A COMPARISON OF COINTEGRATION AND COPULA ASSET ALLOCATION APPROACHES

Y S Stander, D J Marais and I Botha*

Abstract

The empirical performance of cointegration and copula asset allocation techniques are compared against that of the market. Multivariate copula structures are used to derive index-tracking portfolios which are then compared with that of portfolios constructed using cointegration techniques. The results suggest that modelling the long-term relationships between stocks by means of the cointegration approach do not consistently lead to portfolios that outperform the benchmark. Using a short-term asset allocation approach, such as the copula-simulation approach, lead to portfolios that perform at least as well as the cointegration portfolios.

1. Introduction

Cointegration has been explored by various authors as a way to construct index-tracking portfolios that consistently outperform a benchmark. Cointegration allows for simple estimation methods to capture dependencies between non-stationary time series - see the work of Alexander, Giblin & Weddington (2001), Alexander & Dimitriu (2005), Caldeira & Moura (2012) and Chiu & Wong (2012).

Correlation and cointegration are related, but not necessarily interlinked. Alexander *et al*. (2001) argue that correlation is usually derived from stationary series which are based on returns, and is thus a short-term measure. High correlation is not sufficient to ensure the long-term performance of a tracking portfolio. Cointegration on the other hand is based on the hypothesis that the spread in a portfolio of stock prices is mean-reverting in the long-run and thus the portfolios constructed using cointegration should track the benchmark closely over time.

Stock returns are more highly correlated in market downturns than in market upturns, which is referred to as asymmetric dependence (Patton, 2004; Hong, Tu & Zhou, 2007). Patton (2002) proposes modelling the asymmetric dependence with a copula structure by making use of utility functions. The idea is to estimate the

*Respectively, Senior quantitative analyst at one of the major banking groups in South Africa. Finalised her PhD at the University of Johannesburg.
Supervisor, Department of Economics and Econometrics, University of Johannesburg; Senior lecturer, Department of Finance and Investment management, University of Johannesburg; Email: ilseb@uj.ac.za

portfolio weights by maximising the excess return earned by an index-tracking portfolio above the benchmark index. Hennessy & Lapan (2002) also use utility functions, but consider a simplistic *n*-dimensional Archimedean copula with a single generator function across all dimensions.

The focus of this article is to compare the cointegration and copula-based asset allocation approaches. Instead of using the utility-function approach, a new simulation-based copula approach is employed by fitting multidimensional d-vine copula structures to the portfolio of equities.

There are various asset allocation strategies employed in the market. See for instance the summary in Connor & Lasarte (2004) on the strategies followed by hedge funds, which includes long/short strategies, relative value strategies that seeks to take advantage of mispricing between related financial instruments, event-driven strategies that seeks profits from events like corporate restructuring and stock buy-backs, as well as tactical trading strategies where profits are made from views on the future direction of the market. Portfolio managers have to make decisions around allocation between different asset classes like equity and interest-bearing investments; the liquidity in the market; the investment horizon; and the financial goals and the risk aversion of investors to cater for the needs of different investors (Infanger, 2011).

In this article empirical analyses are performed using different trading strategies and sets of stocks over various economic conditions to test the performance of the two approaches over time. Transaction costs are ignored for the two cases, assuming that it would be similar.

In Section 2 the theory behind the cointegration and copula asset allocation techniques is discussed while the two approaches are compared in Section 3 in an empirical study using South African equities traded on the Johannesburg Stock Exchange. The main findings are summarised in Section 4

2. Theory

This section provides a brief overview of the theory behind the cointegration and copula asset allocation techniques as well as how back-testing is performed. In each case a portfolio is constructed that can track the performance of a benchmark index.

