, in the trading period, trading strategies are implemented based on the insights generated in the formation period to test for profitability.

2.1 Two-Dimensional Conventional Method

2.1.1 Framework

Out of several conventional methods, we adopt the most popular conventional method – the distance method. The implementation here is in accordance with Gatev et al. (2006) except for the pairs selection process. According to Gatev et al. (2006), the pairs are first selected based on the minimum distance criterion, which is equivalent to maximum correlation. However, we did not include a pairs selection process in this study and the stock candidates are pre-selected.

In our approach, the standardized price difference between the two stocks is known as spread and it is implied to be symmetrically and normally distributed around the long-term mean value of 0. Therefore, when the prices diverge and spread deviates from 0, one stock is overvalued relative to another and the latter is undervalued relative to the former. Hence, the greater the absolute value of spread, the higher the degree of relative mispricing. This price divergence is expected to be temporary as the spread has a long-run mean value of 0. Hence, the prices are expected to converge to restore the spread to its long-term mean value of 0. This fundamental mechanism is also known as mean-reverting behavior.

Thus, the historical prices of stocks are first normalized and the spread is scaled to 0 at the beginning of the formation period. Then, the difference between the normalized prices are calculated. In addition, the standard deviation of spread is calculated based on the stock prices in the formation period and it is used to measure the degree of relative mispricing in the trading period.

2.1.2 Trading Strategy

At the beginning of the trading period, the spread is rescaled to 0. When the spread deviates from 0 by more than two times the standard deviation of spread, equal long/short positions are simultaneously opened. The rationale is that, two standard deviation covers 95% of the distribution of spread. Hence, any value of spread above (below) two positive (negative) standard deviation indicates significant overvaluation (undervaluation). Then, the trading positions are closed when the spread reverts to the mean value of 0.

2.2 Two-Dimensional Copula

2.2.1 Framework

Copula separates the estimations of individual marginal distributions and joint distribution into two different procedures, which makes it a powerful tool as it eliminates the need to imply normal distribution of stock returns. In addition, copula links the individual marginal distributions to their joint distribution by providing an explicit function to describe their dependency structure. Thus, the focus here is to optimally model the joint distribution of stock returns and measure the degree of relative mispricing based on their joint distribution.

Sklar's theorem (1959) states that, if H is a 2-dimensional joint cumulative distribution function for random variables X_1 and X_2, with respective continuous marginal cumulative distributions $F_1(X_1)$ and $F_2(X_2)$, then there exists a 2-copula C such that

$$H(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big) = P(X_1 \leq x_1, X_2 \leq x_2)$$

Next, let $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$

$$C(u_1, u_2) = P(F_1(X_1) \leq u_1, F_2(X_2) \leq u_2) = H\big(F_1^{-1}(u_1), F_2^{-1}(u_2)\big)$$

Then, the conditional probabilities can be found by taking first derivative of the copula function.

$$P(U_1 \leq u_1 | U_2 = u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2} = P(X_1 \leq x_1 | X_2 = x_2)$$

$$P(U_2 \leq u_2 | U_1 = u_1) = \frac{\partial C(u_1, u_2)}{\partial u_1} = P(X_2 \leq x_2 | X_1 = x_1)$$

Assume that X and Y are a pair of selected stocks with joint distribution H and marginal cumulative distributions $F_X(R_t^X)$ and $F_Y(R_t^Y)$ respectively, where R_t^X and R_t^Y represent their respective returns for a given day t, then according to Sklar's theorem (1959), there exists a 2-copula function such that

$$H(r_t^X, r_t^Y) = C\left(F_x(r_t^X), F_Y(r_t^Y)\right) = P(R_t^X \leq r_t^X, R_t^Y \leq r_t^Y)$$

Subsequently, let $U_1 = F_X(R_t^X)$ and $U_2 = F_Y(R_t^Y)$

$$C(u_1, u_2) = P(F_X(R_t^X) \leq u_1, F_Y(R_t^Y) \leq u_2) = H\left(F_X^{-1}(u_1), F_Y^{-1}(u_2)\right)$$

