## Original Article

# Pairs trading: A copula approach

**Rong Qi Liew**
is a student studying Mathematics and Economics at School of Physical and Mathematical Sciences, Nanyang Technological University (NTU), Singapore.

**Yuan Wu**
is an Associate Professor in the Division of Banking and Finance, Nanyang Business School, NTU, Singapore. His research interests are in Actuarial Science, Data Mining, Application of Statistics to Business and Finance.

**Correspondence:** Rong Qi Liew, Nanyang Technological University (NTU), Block 48 Bendemeer Road #11-1487, 330048, Singapore.

**ABSTRACT**   **Pairs trading is a technique that is widely practiced in the financial industry. Its relevance has been constantly tested with updated samples, and its profitability is acknowledged among practitioners and academics. Yet in pairs trading, the notion of correlation is central, and the use of correlation or cointegration as a measure of dependency is ultimately its Achilles' heel. To overcome this limitation, this article employs the use of copulas, which is much more realistic and robust, to develop trading rules for pairs trading. Copulas are useful extensions and generalizations of approaches for modeling joint distributions and dependence between financial assets. A trading strategy that involves the use of copulas has been compared against two most commonly applied conventional strategies. The empirical results suggest that the proposed strategy is a potentially powerful analytical alternative to the traditional pairs trading techniques.**

## INTRODUCTION

Pairs trading is a well-acknowledged speculative investment strategy in the financial markets that is popularized back in the 1980s. Today, pairs trading is commonly applied by hedge funds and institutional investors as a long/short equity investment strategy. (Vidyamurthy, 2004) Recent researchers (Gatev *et al*, 2006; Do and Faff, 2010) have extended the initial analysis of

pairs trading to more updated samples, and documented economically and statistically significant profits using simple pairs trading rule.

Generally perceived as a form of technical analysis, the objective of pairs trading is to identify the relative overvalued and undervalued positions between two stocks that are closely related, with a long-run relationship. Such relative mispricing occurs if the spread between

the two stocks deviates from its equilibrium, and excess returns will be generated if the pair is mean-reverting (that is, any deviations are temporary, and will return to its equilibrium after a period of adjustment). In this situation, the stategy will simultaneously short the relatively overvalued stock and long the relatively undervalued. The formation of pairs ensues from a cointegration analysis or maximum correlation criteria of the historical prices. Thereafter, pairs trading strategy is implemented to identify the trading signals.

However, a significant shortcoming in the technique is the fundamental assumption of linear association and its use of correlation coefficient or cointegration as a measure of dependency. These basic assumptions may be convenient and useful in application, but they can cause the simple pairs trading signals to be inaccurate. If the data is normally distributed, then linear correlation completely describes dependency. But it is widely acknowledged that financial data are rarely normally distributed in reality, therefore correlation cannot completely describe the dependency. In fact, negative skewness and/or excess kurtosis are frequently observed in most financial assets (Kat, 2003; Crook and Moreira, 2011), resulting in upper and lower tail dependence of different extent. As such, correlation and cointegration are not sufficient in describing the association between financial assets and predicting their future movements.

The objective of this article is to link the use of copula with pairs trading to develop a trading strategy. Because copulas separate marginal distributions from dependence structures, the appropriate copula for a particular application is one that best captures dependence features of the data. (Trivedi and Zimmer, 2007) Hence,

the use of copula is able to capture the co-movement between stocks accurately enough to identify trading signals, which standard linear correlation analysis is not robust enough to accomplish (Ferreira, 2008). Therefore, it is hypothesized that a trading strategy involving the use of copula will bring about more trading opportunities, and potentially more profit than conventional strategies. The proposed strategy will be explored, and compared against conventional strategies.

The rest of the article will be organized as follow. The next section will provide a brief overview of pairs trading. The section 'Trading strategies' will describe the trading strategies studied in this article, and the empirical results will be demonstrated in the section 'Empirical results'. The section 'Conclusion' will conclude the article and provide directions for future studies.

## FOUNDATIONS OF PAIRS TRADING STRATEGIES

The general idea for investing in the marketplace from a valuation point of view is to sell overvalued securities and buy the undervalued. As the true values of the securities in absolute terms are rarely known, pairs trading techniques attempt to resolve this by looking at stock pairs with similar characteristics. Its objective is to identify the relative positions whenever an inefficient market results in the mispricing of securities. This mutual mispricing between two securities is theoretically captured by the notion of spread (Vidyamurthy, 2004).

Currently, there are several different pairs trading techniques applied in the modern financial industry. The two most commonly established techniques are the distance strategy

(Gatev *et al*, 2006; Perlin, 2009; Do and Faff, 2010) and cointegration strategy (Vidyamurthy, 2004; Lin *et al*, 2006; Galenko *et al*, 2012).

Generally, pairs are selected based on a cointegration analysis or minimum distance (equivalently, maximum correlation) criteria. When a suitable pair is identified, the traditional technique will engage in a simultaneous purchase of relatively undervalued stock and sale of relatively overvalued stock in an attempt to create a market-neutral trading system. This is to take advantage of the price divergence in terms of the spread that is expected to revert eventually, known as the mean-reverting behavior (Bock and Mestel, 2008). Hence, pairs trading is also known as a form of long/ short equity investment as the market-neutral strategy holds two stocks of different positions with equal market risk exposure at all times.

It is important to note that all the conventional techniques are essentially found on the assumption of linear association and its use of correlation coefficient or cointegration as a measure of dependency. In addition, the use of correlation and cointegration in pairs trading assumes a symmetric distribution of spread around the mean value of 0. Therefore, pairs trading has its limitations, and these may result in issues where the traditional pairs trading approach produces wrong trading signals or fail to identify profit opportunities (Bock and Mestel, 2008).

On the other hand, copulas provide a powerful framework for modeling dependence structure without rigid assumptions (Ferreira, 2008). It can potentially resolve the concerns mentioned earlier as it separates the estimation of individual marginal behavior and dependence structure into two different procedures. This separation of procedure is extremely valuable

and useful in many different aspects. From an economic perspective, it gives analyst the opportunity to use different marginal distributions to account for the diversity in financial risks (or assets) (Ane and Kharoubi, 2003). Hence, copula can be applied regardless of the form of marginal distributions, providing much greater flexibility for practical application. From a modeling position, the lower the dimensionality of a model or the cardinality of its parameters, the higher the reliability of the estimates. Hence, applying the best-fitting marginal distribution before estimating its joint distribution ensures that all information regarding the dependence structure between random variables are accurately captured without rigid assumptions.

Unlike the conventional approaches, employing a copula-based approach results in a far richer set of information, such as the shape and nature of the dependency between the stock pairs (Ferreira, 2008). This advantage is resulted because of the variety of copula choices that measures upper and lower tail dependencies of different extent, in an environment that considers both linear and non-linear relationship. For example, Gumbel copula produces more correlation at the two extremes of the correlated distribution but has its highest correlation in the maxima tails, whereas Clayton copula produces a tight correlation at the low end of each variable. In addition, copulas possess an attractive property of being invariant under strictly monotone transformations of random variables. In other words, the same copula will be obtained regardless of whether the analyst or researcher is using price series or log price series (Hu, 2003).