2.1 Cointegration

2.1.1 Asset allocation

Granger (1981) showed that the linear combination of two or more non-stationary series can be stationary if the series are cointegrated. The theory is further extended and formalised in Engle & Granger (1987). They define a series with no deterministic component and which is stationary after differencing d times, to be integrated of order d written as $I(d)$. For instance, if $\log(X_t)-\log(X_{t-1})$ is stationary, then $\log(X_t)$ is $I(1)$.

The augmented Dickey-Fuller (ADF) test statistic is used to test the stationarity of the data series. The ADF statistics used in the analyses are derived using the Matlab functions written by LeSage (1999). He uses the methodology discussed in Said & Dickey (1984). A more recent summary on cointegration tests can be found in MacKinnon (2010).

The benchmark index is a weighted average of stocks, which means that it should be possible to identify a linear combination of stocks that is cointegrated with the index. If the portfolio of stocks is sufficiently large and the index weights do not change too much over time, the tracking error will be stationary (Alexander *et al.*, 2001; Alexander & Dimitriu, 2005). A cointegration regression analysis is performed as follows:

$$\log(B_t)=\beta_1\log(X_1)+\cdots+\beta_n\log(X_n)+\varepsilon_t \qquad \dots (1)$$

where

B_t denotes the benchmark index; X_i denotes the stock prices; ε_t is the tracking error; and β_i the regression coefficients to be estimated. The log-prices are used because when taking the first differences of Equation (1), the expected return on the index will equal the expected return on the index-tracking portfolio. If the price series are cointegrated, then the log-price series will also be cointegrated. The Engle-Granger optimisation routine is used to estimate the regression coefficients is

$$\min_{\beta_i,i=1,\dots,n}\left[\log(B_t)-\sum_{i=1}^{n}\beta_i\log(X_i)\right] \qquad \dots (2)$$

with all symbols as defined before. The regression coefficients are normalised to sum to one using:

$$\beta_i^*=\frac{\beta_i}{\sum_{j=1}^{n}\beta_j} \qquad \dots (3)$$

where

β_i^* denotes the adjusted coefficient such that $\sum_{i=1}^{n}\beta_i^*=1$. The normalised coefficients are used as the weights in the index-tracking portfolio. The value of the index-tracking portfolio P_t is derived as

$$P_t = \beta_1^* X_1 + \beta_2^* X_2 + \cdots + \beta_n^* X_n \quad \ldots (4)$$

with all symbols as defined before. The performance of the index-tracking portfolio will be compared to the performance of a portfolio where the weights are estimated using a copula asset allocation technique.

2.1.2 Back-test approach

The back-test is performed by selecting a training period that is used to estimate the portfolio weights, and then selecting a back-test period which is the out-of-sample period that is used to test the performance of the index-tracking portfolio. The steps can be summarised as follows:

Step 1: Select the training period. For instance if the training period is one year, then data from January to December 2002 is used to estimate the portfolio weights.

Step 2: Estimate the cointegration regression coefficients using Equation (2).

Step 3: Calculate the ADF of the training sample tracking error. This is to check whether there is a cointegrated relationship during the training period.

Step 4: Derive the portfolio weights using Equation (3).

Step 5: Calculate the return that an investor would have earned if she invested in the portfolio during the back-test period. If a back-test period b = 10 workdays is used, then the back-test period would cover the first two weeks in January 2003. The return over the back-test period is derived as $R_{t+b} = \log\left(\frac{P_{t+b}}{P_t}\right)$ where P_t denotes the value of the index-tracking portfolio.

Step 6: Move the training period on by *b* days so that the new training period ends at the end of the previous back-test period. Using the training period of 1 year and a back-test period of 10 workdays implies that the new training period will cover middle-January 2002 to middle-January 2003.

Step 7: Repeat steps 2 to 6 until the end of the historical dataset is reached.