Therefore, the first step is to estimate their best-fitted marginal cumulative distributions, R_t^X and R_t^Y, and their respective parameters. Then, different categories of copula are applied to capture their optimal joint distribution and the one that that best describes the dependency structure is selected. After that, first derivatives of the copula function are taken to calculate the conditional probabilities, which are then used to measure the degrees of relative mispricing, $M_t^{X|Y}$ and $M_t^{Y|X}$.

$$M_t^{X|Y} = \frac{\partial C(u_1, u_2)}{\partial u_2} = P(R_t^X \leq r_t^X \middle| R_t^Y = r_t^Y)$$

$$M_t^{Y|X} = \frac{\partial C(u_1, u_2)}{\partial u_1} = P(R_t^Y \leq r_t^Y \middle| R_t^X = r_t^X)$$

where $0 \leq M_t^{X|Y} \leq 1$ and $0 \leq M_t^{Y|X} \leq 1$.

Generally speaking, a conditional probability of 0.5 means that the underlying stock is fairly-valued relative to another stock that is being conditional on. On the other hand, when the conditional probability is higher than 0.5, the underlying stock is overvalued relative to another stock that is being conditional on. Conversely, when the conditional probability is lower than 0.5, the underlying stock is undervalued relative to another stock that is being conditional on. The higher (lower) the conditional probability, the higher the degree of relative overvaluation (undervaluation).

2.2.2 Trading Strategy

The conditional probabilities, $M_t^{X|Y}$ and $M_t^{Y|X}$, only measure the degrees of relative mispricing for a single day. Thus, it is necessary to sum up the degree of relative mispricing for each day to determine the overall degree of relative mispricing. Let MI_{2D}^X and MI_{2D}^Y be the overall mispricing indexes of stock X and stock Y respectively. Since a conditional probability of 0.5 means that the two underlying stocks are fairly-valued, then only $(M_t^{X|Y} - 0.5)$ and $(M_t^{Y|X} - 0.5)$ are added to MI_{2D}^X and MI_{2D}^Y respectively. At the beginning of the trading period, MI_{2D}^X and MI_{2D}^Y are scaled to 0. Assume that T_{2D} and SL_{2D} (T for trigger and SL for stop-loss) are the respective threshold levels of relative mispricing for opening trading positions, and closing trading positions to prevent substantial losses when stocks diverge. In addition, trading positions are closed when MI_{2D}^X returns to 0 if they are opened based on MI_{2D}^X or when MI_{2D}^Y returns to 0 if they are opened based on MI_{2D}^Y. Thus, there are four different scenarios of opening and closing trading positions, which are summarized in Table 1.

[Insert Table 1 here.]

Note that all long and short positions are taken in equal weightages to ensure that the trades are market risk-free. At the end of the trading period, all trading positions are closed regardless of the values of MI_{2D}^{X} and MI_{2D}^{Y}. It should be noted that T_{2D} and SL_{2D} are both pre-specified values. Hence, there are infinite combinations and back-testing should be performed to determine the optimal values of T_{2D} and SL_{2D}.

2.3 Multi-Dimensional Copula

2.3.1 Framework

For simplicity, we adopt 3-dimensional copula to represent multi-dimensional copula. Similar to the 2-dimensional copula method, the focus here is to optimally model the joint distribution of returns of stocks X, Y and Z and measure the degrees of relative mispricing based on their joint distribution. Therefore, we first apply Sklar's theorem (1959) in 3 dimensions.

$$H(r_t^X, r_t^Y, r_t^Z) = C\left(F_x(r_t^X), F_Y(r_t^Y), F_Z(r_t^Z)\right) = P(R_t^X \leq r_t^X, R_t^Y \leq r_t^Y, R_t^Z \leq r_t^Z)$$

Then, the marginal cumulative distributions of R_t^X, R_t^Y, and R_t^Z as well as their respective parameters are estimated. Subsequently, we apply the Bernstein copula (Sancetta and Satchell, 2004) to estimate their joint distribution. The Bernstein copula generalizes the families of polynomial copulas such that

$$C_B(u_1, \dots, u_k) = \sum_{v_1} \dots \sum_{v_k} \alpha\left(\frac{v_1}{m_1}, \dots, \frac{v_k}{m_k}\right) P_{v_1, m_j}(u_1) \dots P_{v_k, m_j}(u_k)$$

where $\alpha\left(\frac{v_1}{m_1}, \dots, \frac{v_k}{m_k}\right)$ is a real valued constant indexed by $(v_1, \dots, v_k)$,