As a whole, copula is unique as it allows modeling dependence structure to be split into

two separate procedures. First, the choice of best-fitting marginal distribution is provided to describe the variables. Subsequently, a suitable copula is applied to establish the dependence structure. This two-step approach provides more alternatives in model specification, and an explicit dependence function obtained will provide more delicate description of dependence (Hu, 2003). These functionalities of copula ensure high accuracy and reliability of estimations, both of which are essential for financial analysis and application. Hence, the concept of copula will be explored as an alternative in this article where non-linear environment can be considered. For more details about copula, please refer Appendix A.

## TRADING STRATEGIES

In this section, the trading rules customized for three approaches, namely, copula, distance and cointegration, will be elaborated. In all three approaches, there are two different time periods, the formation period and trading (or backtesting) period. Historical data during formation period are used to observe price behavior and estimate the distribution and relevant parameters required for each approach. Using the estimated distributions and parameters from formation period, strategies are implemented during the trading period to test for profitability. There is no fixed guideline to the length of each time period, hence this article will use a 2-year and a subsequent 1-year phase as the formation and trading period, respectively.

Ultimately, all three approaches aim to identify the relative positions of the stock pairs and simultaneously long the undervalued stocks and short the overvalued, in attempt to establish a market-neutral trading system. The use of same formation period, trading period and stock pairs are maintained throughout the three approaches in this article due to the intention to make a comparison between the three.

## Copula approach

The objective of the pairs trading technique using copula approach is to apply the optimal copula between two stock returns, and identify the relative positions between stock pairs. In general, application of the strategy for copula approach will require the marginal distributions, the relevant copula function and the conditional probability distribution functions, which can be functions of copula.

Using data of the stocks during formation period, marginal distribution functions and the respective parameters are estimated based on the value of its cumulative log-returns. This can be done using any standard statistical analysis software that estimates the best-fitting marginal distribution. After applying the marginal distributions and relevant estimated parameters for each stock returns, the cumulative distribution function values obtained of each stock, $u$ and $v$, provides the information for a relevant copula function to be selected.

The last key information that is required for the strategy will be the conditional probability functions. By definition, the conditional probabilities $P(U \leqslant u \,|\, V = v)$ and $P(V \leqslant v \,|\, U = u)$ are derivatives of the copula with respect to $v$ and $u$, respectively.

$$P(U \leqslant u | V = v) = \frac{\partial C(u, v)}{\partial v}$$

$$P(V \leqslant v | U = u) = \frac{\partial C(u, v)}{\partial u}$$

As a general guideline, stocks are identified as being relatively undervalued if the conditional probability is less than 0.5 and relatively overvalued if the conditional probability is greater than 0.5. In addition, the values of the conditional probabilities are also an indication of its certainty or confidence regarding the position of stocks (Ferreira, 2008). Therefore, the execution of trade should be done when one of the conditional probabilities is close to 1. Hence, the use of conditional probability functions is essential for the strategy. For more information on the formulas of conditional probability functions, please refer Appendix B (Table B1).

For demonstration purposes, this article selects the upper bound of 0.95 and lower bound of 0.05 for the threshold of conditional probabilities as trading triggers in this approach. A position is opened during the trading period when one of the conditional probability values is above the upper bound, whereas the other is below the lower bound. Subsequently, exit position is assumed once the positions revert (that is, when the conditional probabilities cross the boundary of 0.5).

## Distance approach

In this article, we implemented the same trading strategy framework as described in Gatev *et al* (2006) for the distance approach. Upon finding two stocks that move together, long/short positions are taken when the stocks diverge abnormally. This divergence is determined by the difference between standardized price gap of the two securities, often known as spread. It acts as a signal to the open and close positions of the pairwise stocks. During the trading period, position is opened when the spread widens by more than two historical standard deviations, as estimated during the formation period. Thereafter, the positions are closed when spread of the stocks reverts. If reversion does not occur before the end of the trading period, profits or losses are calculated at the end of the last trading day of the trading period.

## Cointegration approach

Similar to the previous trading strategy, the main concern of this aproach is the movement of spread. However, instead of emphasizing on the distance between the standardized prices of stock pairs, the spread considered is based on the notion of error correction. The idea of error correction is based on the long-run equilibrium in a cointegrated system; that is, the long-run mean of the linear combination of two time series (Vidyamurthy, 2004). If there is a deviation from the long-run mean, then it is expected for one or both the time series to adjust in order for the long-run equilibrium to be restored.

Using cointegration as a theoretical basis, the general framework of cointegration approach in pairs trading consist of two parts. First is to generate the spread based on the actual cointegration error term of long-run relationship. This is estimated using regression analysis based on log-price series data from the formation period. Using the standard deviation of spread during formation period, a threshold of two standard deviation is set up for the trading strategy. Once the spread deviates from its long-term equilibrium and exceeds the threshold during the trading period, long/short positions are taken. Thereafter, positions are closed after the spread converges to its long-run equilibrium value of 0. For more details of the strategy framework conducted in this article, please refer Vidyamurthy (2004).

## EMPIRICAL RESULTS

This section will provide details of the actual implementation by demonstrating the three strategies using the pairs that have been applied. For reasons of space, only one example of Brookdale Senior Living Inc. and Emeritus Corporation (*BKD-ESC*) will be illustrated in detail. The stock pair is verified as one that is highly correlated and cointegrated. It is also one of the stock pairs listed on www.pairslog .com in the healthcare sector, specifically in long-term care facilities.

This article investigates the time period from 1 December 2009 to 30 November 2012. Data from the first 24 months are used to find the relevant parameters, and information obtained is applied on the trading period, which is the subsequent 12 months. The trading strategies of different approaches are studied and demonstrated, without the procedure of pairs

selection. Hence, only stock pairs that are widely discussed online or speculated in several literatures have been considered. Each stock pair has the same Schwarz Information Criterion (SIC) code to ensure industrial neutrality, or at least to reduce the industrial risk as well.

### Distance approach

A detailed figure for the conventional distance strategy is featured below to illustrate the standardized prices and spread of the values.

The overall trend of the standardized prices during the formation period and trading period are illustrated in Figure 1(a) and (c), respectively. Both diagrams indicate that the stock prices are highly correlated as they share similar co-movement in both time periods. On the other hand, the behavior of spread during formation period is displayed in
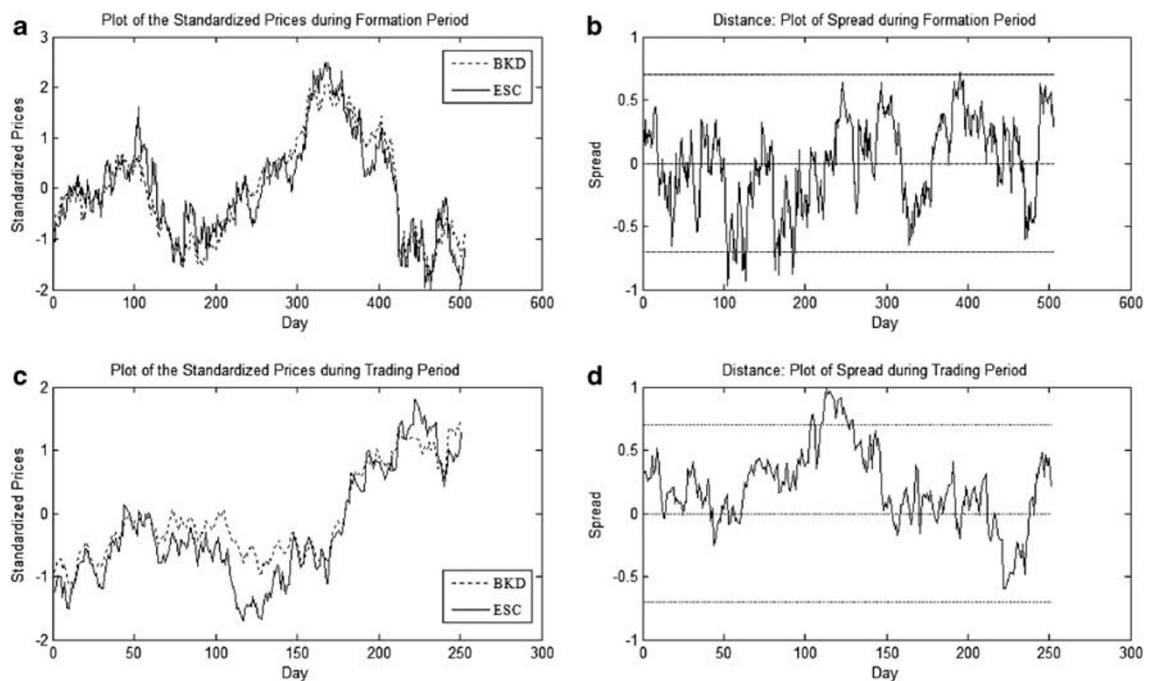