2.2 Copulas

2.2.1 Fitting the multivariate dependence structure

Abe Sklar originally defined a copula by showing the relationship between the multivariate distribution function and the copula function which joins the marginal distribution functions (Sklar, 1959). In this article the focus is on Archimedean copulas, which are derived from generator functions that have very specific properties. Let ϕ be a continuous and strictly decreasing generator function from [0,1] to $[0,\infty)$ such that $\phi(1)=0$. The Archimedean copula function C is then given by

$$C(u,v)=\phi^{[-1]}\left(\phi(u)+\phi(v)\right)$$

where

$\phi^{[-1]}$ is the pseudo-inverse of φ and is defined as

$$\phi^{[-1]}(t)=\begin{cases}\phi^{(-1)}(t) & ,0\le t\le\phi(0)\\ 0 & ,\phi(0)\le t\le\infty.\end{cases}$$

C is a copula if and only if ϕ is convex, that is $\phi''>0$. If $\phi(0)=\infty$ then ϕ is called a strict generator function and it follows that $\phi^{[-1]}(t)=\phi^{(-1)}(t)$ (Nelsen, 2006).

Examples of the generator functions for some of the more popular Archimedean copulas are:

- Clayton: $\phi(t)=\frac{1}{\alpha}\left(t^{-\alpha}-1\right)$, $\alpha\in[-1,\infty)/\{0\}$
- Gumbel: $\phi(t)=(-\ln t)^{\alpha}$, $\alpha\in[1,\infty)$
- Frank: $\phi(t)=-\ln\left(\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}\right)$, $\alpha\in(-\infty,\infty)/\{0\}$
- N14: $\phi(t)=\left(t^{-1/\alpha}-1\right)^{\alpha}$, $\alpha\in[1,\infty)$

The parameter α denotes the strength of the relationship between two variables where a higher α implies a greater dependence. A good summary of the properties of the bivariate Archimedean copulas can be found in Joe (1997) and Nelsen (2006).

The parameters of the bivariate Archimedean copulas are estimated using the canonical maximum likelihood approach (see Cherubini, Luciano & Vecchiato (2004) for a detailed discussion). With this approach no assumption regarding the marginal distributions are made. The empirical marginal distribution functions are determined and then the copula parameter is estimated by:

$$\hat{\alpha}=\arg\max\sum_{t=1}^{n}\ln c\left(u_t,v_t\right)$$

where

n denote the number of observations in the sample and $\hat{\alpha}$ denotes the estimated copula parameter. The copula density function $c(u,v)$ is defined by

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}.$$

The Kolmogorov-Smirnov goodness-of-fit test can be used to indicate the best fit copulas. This test compares the distance between the parametric distribution function of the Archimedean copula and its nonparametric counterpart under the null hypothesis. The parametric distribution function $K_C(t)$ of the random variable $C(U,V)$ can be derived from the generator function $K_C(t) = t - \frac{\phi(t)}{\phi'(t_+)}$. The non-parametric counterpart of this distribution function is:

$$\hat{K}_C(v) = \frac{1}{n}\sum_{j=1}^{n} \text{ind}\left(V_j \le v\right)$$

where

$$V_j = \hat{C}\left(u_j, v_j\right) \text{ for } j = 1, \ldots, n$$

where

n denotes the number of observations in the sample. $\hat{C}$ denotes the empirical copula estimate. More information on the empirical copula can be found in Cherubini *et al*. (2004). Please refer to Genest & Rémillard (2005), Genest, Rémillard & Beaudoin (2007) and Patton (2012) for the details around the goodness-of-fit test and approaches to derive the p-value using bootstrap techniques.

The copula asset allocation approach considered in this article is a simulation-based approach, where the simulation study is performed by fitting an (n+1)-dimensional copula structure to model the dependence between the n equity price series and the index. The dependence structure is fitted to the log-returns of the price series derived as $\log\left(\frac{X_{i,t}}{X_{i,t-1}}\right)$ where $X_{i,t}$ denotes the price for equity i at period t (typically day t).