$$v_j \in \mathbb{N}_+ \text{ such that } 0 \leq v_j \leq m_j,$$

$$P_{v_j, m_j}(u_j) \equiv \binom{m_j}{v_j} (u_j)^{v_j} (1 - u_j)^{m_j - v_j} \text{ and}$$

$$u_j = F_j(x_j).$$

It should be noted that practitioners are free to choose the values of v and m. In this paper, a value of 10 is chosen for all m's, and all v's start from 1 with a step increment of 1. Thus, in the case of three stocks, the Bernstein copula and the degrees of mispricing are defined as follows

$$C_B(u_1, u_2, u_3) = \sum_{v_1} \dots \sum_{v_3} \alpha\left(\frac{v_1}{10}, \dots, \frac{v_3}{10}\right) P_{v_1,10}(u_1) \dots P_{v_3,10}(u_3)$$

$$M_t^{X|Y,Z} = \frac{\partial C_B(u_1, u_2, u_3)}{\partial u_2\, \partial u_3} = P(R_t^X \le r_t^X \middle| R_t^Y = r_t^Y, R_t^Z = r_t^Z)$$

$$M_t^{Y|X,Z} = \frac{\partial C_B(u_1, u_2, u_3)}{\partial u_1\, \partial u_3} = P(R_t^Y \le r_t^Y \middle| R_t^X = r_t^X, R_t^Z = r_t^Z)$$

$$M_t^{Z|X,Y} = \frac{\partial C_B(u_1, u_2, u_3)}{\partial u_1\, \partial u_2} = P(R_t^Z \le r_t^Z \middle| R_t^X = r_t^X, R_t^Y = r_t^Y)$$

Similarly, a conditional probability of 0.5 means that the underlying stock is fairly-valued relative to the other two stocks that are being conditional on. On the other hand, when the conditional probability is higher than 0.5, the underlying stock is overvalued relative to the other two stocks that are being conditional on. Conversely, when the conditional probability is lower than 0.5, the underlying stock is undervalued relative to the other two stocks that are being conditional on. The higher (lower) the conditional probability, the higher the degree of relative overvaluation (undervaluation).

2.3.2 Trading Strategy

The overall mispricing indexes of stocks X, Y and Z are defined as MI_{MD}^X, MI_{MD}^Y and MI_{MD}^Z. They are determined by adding $(M_t^{X|Y,Z} - 0.5)$, $(M_t^{Y|X,Z} - 0.5)$ and $(M_t^{Z|X,Y} - 0.5)$ to MI_{MD}^X, MI_{MD}^Y and MI_{MD}^Z every day respectively. Note that MI_{MD}^X, MI_{MD}^Y and MI_{MD}^Z are scaled to 0 at the beginning of the trading period. Similar to the case of 2-dimensional copula, assume that

T_{MD} and SL_{MD} are the respective threshold levels of relative mispricing for opening trading positions, and closing trading positions to prevent substantial losses when stocks diverge. Furthermore, trading positions are closed when the corresponding mispricing index returns to 0. There are six different scenarios of opening and closing trading positions, which are summarized in Table 2.

[Insert Table 2 here.]

Note that in each scenario, all long and short positions are taken in equal weightages to ensure that the trades are market risk-free. At the end of the trading period, all trading positions are closed regardless of the values of MI_{MD}^{X}, MI_{MD}^{Y} and MI_{MD}^{Z}. It should be noted that T_{MD} and SL_{MD} are both pre-specified values. Hence, there are infinite combinations and back-testing should be performed to determine the optimal values of T_{MD} and SL_{MD}.

3. Empirical Results

3.1 Data

Note that we did not include a stocks selection process in this study as each method has its own measure of dependence. For demonstration purpose, the stock candidates in this study are Oversea-Chinese Banking Corporation (OCBC), United Overseas Bank (UOB) and Development Bank of Singapore (DBS). They are all from the banking sector in Singapore and are listed as O39, U11 and D05 respectively on the Singapore Exchange (SGX). Furthermore, they are very similar in terms of business operations. For instance, while they have their main operations in Singapore, they also have relatively smaller operations in many Asian countries.