**Figure 1**: Distance trading strategy.

Figure 1(b), whereas the spread during trading period is illustrated in Figure 1(d). When compared against Figure 1(b), the pattern of spread illustrated in Figure 1(d) seems very different even though both figures indicate an asymmetric distribution of the spread. In fact, the behavior of spread illustrated in both figures seem to suggest two total different time series. This is likely due to the underlying assumptions of the approach that uses linear correlation coefficient as a measure of dependence. In addition, the approach presumed that the spread follows a symmetric distribution around 0. These assumptions ignored the complexity of the dependence structure that associates financial assets, therefore the use of simple correlation is insufficient in describing the dependency between the financial assets.

As a result, this approach could inaccurately translate the spread of formation period and trading period into two distinctively different time series even though the financial assets are closely related in both periods. Hence, the limitations of this approach in capturing the dependency could result in the discrepancy in behavior of the spread between the two time periods. This potentially brings about incorrect or the lack of trading signals and failure to identify profitable trading opportunities.

## Cointegration approach

In Figure 2, four diagrams are illustrated for the cointegration trading strategy. Note that the stock pair *BKD-ESC* has been tested and
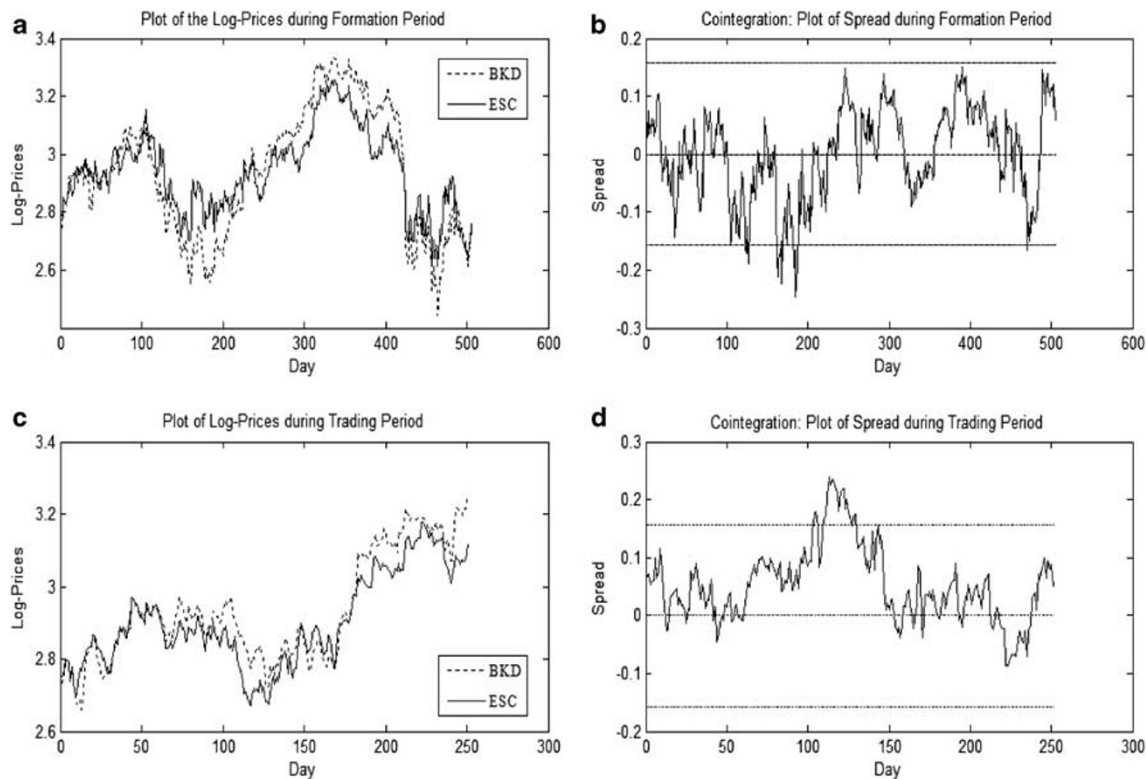


**Figure 2**: Cointegration trading strategy.

verified as a cointegrated stock pair at 5 per cent level of confidence.

Figure 2(a) shows a plot of the log-price series of the stock pair *BKD-ESC* during formation period and Figure 2(b) displays the spread during the same time period. The same applies for Figure 2(c) and Figure 2(d), respectively, but the values plotted are data from trading period.

Note that the spread of concern in this approach is obtained from the actual cointegration error term of the long-run relationship between the two log-price series, estimated using regression. Hence, the spread itself is a linear combination of the two log-price series. As the stock pair is verified to be cointegrated, it implies that the spread should thereotically be a stationary time series that is randomly distributed about the long-term mean value of 0. However, this is not illustrated in Figures 2(b) and (d). In fact, the spread behavior from the two time periods are distinctively different, with an asymmetric distribution. Hence, linear association and its use of cointegration as a measure of dependence in this approach is insufficient in capturing the association between the financial assets. This results in an inconsistency of spread values obtained during the two time periods. Note that the possibility of a structural change cannot be dismissed, but the limitatons of the approach itself seems more reasonable. The limitations of this conventional trading strategy could in turn result in an inaccurate perspective of the association between the assets concerned, possibly causing the lack of trading opportunities and therefore a lower profit in the trading strategy.

## Copula approach

As mentioned in the section 'Distance approach', this approach will first require the

**Table 1:** SIC, AIC and HQIC test values of copulas using formation period data

|            | SIC       | AIC       | HQIC      |
|------------|-----------|-----------|-----------|
| Gumbel     | −1098.4   | −1102.7   | −1101.0   |
| Student-*t* | −1028.8  | −1037.2   | −1033.9   |
| Normal     | −1030.4   | −1034.6   | −1032.9   |
| Frank      | −1003.2   | −1007.4   | −1005.7   |
| Clayton    | −653.1    | −657.3    | −655.7    |

marginal distributions of each variable; in this case, the cumulative log returns of *BKD* and *ESC* from the historical data. Using standard statistical analysis software, the marginal distributions fitted to the cumulative log returns of *BKD* and *ESC* are Error and Generalized Logistic distributions, respectively. Using the distribution functions and parameters estimated, the values of $u$ and $v$ are computed.