The d-vine structure is used to estimate the parameters of the multivariate copula. Aas, Czado, Frigessi & Bakken (2009) show how to use the bivariate Archimedean copulas to build the multivariate structure and the algorithm to sample from the d-vine structure. The d-vine structure is illustrated in Figure 1. To derive the multivariate copula, bivariate copulas are iteratively fitted to different variable-pairs. For instance, the second level indicates that the bivariate copulas are fitted to {U;V}, {V;W} and {W;X} respectively. The third level shows that the bivariate copulas are fitted to {U|V; W|V} and {V|W; X|W} using the conditional copula function. The process is repeated until the inverted apex of the tree is reached.

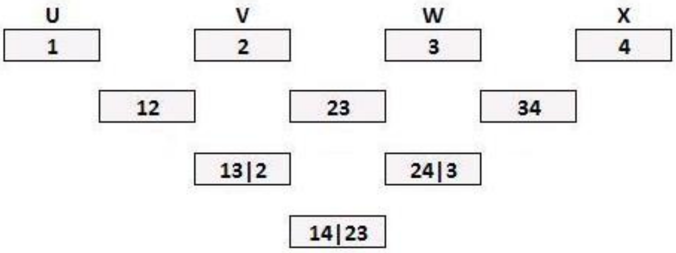


Figure 1: An example of a d-vine structure for four variables

2.2.2 Asset allocation

The index-tracking portfolio weights are estimated as follows with the copula simulation approach:

- Generate N uniform values from the (n+1)-dimensional dependence structure and denote these by $U_1,\ldots,U_{n+1}$ where the first n variables denote the equity price-return series and U_{n+1} denotes the benchmark index return.

- Use the probability integral transform to obtain values from the equity return distributions $R_i^* = F_i^{-1}(u_i)$ where F_i^{-1} denotes the inverse of the cumulative equity return distribution for equity i. The simulated index return values are obtained similarly and are denoted by R^{B*}. In each case the empirical distribution function is used.

- Estimate the regression coefficients by minimising the difference between the returns of the benchmark index and the returns of the index-tracking portfolio using the following optimisation routine:

$$\min_{\beta_i, i=1,\ldots,n} \left[R^{B*} - \sum_{i=1}^{n} \beta_i R_i^* \right] \quad \ldots (5)$$

where

$\beta_1,\ldots,\beta_n$ are the weights to be estimated.

- Normalise the regression weights to sum up to 1 using Equation (3).

By generating values from the copula dependence structure (instead of just using historical observations), it is possible to simulate sets of observations that have not occurred historically. This approach is computationally intensive because of the fact that the multivariate dependence structure has to be estimated and returns generated from the (n +1)-dimensional structure.

2.2.3 Back-test approach

The back-test of the copula simulation approach is performed following these steps:

Step 1: Select the training period.

Step 2: Estimate the copula parameters of the multivariate dependence structure of the index and equity returns using the data in the training period. It is assumed that the copula dependence structure is known; only the copula parameters have to be updated.

Step 3: Calculate the p-values of the Kolmogorov-Smirnov test statistics to determine whether the copula at each node fits the data adequately.

Step 4: Generate N values from the multivariate dependence structure.

Step 5: Estimate the portfolio weights by minimising the difference between the returns of the benchmark index and the index-replicating portfolios using Equation (5).

Step 6: Calculate the portfolio return that an investor would earn during the back-test period. The return over the back-test period is derived as $R_{t+b} = \log\left(\frac{P_{t+b}}{P_t}\right)$.

Step 7: Move the training period on by b days so that the new training period ends at the end of the previous back-test period.

Step 8: Repeat steps 2 to 7 until the end of the historical dataset is reached.

2.3 Deriving annual portfolio returns

The annual returns are derived by rebalancing the portfolio every two weeks and accruing profits and losses until the end of the year, when the excess income is paid out so that at the beginning of the next year the portfolio starts with the initial value again. Mathematically this can be expressed as:

$$R_j^{annual} = \prod_{i=1}^{n} \exp\left(R_i\right) - 1 \qquad \text{… (6)}$$

where

R_j^{annual} denotes the annual return for year *j*, R_i denotes the return earned over period *i* and *n* denotes the total number of investment periods in a year.