The sample period is from 4 January 2010 to 31 December 2014. The daily stock prices are extracted from Yahoo Finance. The stock candidates are proven to be highly correlated throughout the sample period as shown in Figure 1 and they have an average correlation coefficient of 0.9332. The data from the sample period are further divided into 8 time periods, with each time period consisting of a formation period and a trading period. The formation period is defined as a 12-month period, which is equivalent to 252 days, and the trading period is defined as a 6-month period, which is equivalent to 126 days. Consequently, the number of days in a month is taken as 21 days. Note that there is no general guideline for how long each period should be fixed at.

[Insert Figure 1 here.]

3.2 Excess Returns

In this paper, we adopt the return on committed capital as our return measurement, which means that all money prepared for potential trades is considered in the principal amount even if the if the money has not been used to open a position (Gatev et al., 2006). Note that for the 2-dimensional methods, three possible pairs can be made out of the three stock candidates while there is only one possible group for the multi-dimensional method. Therefore, the excess returns of 2-dimensional methods are calculated by averaging the excess returns of the three possible pairs. Then, for all the aforementioned strategies, the monthly excess returns are calculated by compounding daily stock returns over the 21 days in that month. Subsequently, the 6-month trading period excess returns are calculated by averaging the six respective monthly excess returns.

3.3 Optimization

As mentioned earlier, practitioners are free to choose the threshold trigger (T) and stop-loss (SL) values. However, instead of choosing an arbitrary value, we ran through an optimization with 100 combinations of T and SL for every method to locate the optimal values. The optimization results for the distance, 2-dimensional copula and multi-dimensional copula methods are shown in Table 3, Table 4 and Table 5 respectively. The returns are the averages of 8-time-period excess returns. The average return that corresponds to the optimal values of T and SL is highlighted in bold in each table. Note that for all three methods, the average returns that correspond to the selected optimal values may not necessary be the highest. This is to increase the representation of the optimal values and to avoid optimal values that give higher returns but are either too high or too low and may only be optimal to this selected sample. After all, the main purpose of these optimizations is to locate a range in which the optimal values of T and SL lie.

[Insert Table 3 here.]

[Insert Table 4 here.]

[Insert Table 5 here.]

3.4 Main Results

Table 6 shows the 8-time-period average excess returns of the distance, 2-dimensional copula and multi-dimensional copula strategies. The excess returns of all 3 strategies are positive but only the excess returns of the distance and multi-dimensional copula strategies are significant at 10% significance level. In addition, the multi-dimensional copula strategy clearly outperforms the 2-dimensional copula and the distance strategies. Meanwhile, the average excess return of 2-dimensional copula strategy is higher than that of the conventional distance strategy. Furthermore, as seen in Figure 2, the cumulative returns of 2-dimensional and multi-dimensional copula strategies are increasing consistently while the cumulative return of the distance strategy tends to be increasing at a smaller magnitude. Although the cumulative return of 2-dimensional copula strategy tends to increase more than that of the multi-dimensional copula strategy, the latter shows less variations, indicating that the strategy is less risky.

[Insert Table 6 here.]

[Insert Figure 2 here.]

Furthermore, Table 7 presents the trading statistics of the 3 pairs trading strategies. It was found that the multi-dimensional copula strategy generates the highest average number of trades per pair, followed closely by the 2-dimensional copula strategy. This indicates that the copula methods are able to identify more trading opportunities than the conventional distance method. However, trading positions are open for a longer period of time (in months) for the 2-dimensional and multi-dimensional copula strategies. Having said that, there is greater certainty about the duration of an open position for the copula strategies as the standard deviations are significantly lower than that of the distance method.

[Insert Table 7 here.]

To examine whether the average excess returns generated by the 3 different strategies are due to any risk factors, the return series are ran against the Fama-French 3 factors (1993) and the Carhart momentum factor (1997), and the results are summarized in Table 8. The results show that the risk-adjusted return of the multi-dimensional copula strategy is positive and significant the 10% significance level. In addition, it is hardly explained by the risk factors and it is consistent with the raw average excess return in Table 6. Meanwhile, the risk-adjusted returns of the 2-dimensional copula and distance strategies are 21% and 27% lower than their respective raw average excess returns in Table 6, indicating that a relatively larger portion of their raw excess returns are explained by risk factors.

[Insert Table 8 here.]