After obtaining the marginal distribution and relevant information, we identify the relevant copula function. By using the software *MATLAB*, suitability of each copula is tested based on the data obtained from the formation period, and the respective parameters are estimated. The SIC, Akaike Information Criterion (AIC) and Hannan-Quinn Information Criterion (HQIC) are common statistical tools used to identify the best-fitting copula. Table 1 illustrates the SIC, AIC and HQIC test values of the five copulas that are most commonly applied in financial research. Note that these are the same five copulas available on softwares such as *MATLAB* and *ModelRisk*.

As shown in Table 1, the dependence structure of Gumbel copula fits the formation period data best as it has the lowest SIC, AIC and HQIC test values among the tested.
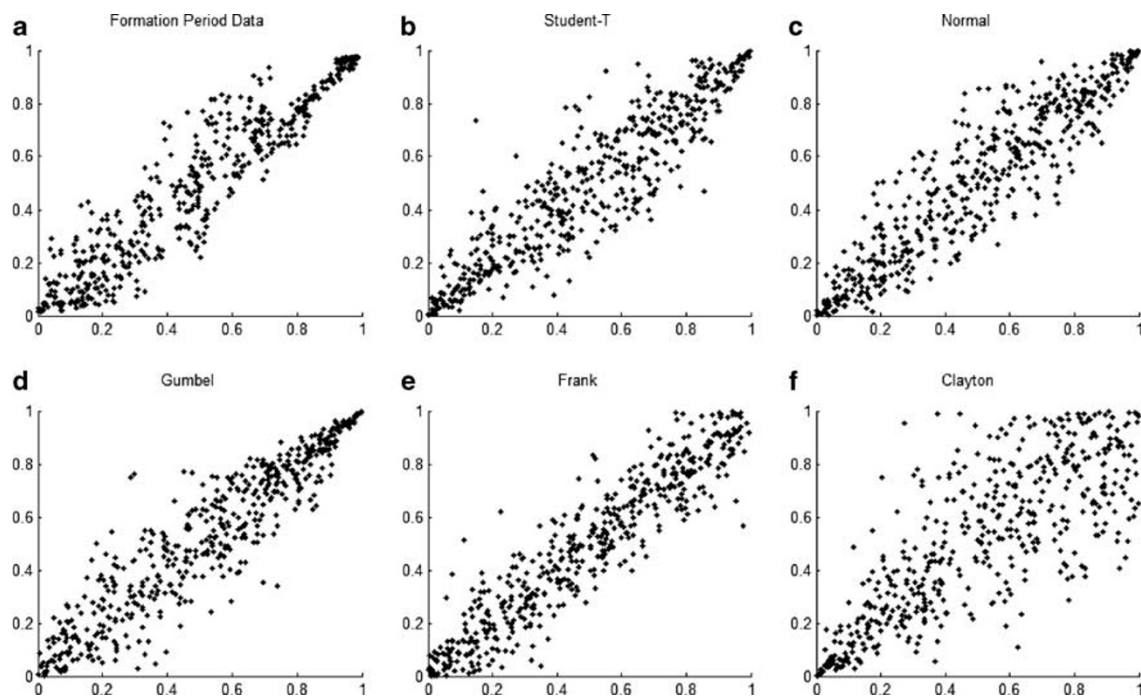
**Figure 3**: Plots of *u* and *v* values of formation period data and copulas fitted.

Figure 3 comprises six different scatter plots of *u* and *v*. Figure 3(a) is the plot obtained based on the formation period data, whereas Figures 3(b) to (f) are theoretical plots obtained from each of the five commonly applied copulas. With reference to the six diagrams, the plot of Gumbel in Figure 3(d) clearly matches the dependence structure illustrated in Figure 3(a) based on the density at the extreme ends and the overall trend. Hence, this further verifies that the Gumbel copula and its estimated parameters are suitable for application.

Note that extreme tail dependencies of different extent can be captured by copulas as seen from the different options provided by the copulas displayed in Figure 3. On the other hand, these crucial characteristics cannot be captured by correlation and cointegration as both indicators only measure the linear

association and assumes that observations follow a symmetrical pattern. Hence, copula will be able to make estimations and predictions that are closer to the reality.

Using the same marginal distributions and its parameters, the suitability of each copula is tested again, this time using data from the trading period (Table 2).

Clearly, the statistical test results suggest that Gumbel copula still matches the dependence structure of the data best, and Figure 4 further supports the choice. This consistency is important as it ensures that the long-term relationship of variables and its dependency features are appropriately captured during the formation period so that accurate predictions and profitable trade executions can be made during the trading period. Unlike the copula approach, both conventional

approaches have clearly failed to illustrate such consistency. This is indicated from the dramatic change in behavior of spread between the two time periods, possibly due to the conventional assumptions and limitations of the approaches. On the other hand, the use of copula dismisses the need for rigid assumptions,

thus it is an approach that is more sensitive and robust in capturing the true dependency and association features of data. In turn, trading signals could be accurately and reliably identified, providing more trading opportunities that may result in higher profit. Table 3 illustrates the details of results obtained for the three strategies executed on the stock pair *BKD-ESC*.

From Table 3, it can be seen that all three measures, Linear Correlation Coefficient, Kendall's Tau and Spearman Rho, indicate that the association between variables are consistent between the formation and trading period. Slight differences in the values are noticed, but inevitable during the two time periods. It generally indicates that the association between the variables should remain similar in the trading period as

**Table 2:** SIC, AIC and HQIC test values of copulas using trading period data

|  | *SIC* | *AIC* | *HQIC* |
|---|---|---|---|
| Gumbel | −298.7534 | −302.2628 | −300.8601 |
| Frank | −283.3148 | −286.8242 | −285.4215 |
| Student-*t* | −251.9826 | −258.9851 | −256.1961 |
| Normal | −138.1833 | −141.6927 | −140.2900 |
| Clayton | −125.9552 | −129.4646 | −128.0619 |
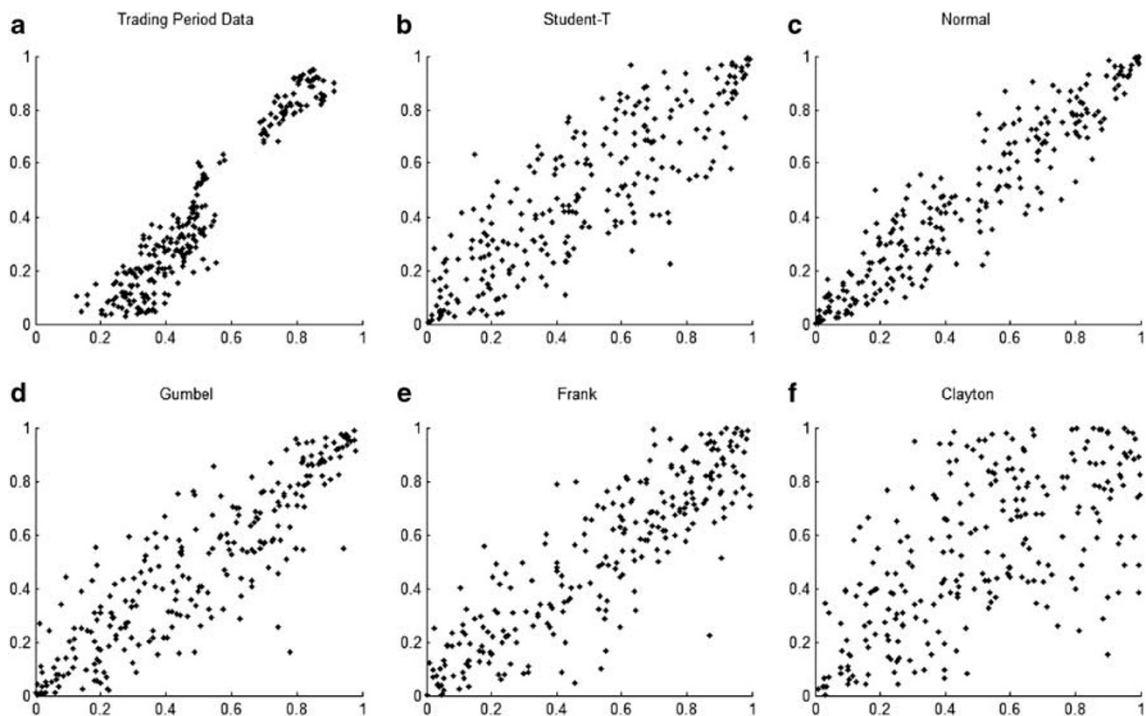