3. Comparing the cointegration and copula asset allocation techniques

Empirical analyses are performed to compare the performance of index-tracking portfolios constructed using cointegration techniques with the performance of portfolios constructed using copula techniques. Two different trading strategies are tested. The first trading strategy only allows long positions in the equities. Allowing long positions only implies that the investor is allowed to buy stocks into the portfolio, but not to sell stocks short (i.e. selling stocks not owned). The second trading strategy allows long and short positions. In all cases it is assumed that the portfolios are rebalanced every two weeks.

3.1 Market data

The ALSI40 is used as the benchmark index when the index-tracking portfolios are constructed. The ALSI40 index is derived from the most liquid stocks traded on the Johannesburg Stock Exchange (JSE) in South Africa. Please refer to JSE (2012) for details around how the index is constructed.

The sample of stocks that will be used in the back-test analysis is compiled by first selecting the most liquid stocks traded on the JSE during June 2011. The sample is then adjusted to only include stocks with a full price series from January 2002 to December 2010. This is done to ensure that an adequate historical dataset is available to be able to do a full back-test analysis over various economic conditions. All prices are in South Africa Rand, denoted by ZAR, and the equities are referenced by using their JSE stock codes.

Once the sample of viable stocks has been identified, the next step is to select the number of assets to include when deriving the index-tracking portfolio. Alexander *et al*. (2001) choose the optimal portfolio size by analysing the information ratio over different training periods and using different portfolio sizes. The information ratio is defined as $IR = \frac{\mu_\varepsilon}{\sigma_\varepsilon}$ where μ_ε is the mean tracking error and σ_ε the standard deviation of the tracking error.

Figure 2 depicts two contour plots derived using the same dataset. The contour plot on the right was derived by first sorting the equities according to their market capitalisation (equities with the biggest market cap are included first). The plot shows a relatively good information ratio (high values) when about 15 to 20 stocks are included using a training period between 1 and 3 years. The contour plot on the left was derived by sorting the equities by volatility. Using the highly volatile stocks first seems to indicate that fewer equities need to be included in the analysis but that longer training periods have to be used. The issue with using this approach to determine the portfolio size is that the two contour plots show very different results depending on the order in which the equities are considered. It was decided to limit the portfolio size to 10 stocks in the back-test analyses.

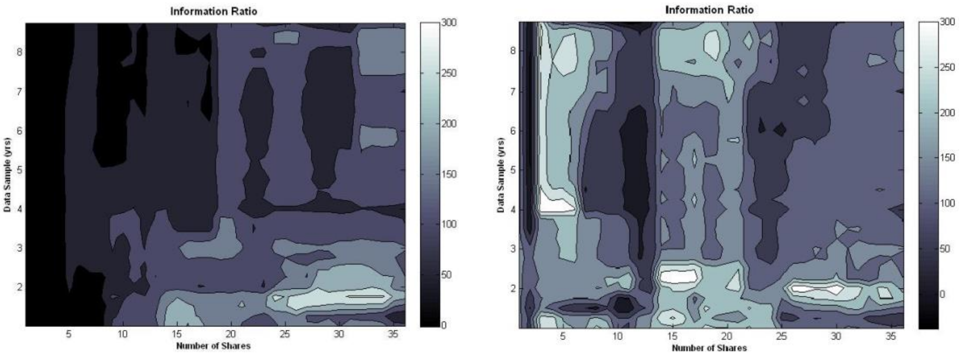


Figure 2: Contour plots of the information ratio derived using an out-of-sample test period of 1 month

There are various approaches that can be considered to select which 10 stocks to include in the index-tracking portfolio. Stock selection strategies include forming portfolios based on earnings-to-price ratios, book-to-market ratios, price momentum or earnings revisions (Van der Hart, Slagter & Van Dijk, 2001). In this paper two sets of stocks were selected based on the stocks' Sharpe ratios and market capitalisations using information as at December 2002.