4. Conclusion

Pairs trading is a well-known statistical arbitrage strategy that has always been implemented in 2 dimensions where two stocks are traded as a pair. Despite having documented significant consistent profits over time, there has been a significant decline in its profitability, especially the most popular conventional distance method. Thus, the copula method is proposed to serve as a solution to the declining profitability of 2-dimensional pairs trading strategies. The copula method is able to do so because it eliminates the need to imply that the financial data are normally distributed, and it provides an explicit function of joint distribution to describe dependence. However, in 2-dimensional pairs trading strategies, only one stock is used as a benchmark for another, meaning that only a portion of the full dependency information is considered. Therefore, there could be a case where the underlying stock is only locally mis-valued but intrinsically fairly-valued when it is being benchmarked against a comprehensive group of co-moving stocks. In this paper, multi-dimensional pairs trading is proposed to increase dependency information and generalize 2-dimensional pairs trading techniques.

Other than that, we also recognize the advantages of copula and therefore, have extended its application to multi-dimensional pairs trading. In comparison with the distance and 2-dimensional copula strategies, the multi-dimensional copula strategy has shown the highest average excess return and risk-adjusted return. On top of that, the 2-dimensional and multi-dimensional copula strategies have identified more trading opportunities and are more certain of the duration of an open position than the distance strategy.

Although the results have generally shown that the proposed multi-dimensional pairs trading strategy is superior to the 2-dimensional pairs trading strategies, there are certain rooms for improvements. For instance, a longer sample period could be adopted to increase the credibility of the results as a superior strategy should be able to generate consistent significant excess returns over a long period of time. In addition, a stocks selection process could be included to demonstrate the ability of the three different strategies to identify high-quality converging stocks. For example, the distance method captures only the linear associations as simple linear correlation coefficient is used, while the copulas are able to capture both linear and non-linear associations between random variables. Furthermore, moving from 2-dimensional copula to multi-dimensional copula, there may be an improvement of stocks selection ability as more dependency information is being utilized. Owing to processing power constraints, we are unable to include these aspects in this study. We believe that these unresolved issues are interesting topics that encourage further research and the resolutions of these issues will enhance the proposed strategy as well as establish the role of multi-dimensional pairs trading.

References

Andrade, S., Di Pietro, V., & Seasholes, M. (2005). Understanding the profitability of pairs trading. *Unpublished working paper, UC Berkeley, Northwestern University*.

Bock, M., & Mestel, R. (2009). A regime-switching relative value arbitrage rule. In *Operations Research Proceedings 2008* (pp. 9-14). Springer Berlin Heidelberg.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57-82.

Crook, J., & Moreira, F. (2011). Checking for asymmetric default dependence in a credit card portfolio: A copula approach. *Journal of Empirical Finance*, *18*(4), 728-742.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 703-738.

Do, B., & Faff, R. (2010). Does simple pairs trading still work?. *Financial Analysts Journal*, *66*(4), 83-95.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3-56.

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies*, *19*(3), 797-827.

Kat, H. M. (2003). The dangers of using correlation to measure dependence. *The Journal of Alternative Investments*, *6*(2), 54-58.

Liew, R. Q., & Wu, Y. (2013). Pairs trading: A copula approach. *Journal of Derivatives & Hedge Funds*, *19*(1), 12-30.

Perlin, M. S. (2009). Evaluation of pairs-trading strategy at the Brazilian financial market. *Journal of Derivatives & Hedge Funds*, *15*(2), 122-136.

Pizzutilo, F. (2013). A Note on the Effectiveness of Pairs Trading For Individual Investors. *International Journal of Economics and Financial Issues*, *3*(3), 763-771.

Sancetta, A., & Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, *20*(03), 535-562.

Sklar, A. (1959). "Fonctions de répartition á n dimensions et leurs marges." *Publications de l'Institut de Statistique de L'Universite de Paris*, *8*, 229-231.

Xie Wenjun, Liew Rong Qi, Wu Yuan & Zou Xi: "Pairs Trading with Copulas", *European Financial Management Association (EFMA) Annual Meeting 2014, Rome, Italy*.