**Figure 4**: Plots of *u* and *v* values of trading period data and copulas fitted.

**Table 3:** Results of the trading strategies

|  | Distance | Cointegration | Copula |
|---|---|---|---|
| Correlation coefficient of log-price series during formation period |  | 0.9323 |  |
| Correlation coefficient of log-price series during trading period |  | 0.9430 |  |
| Kendall's $\tau$ of log-price series during formation period |  | 0.7825 |  |
| Kendall's $\tau$ of log-price series during trading period |  | 0.7620 |  |
| Spearman $\rho$ of log-price series during trading period |  | 0.9367 |  |
| Spearman $\rho$ of log-price series during trading period |  | 0.9231 |  |
| *Panel A: Trading strategy based on thresholds designated in the section 'Empirical results'* |  |  |  |
| Profit (Capital: 10 000) | 1760.10 | 1760.10 | 2549.10 |
| Number of transactions | 1 | 1 | 2 |
| Returns of trading period | 0.1927 | 0.1927 | 0.267 |
| *Panel B: Trading strategy in Panel A with 1-day waiting period* |  |  |  |
| Profit (Capital: 10 000) | 1828.00 | 1828.00 | 2700.80 |
| Number of transactions | 1 | 1 | 2 |
| Returns of trading period | 0.1944 | 0.1944 | 0.2774 |

compared with the formation period. Yet this is not displayed in the trading strategies for distance (Figure 1(b) against Figure 1(d)) and cointegration approach (Figure 2(b) against Figure 2(d)). Such consistency is only demonstrated in the copula approach, as the same copula is preferred in both time periods. In addition, Table 4, which summarizes the results of three examples studied during the research process, indicates that copula approach does bring about higher returns and more trading transactions. This justifies that copula is able to estimate and predict the movement of financial assets more precisely as it measures the association and dependency without any rigid assumptions or limitations. Please refer Appendix C and D (Figures C1–3 and D1–3) for relevant graphs of the additional examples in Table 4.

## Transactions costs

In real-world applications, it is certain that transaction costs have to be taken into consideration. Hence, this article will address the issues regarding bid–ask spreads and short sales lost.

According to Gatev *et al* (2006), the use of same prices for the start of trading and returns may be biased upward because of the fact that transactions executed in the study are implicitly buying at bid quotes (losers) and selling at ask quotes (winners). Hence, to address this concern, this article has done the same as Gatev *et al* (2006) for the example on stock pair *BKD-ESC*, and initiated positions on the day following the divergence and liquidates on the day following the convergence. The values of profits, number of transactions and returns are computed and illustrated in Panel B

**Table 4:** Summary results of the trading strategies based on thresholds designated in the section 'Empirical results'

| Strategy | Transactions | Profit | Returns |
| --- | --- | --- | --- |
| *Panel A: Trading strategies executed on BKD-ESC* | | | |
| Distance | 1 | 1760.10 | 0.1927 |
| Cointegration | 1 | 1760.10 | 0.1927 |
| Copula | 2 | 2549.10 | 0.267 |
| | | | |
| *Panel B: Trading strategies executed on APA-DVN* | | | |
| Distance | 1 | 992.71 | 0.117 |
| Cointegration | 1 | 992.71 | 0.117 |
| Copula | 2 | 1113.20 | 0.1176 |
| | | | |
| *Panel C: Trading strategies executed on BOH-CYN* | | | |
| Distance | 1 | 1207.20 | 0.1098 |
| Cointegration | 1 | 1239.80 | 0.1117 |
| Copula | 1 | 1242.90 | 0.1166 |

of Table 3. However, the authors understand that it is meaningless to carry out such precaution on a few stock pairs. In order for the precaution to be effective in addressing the issue of bid–ask spread, it is essential for the research to be extended to an entire market. Yet this is currently out of the scope of this article, thus it can be one of the aspects that the authors are considering for future studies.

On the other hand, the concerns regarding cost of short–selling is dismissed in this article as it has been verified by Gatev *et al* (2006) that pairs trading profits are robust to short–selling costs. Hence, it is presumed that the impact of short–selling costs is mitigated by the use of liquid stocks that trade every day over a period of 1 year and it is therefore viewed as insiginificant.

## CONCLUSION

Pairs trading is a technique that is commonly applied in the financial industry. In this article, the concept of copulas in pairs trading is explored to overcome the limitations of traditional pairs trading strategies. The use of copulas in constructing joint distributions separates the estimation of individual marginal distributions from the dependence structure. This brings about much greater flexibility in the framework when specifying joint distributions, while providing richer information regarding the dependency between financial assets. Hence, copulas will bring about estimations that are more realistic and of better precision.

The empirical results demonstrate that the copula approach for pairs trading is superior to the conventional. It is observed that the use of copula in pairs trading provides more trading opportunities in practical application, and with greater confidence, as it does not require any rigid assumptions. The use of correlation or cointegration as a measure of dependency is also disregarded, thus the proposed strategy offers a potentially powerful analytical alternative to the traditional pairs trading techniques.

Despite the overall superiority of results obtained from copulas, this is only a preliminary study, and it is certainly imperfect. Copulas provide flexibility, and the strategy based on copula is considered relatively easy to implement for such a sophisticated approach. However, it is still a new approach in the trading area after all. There is much more discovery and further improvements to be done to overcome the limitations of the current work. Furthermore, the use of copulas may also be explored in the selection of stock pairs. Pairs selection is the first and most essential

step to pairs trading, thus improving the conventional selection process will be a new leap forward in this area.

# REFERENCES

Ane, T. and Kharoubi, C. (2003) Dependence structure and risk measure. *Journal of Business* 76(3): 411–438.

Bock, M. and Mestel, R. (2008) A regime-switching relative value arbitrage rule. University of Graz, Institute of Banking and Finance.

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

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

Ferreira, L. (2008) New tools for spread trading. *Futures* 37(12): 38–41.

Galenko, A., Popova, E. and Popova, I. (2012) Trading in the presence of cointegration. *The Journal of Alternative Investments* 15(1): 85–97.