The stocks selected based on the highest Sharpe ratios, are those stocks with the highest return per unit of risk. The selected stocks are {HAR; GFI; TRU; AEG; MUR; APN; PPC; TBS; PIK; MSM} as indicated by the JSE stock codes.

By selecting stocks with the highest market cap, it is expected that the portfolio should perform very similar to the ALSI40, because it would be aligned with the manner in which the ASLI40 is derived. The selected stocks in this portfolio are {SBK; FSR; IMP; NPN; SAB; ANG; SOL; OML; BIL; AGL}.

The back-test analyses can be distorted significantly by stock splits, so the stock price data was transformed to eliminate the effect of the stock splits. Assuming a stock split occurred on day *t*, resulting in a significant decrease in the stock price from day t-1 to day *t* (denoted by the return R_t). Rather than using the extreme return (due to the stock split), use the average of the returns R_{t-1} and R_{t+1} to replace the value of R_t. Then recalculate all the stock prices from day 1 to day t using $X_t = X_{t-1} \exp(R_t)$. An example of an equity where the historical prices were adjusted for a stock split is shown in Figure 3. The price series are not adjusted for dividends

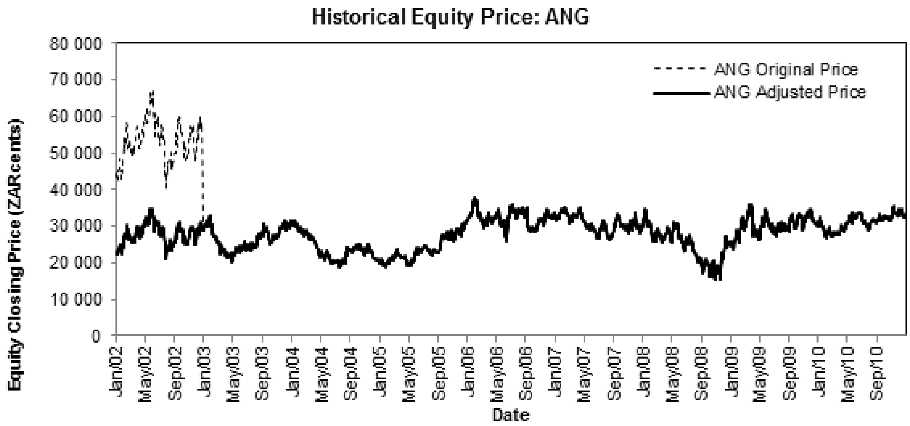


Figure 3: An example of how the closing prices of the ANG equity were adjusted to reduce the effect of the stock split

3.2 Asset allocation using cointegration

3.2.1 Statistical tests for cointegration

The first step in the cointegration analysis is to ensure that the benchmark index is cointegrated with the portfolio of equities. The ADF statistics in Tables 1 and 2 indicate that none of the equity price series are stationary; however, the log-returns denoted by I(1) are stationary. These results are in line with the results by Alexander & Dimitriu (2002), who found that stock prices and stock market indices are usually integrated of order 1.

Table 1: ADF statistics of the closing prices denoted by I(0) and log-returns of the equities denoted by I(1) chosen on the basis of their market capitalisation and that of the benchmark index

	I(0)	I(1)
ALSI	0,84	-34,02
SBK	1,08	-37,41
FSR	0,81	-36,54
IMP	0,52	-35,89
NPN	2,58	-34,55
SAB	1,31	-36,33
ANG	0,13	-35,52
SOL	1,00	-34,61
OML	-0,19	-36,51
BIL	1,12	-35,22
AGL	0,32	-34,11