Table 1. Trade Conditions and Positions for 2-Dimensional Copula

Open Trade Condition	Trading Positions		Close Trade Condition	Stop Loss Condition
	Long	Short		
$\mathrm{MI}_{2D}^{X} > \mathrm{T}_{2D}$	Stock Y	Stock X	MI_{2D}^{X} falls to 0	MI_{2D}^{X} rises to SL_{2D}
$\mathrm{MI}_{2D}^{X} < -\mathrm{T}_{2D}$	Stock X	Stock Y	MI_{2D}^{X} rises to 0	MI_{2D}^{X} falls to $-\mathrm{SL}_{2D}$
$\mathrm{MI}_{2D}^{Y} > \mathrm{T}_{2D}$	Stock X	Stock Y	MI_{2D}^{Y} falls to 0	MI_{2D}^{Y} rises to SL_{2D}
$\mathrm{MI}_{2D}^{Y} < -\mathrm{T}_{2D}$	Stock Y	Stock X	MI_{2D}^{Y} rises to 0	MI_{2D}^{Y} falls to $-\mathrm{SL}_{2D}$

Table 2. Trade Conditions and Positions for Multi-Dimensional Copula

Open Trade Condition	Trading Positions		Close Trade Condition	Stop Loss Condition
	Long	Short		
$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{X}} > \mathrm{T}_{\mathrm{MD}}$	Stock Y, Stock Z	Stock X	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{X}}$ falls to 0	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{X}}$ rises to $\mathrm{SL}_{\mathrm{MD}}$
$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{X}} < -\mathrm{T}_{\mathrm{MD}}$	Stock X	Stock Y, Stock Z	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{X}}$ rises to 0	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{X}}$ falls to $-\mathrm{SL}_{\mathrm{MD}}$
$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Y}} > \mathrm{T}_{\mathrm{MD}}$	Stock X, Stock Z	Stock Y	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Y}}$ falls to 0	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Y}}$ rises to $\mathrm{SL}_{\mathrm{MD}}$
$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Y}} < -\mathrm{T}_{\mathrm{MD}}$	Stock Y	Stock X, Stock Z	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Y}}$ rises to 0	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Y}}$ falls to $-\mathrm{SL}_{\mathrm{MD}}$
$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Z}} > \mathrm{T}_{\mathrm{MD}}$	Stock X, Stock Y	Stock Z	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Z}}$ falls to 0	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Z}}$ rises to $\mathrm{SL}_{\mathrm{MD}}$
$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Z}} < -\mathrm{T}_{\mathrm{MD}}$	Stock Z	Stock X, Stock Y	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Z}}$ rises to 0	$\mathrm{MI}_{\mathrm{MD}}^{\mathrm{Z}}$ falls to $-\mathrm{SL}_{\mathrm{MD}}$

Table 3. Optimization Result for the Distance Method

T SL	1.50	1.60	1.70	1.80	1.90	2.00	2.10	2.20	2.30	2.40	2.50
3.00	0.2370%	0.2310%	0.2470%	0.1520%	0.1580%	0.2030%	0.1670%	0.2650%	0.2460%	0.2520%	0.0830%
3.20	0.2080%	0.2020%	0.2170%	0.1500%	0.1520%	0.1980%	0.1610%	0.2530%	0.2250%	0.2300%	0.0550%
3.40	0.3170%	0.3110%	0.3140%	0.2270%	0.2300%	0.2760%	0.2490%	0.2850%	0.2670%	0.2810%	0.1550%
3.60	0.3590%	0.3630%	0.3660%	0.2810%	0.2840%	0.3300%	0.3030%	0.3410%	0.3230%	0.3360%	0.2180%
3.80	0.3590%	0.3520%	0.3560%	0.2710%	0.2740%	0.3190%	0.3160%	0.3540%	0.3280%	0.3410%	0.2240%
4.00	0.3630%	0.3660%	**0.3690%**	0.2850%	0.2870%	0.3330%	0.3300%	0.3670%	0.3410%	0.3550%	0.2380%
4.20	0.3610%	0.3630%	0.3660%	0.2810%	0.2840%	0.3300%	0.3270%	0.3640%	0.3380%	0.3510%	0.2340%
4.40	0.3600%	0.3540%	0.3570%	0.2720%	0.2750%	0.3210%	0.3180%	0.3550%	0.3280%	0.3420%	0.2250%
4.60	0.3570%	0.3510%	0.3540%	0.2690%	0.2720%	0.3180%	0.3150%	0.3520%	0.3260%	0.3390%	0.2220%
4.80	0.3330%	0.3610%	0.3480%	0.2810%	0.2780%	0.3240%	0.3210%	0.3570%	0.3300%	0.3430%	0.2250%
5.00	0.3320%	0.3530%	0.3360%	0.2790%	0.2760%	0.3220%	0.3190%	0.3550%	0.3280%	0.3410%	0.2230%