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

Hu, L. (2003) *Dependence Patterns across Financial Markets: A Mixed Copula Approach*. Columbus, OH: The Ohio State University, Department of Economics.

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

Lin, Y-X., McCrae, M. and Gulati, C. (2006) Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Journal of Applied Mathematics and Decision Sciences* 2006(4): 1–14.

Nelsen, R.B. (2006) *An Introduction to Copulas*, 2nd edn. New York: Springer.

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

Sklar, A. (1959) *Fonctions de répartition à n dimensions et leurs marges*, Vol. 8. Publications de l'Institut de Statistique de L'Université de Paris, pp. 229–231.

Trivedi, P.K. and Zimmer, D.M. (2007) Copula modeling: An introduction for practitioners. *Foundations and Trends® in Econometrics* 1(1): 1–111.

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

# APPENDIX A

## Copula and Sklar's Theorem

In the statistics literature, the idea of a copula arose as early as the nineteenth century in the context of discussions regarding non–normality in multivariate cases (Hu 2003). Currently, copulas are known as tools used in probability theory and statistics for modeling dependence between random variables.

The behavior of two random variables, $X$ and $Y$, with the respective marginal distributions of $F(x)$ and $G(y)$, is appropriately described by their joint distribution $H(x,y) = P(X \leqslant x, Y \leqslant y)$. As $U = F(X)$ and $V = G(Y)$ are uniformly distributed between 0 and 1, their joint distribution for every $(u, v) \in [0, 1]^2$ can be expressed as

$$\begin{aligned} C(u, v) &= P(F(X) \leqslant u, G(Y) \leqslant v) \\ &= P\big(X \leqslant F^{-1}(u), Y \leqslant G^{-1}(v)\big) \\ &= H\big(F^{-1}(u), G^{-1}(v)\big) \end{aligned}$$

where $F^{-1}$ and $G^{-1}$ are the quasi-inverse of the margins $F(x)$ and $G(y)$.

Under this construction, $C$ is a distribution function of two random variables with uniform unit values. Such a function is formally defined as a copula. In general sense, copulas are known as functions that combine individual one-dimensional marginal distributions to form a multivariate distribution function that describes both the linear and non–linear relationship between variables. It ensures that the dependency between variables are accurately captured and described in terms of a function.

The word 'copula' was first employed in a statistical sense by Sklar (1959), in the theorem that now bears his name. His idea was to separate a joint distribution function into one part that

describes the dependence structure (the copula) and the other that describes marginal behavior.

*Sklar's theorem*: For a joint distribution function $H(x, y)$ with margins $F(x)$ and $G(y)$, the theorem states that there exists a copula $C$ such that for all $x, y$ in $\bar{R}^2$, where $\bar{R}$ is the extended real line $[-\infty, \infty]$, the following equation applies.

$$H(x, y) = C(F(x), G(y))$$

If $F(x)$ and $G(y)$ are continuous then $C$ is unique; otherwise $C$ is uniquely determined on $\text{Ran}(F) \times \text{Ran}(G)$, where Ran denotes range. Conversely, if $C$ is a copula, and $F(x)$ and $G(y)$ are distribution functions then the function $H(x, y)$ defined by the above equation is a joint distribution function with margins $F(x)$ and $G(y)$. Certainly, the notion of copula can be extended to higher dimensions: $n$-dimensional copulas are joint distribution functions of $n$ random variables with unit uniform marginal. Formal proofs can be found in Nelsen (2006).

In words, the Sklar's theorem translates that every copula is a joint distribution function with margins that are uniform in the domain of the copula. An essential implication is that a copula can be constructed from any joint distribution function with continuous marginal distributions. This is ultimately Sklar's idea of the concept of copula where the best-fitting marginal distributions are estimated before a joint distribution function that describes the dependence structure (a copula) is estimated and then applied.
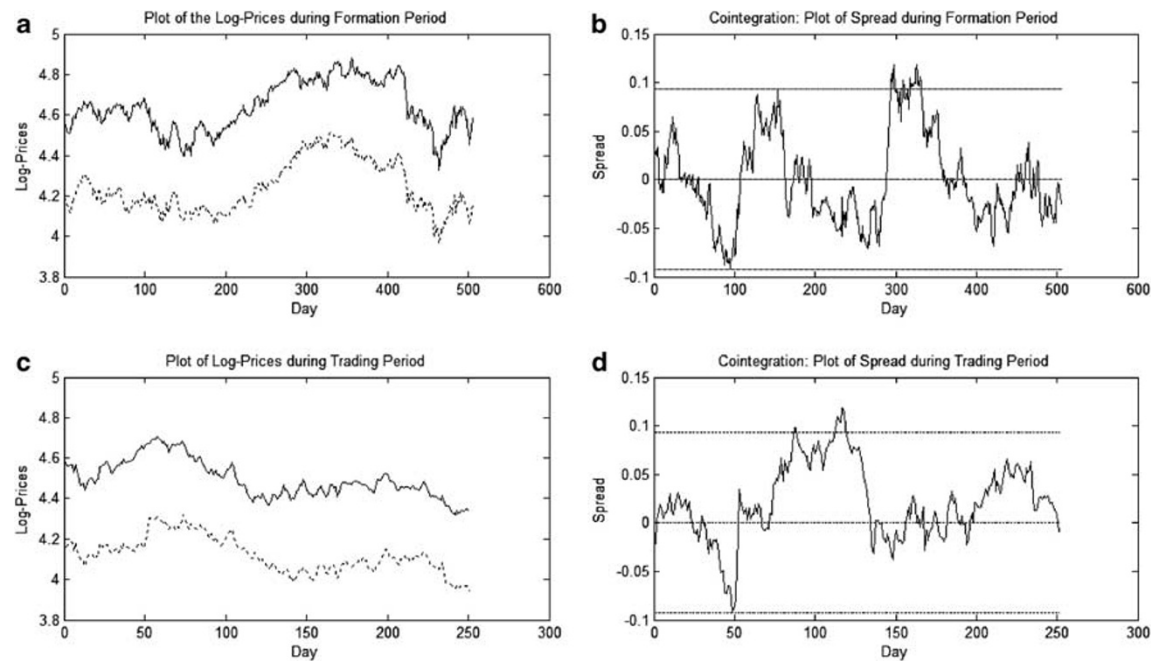
# APPENDIX B

**Table B1:** Conditional probability formulas of commonly applied copulas

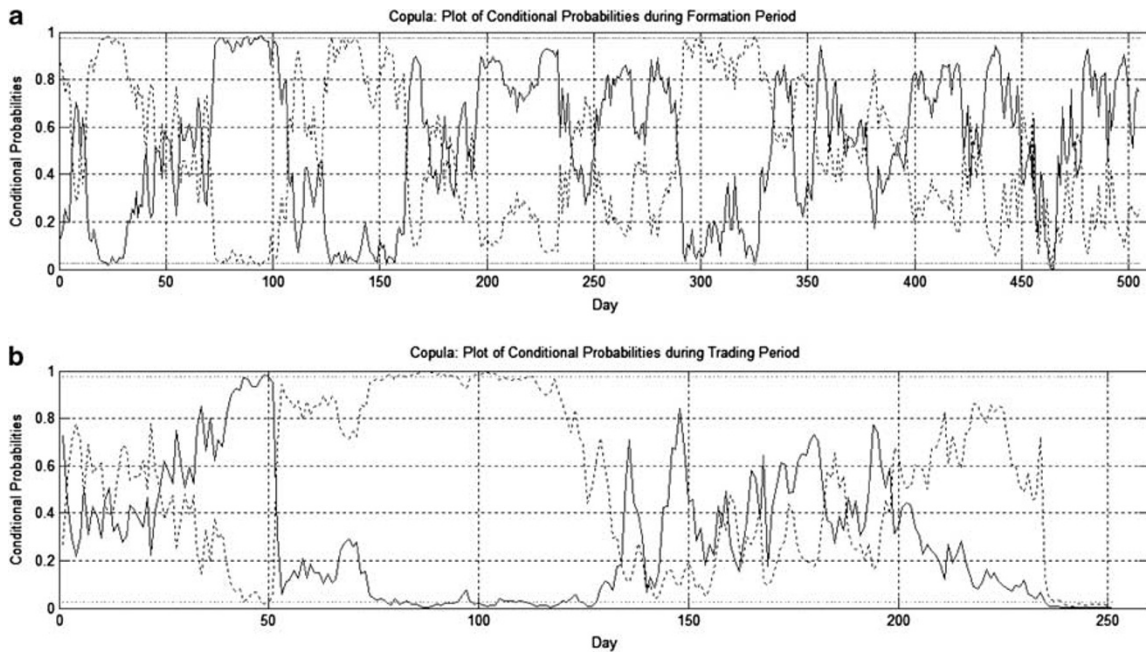| S/N | Copula | $\theta \in$ | $P(U \leqslant u \mid V = v) = \dfrac{\partial C}{\partial v}$ | $P(V \leqslant v \mid U = u) = \dfrac{\partial C}{\partial u}$ |
|---|---|---|---|---|
| 1 | Student-$t$ | $\theta_1 \in (0, \infty)$ $\theta_2 \in (-1, 1)$ | $t_{\theta_1 + 1}\left( \sqrt{\dfrac{\theta_1 + 1}{\theta_1 + t_{\theta_1}^{-1}(v)^2}} * \dfrac{t_{\theta_1}^{-1}(u) - \theta_2 t_{\theta_1}^{-1}(v)}{\sqrt{1 - \theta_2^2}} \right)$ | $t_{\theta_1 + 1}\left( \sqrt{\dfrac{\theta_1 + 1}{\theta_1 + t_{\theta_1}^{-1}(u)^2}} * \dfrac{t_{\theta_1}^{-1}(v) - \theta_2 t_{\theta_1}^{-1}(u)}{\sqrt{1 - \theta_2^2}} \right)$ |
| 2 | Normal | $(-1, 1)$ | $\Phi\left( \dfrac{\Phi^{-1}(u) - \theta \Phi^{-1}(v)}{\sqrt{1 - \theta^2}} \right)$ | $\Phi\left( \dfrac{\Phi^{-1}(v) - \theta \Phi^{-1}(u)}{\sqrt{1 - \theta^2}} \right)$ |
| 3 | Gumbel | $[1, \infty]$ | $C_\theta(u, v) * \left[ (-lnu)^\theta + (-lnv)^\theta \right]^{\left(\frac{1-\theta}{\theta}\right)} * (-lnv)^{\theta - 1} * \dfrac{1}{v}$ | $C_\theta(u, v) * \left[ (-lnu)^\theta + (-lnv)^\theta \right]^{\left(\frac{1-\theta}{\theta}\right)} * (-lnu)^{\theta - 1} * \dfrac{1}{u}$ |
| 4 | Frank | $(-\infty, \infty) \backslash \{0\}$ | $\dfrac{e^{-\theta v}(e^{-\theta u} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1)}$ | $\dfrac{e^{-\theta u}(e^{-\theta v} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1)}$ |
| 5 | Clayton | $[-1, \infty) \backslash \{0\}$ | $v^{-(\theta+1)}\left( u^{-\theta} + v^{-\theta} - 1 \right)^{-\frac{1}{\theta} - 1}$ | $u^{-(\theta+1)}\left( u^{-\theta} + v^{-\theta} - 1 \right)^{-\frac{1}{\theta} - 1}$ |
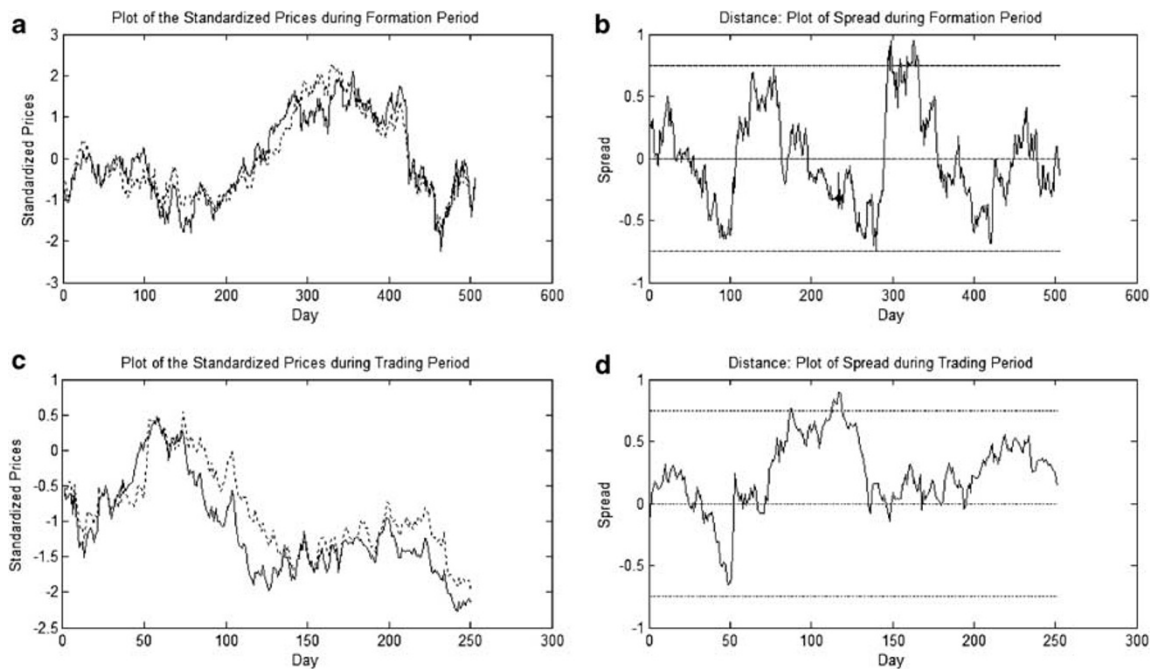
**Figure C1**: Distance trading strategy.
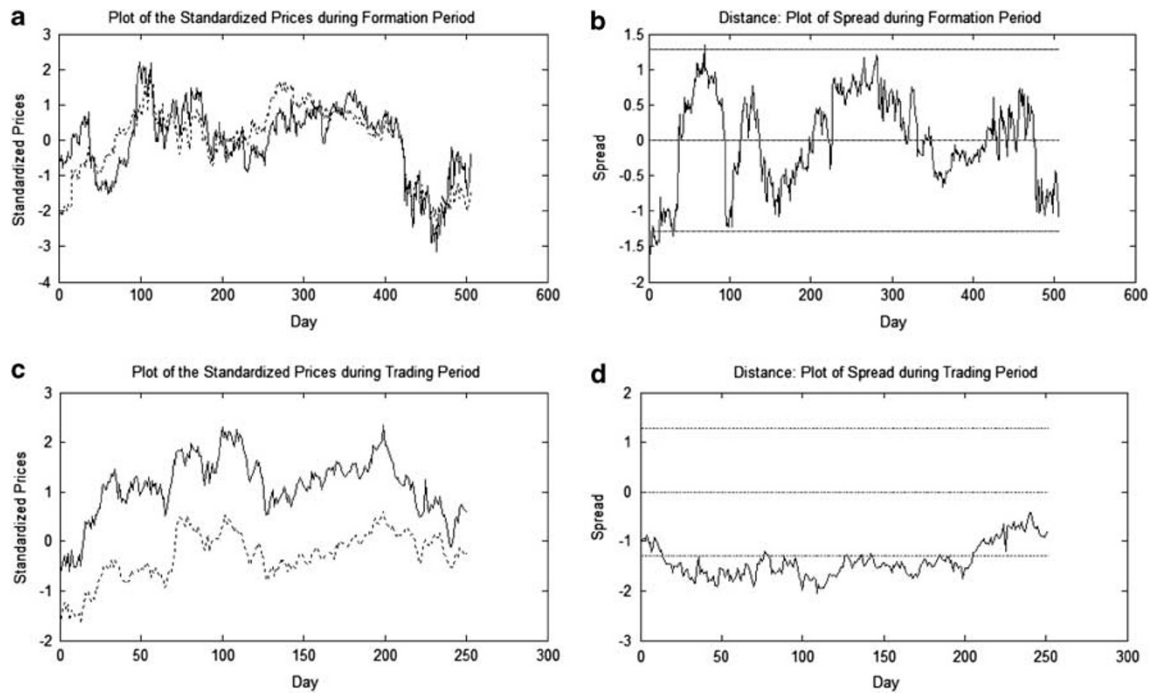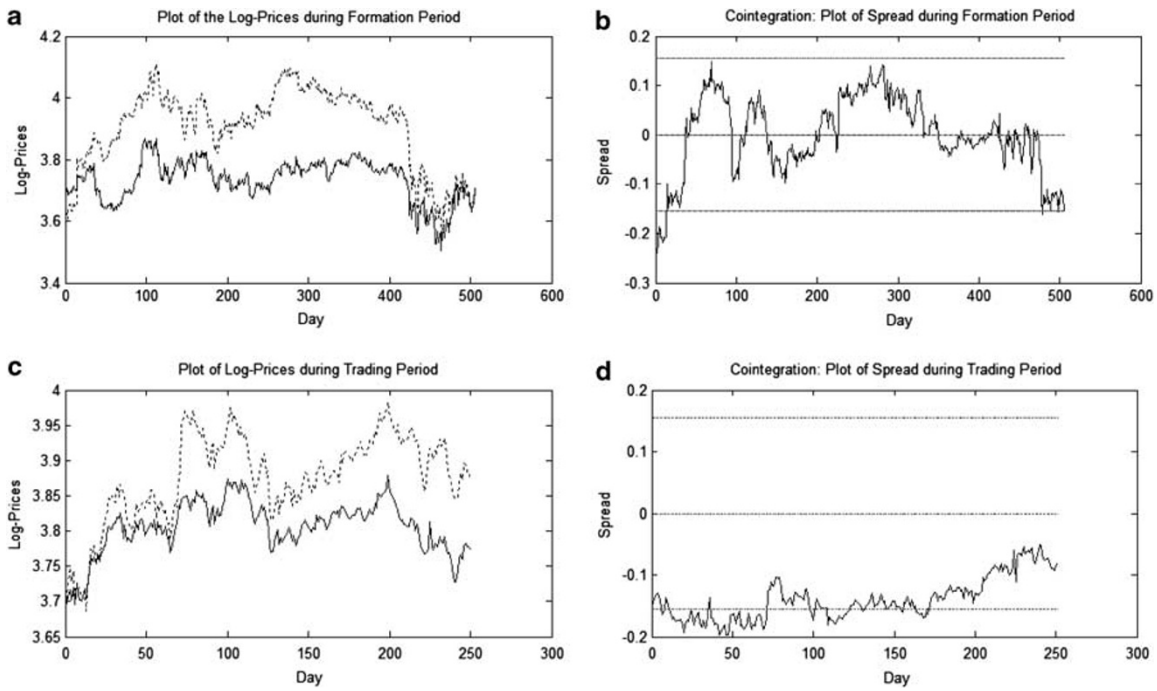