Table 4. Optimization Result for the Two-Dimensional Copula Method

SL \ T	0.55	0.56	0.57	0.58	0.59	0.6	0.61	0.62	0.63	0.64	0.65
1.50	-0.184%	0.125%	0.098%	0.168%	0.353%	0.206%	0.198%	0.244%	0.177%	0.056%	-0.007%
1.60	-0.404%	-0.314%	-0.194%	-0.227%	-0.053%	-0.105%	-0.137%	-0.277%	-0.243%	-0.385%	-0.379%
1.70	-0.116%	-0.042%	-0.062%	-0.041%	0.077%	0.010%	0.006%	-0.076%	-0.070%	-0.142%	-0.083%
1.80	0.009%	0.074%	0.079%	0.081%	0.079%	0.086%	0.090%	-0.001%	0.000%	-0.138%	-0.120%
1.90	0.325%	0.418%	0.451%	**0.467%**	0.449%	0.340%	0.340%	0.336%	0.352%	0.224%	0.151%
2.00	-0.005%	0.061%	0.166%	0.195%	0.293%	0.266%	0.270%	0.168%	0.179%	0.037%	-0.042%
2.10	-0.097%	-0.080%	0.060%	0.151%	0.264%	0.302%	0.255%	0.224%	0.242%	0.153%	0.090%
2.20	-0.228%	-0.218%	0.032%	0.150%	0.276%	0.271%	0.250%	0.214%	0.236%	0.124%	0.106%
2.30	-0.150%	-0.101%	0.137%	0.052%	0.179%	0.125%	0.141%	0.107%	0.146%	0.112%	0.032%
2.40	-0.070%	0.001%	0.000%	-0.090%	0.028%	-0.070%	-0.060%	-0.043%	-0.024%	-0.051%	-0.090%
2.50	0.025%	0.096%	0.132%	0.045%	0.022%	-0.040%	-0.038%	-0.015%	0.008%	-0.007%	0.048%

Table 5. Optimization Result for the Multi-Dimensional Copula Method

SL \ T	0.55	0.56	0.57	0.58	0.59	0.6	0.61	0.62	0.63	0.64	0.65
1.00	0.1845%	0.1715%	0.1772%	0.1983%	0.1110%	0.0650%	0.1010%	0.0380%	-0.0200%	0.0030%	-0.0160%
1.10	0.4077%	0.3342%	0.3783%	0.4302%	0.3340%	0.3830%	0.4040%	0.4560%	0.4430%	0.4880%	0.4620%
1.20	0.5195%	0.4881%	0.4746%	0.5151%	0.4990%	0.5340%	0.5170%	0.5240%	0.4950%	0.4920%	0.4620%
1.30	0.3697%	0.3723%	0.3629%	0.4082%	0.3640%	0.3760%	0.3730%	0.3800%	0.3570%	0.3430%	0.3210%
1.40	0.5126%	0.5268%	0.5096%	0.5285%	0.4730%	0.4510%	0.4890%	0.4910%	0.4820%	0.4600%	0.4670%
1.50	0.5402%	0.4817%	0.4930%	0.5195%	0.4960%	0.4820%	**0.5220%**	0.4810%	0.3820%	0.3830%	0.3390%
1.60	0.4699%	0.4602%	0.4341%	0.4532%	0.4294%	0.4248%	0.4408%	0.4564%	0.4563%	0.4571%	0.4531%
1.70	0.3152%	0.3000%	0.2934%	0.2967%	0.3046%	0.3297%	0.3303%	0.3245%	0.3009%	0.3127%	0.3045%
1.80	0.1803%	0.1801%	0.1948%	0.2226%	0.1825%	0.1908%	0.2137%	0.2327%	0.2061%	0.2263%	0.1800%
1.90	-0.1123%	-0.0996%	-0.0853%	-0.0598%	-0.1383%	-0.1153%	-0.1083%	-0.0911%	-0.1024%	-0.0983%	-0.1256%
2.00	0.0796%	0.0513%	0.0281%	0.0649%	0.0207%	0.0264%	0.0424%	0.0492%	0.0152%	0.0154%	0.0211%