**Figure C2**: Cointegration trading strategy.

**Figure C3**: Conditional probabilities of copula trading strategy.
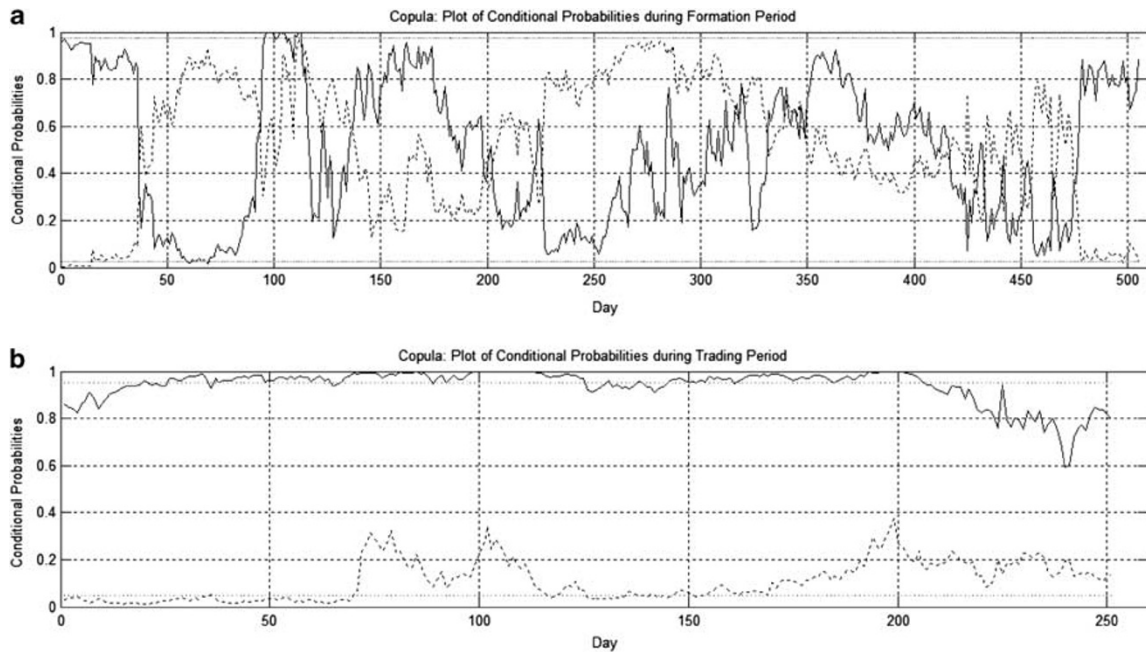
# APPENDIX C

## GRAPHS OF STOCK PAIR *APA-DVN*

# APPENDIX D

## GRAPHS OF STOCK PAIR *BOH-CYN*



**Figure D1**: Distance trading strategy.

**Figure D2**: Cointegration trading strategy.



**Figure D3**: Conditional probabilities of copula trading strategy.