Table 6. Comparison of Excess Returns of Different Pairs Trading Strategies

	Distance	Two-Dimensional Copula	Multi-Dimensional Copula
Average excess returns	0.0037*	0.0047	0.0052*
t-Statistic	1.6715	1.6592	1.7322
Skewness	1.3313	-0.4593	0.9197
Minimum	-0.0229	-0.0535	-0.02425
Maximum	0.0548	0.0572	0.05413

* represents 10% significance level.

Table 7. Comparison of Trading Statistics of Different Pairs Trading Strategies

	Distance	Two-Dimensional Copula	Multi-Dimensional Copula
Average no. of pairs traded per trading period	2.25	3	1
Average no. of trades per pair	1.0417	6.7917	6.7932
Standard Dev of number of round trips per pair	0.7506	2.1260	2.6121
Average time pairs are open in months	2.2321	4.2639	3.3330
Standard Deviation of time open, per pair, in months	1.8106	0.5384	0.7300

Table 8. Risk-Adjusted Returns of Different Pairs Trading Strategies

	Distance		Two-Dimensional Copula		Multi-Dimensional Copula	
	Coefficient	t-statistics	Coefficient	t-statistics	Coefficient	t-statistics
Alpha	0.0027	1.0306	0.0037	1.0403	0.0050*	1.8570
Mkt-RF	-0.0001	-0.3160	0.0002	0.3709	-0.0003	-0.7186
SMB	-0.0003	-0.3143	-0.0010	-0.6861	0.0006	0.4997
HML	0.0006	0.5150	0.0010	0.5668	0.0006	0.4391
WML	0.0012	1.5546	-0.0001	-0.1074	-0.0011	-1.3113

* represents 10% significance level.

Figure 1. Stock Prices over Time

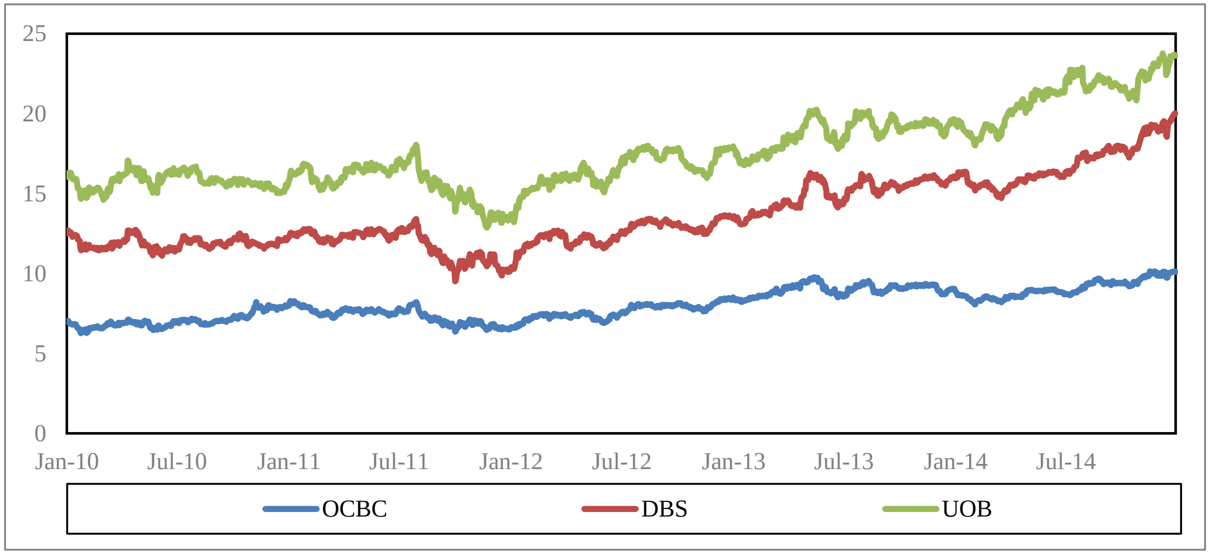

Figure 2. Cumulative Returns over Time