Table 2: ADF statistics of the closing prices and log-returns of the equities chosen on the basis of their Sharpe ratios

	I(0)	I(1)
HAR	-0,06	-33,37
GFI	0,24	-35,46
TRU	2,52	-35,43
AEG	1,16	-33,74
MUR	0,94	-33,92
APN	2,46	-35,27
PPC	1,16	-37,67
TBS	1,38	-36,14
PIK	1,55	-37,77
MSM	2,18	-34,16

Figures 4 and 5 shows the tracking error ε_t derived using Equation (1) on the full historical dataset for the portfolios derived on the basis of the market capitalisation and the Sharpe ratio respectively. The ADF statistics of the tracking errors are -5,27 and -5,07 respectively and indicate that the benchmark index and the portfolios are cointegrated over the full period from January 2002 to December 2010. The assumption of a cointegrated relationship implies that the relationship between the stocks and the benchmark index is relatively stable over time. However, the figures show that there are periods with a clear positive trend and others with a clear negative trend. From this it can be concluded that there exist subintervals where the data may not be cointegrated.

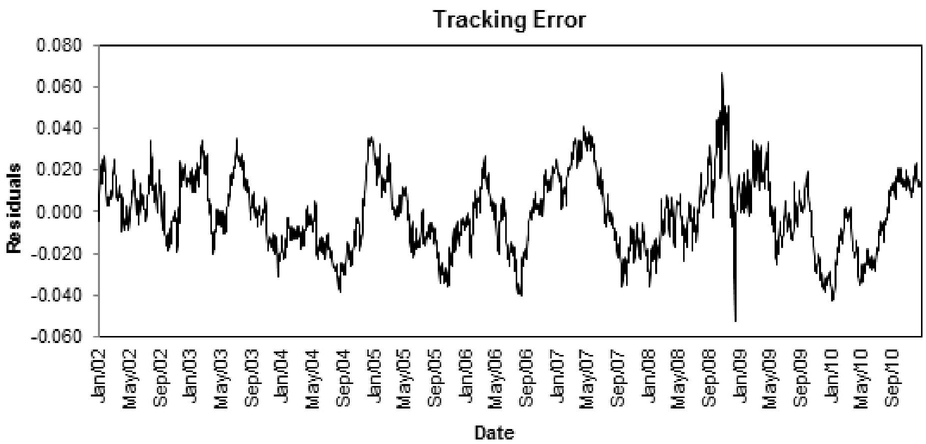


Figure 4: The tracking error ε_t on the full historical dataset of the portfolio of stocks chosen on the basis of their market capitalisation

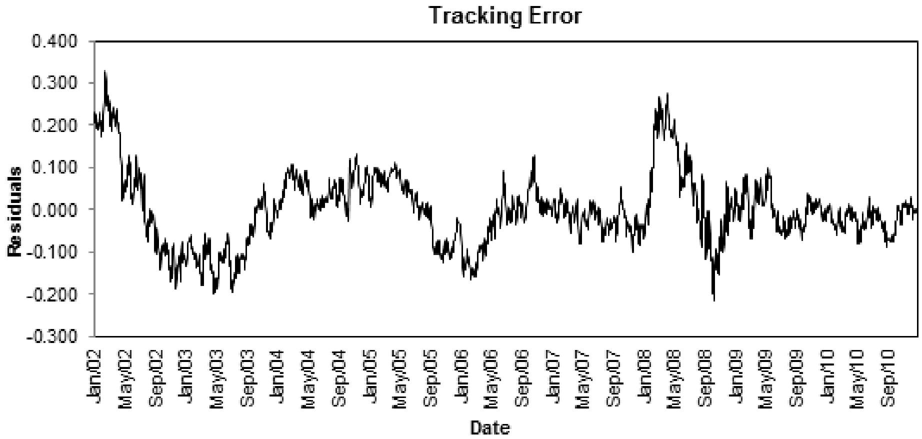


Figure 5: The tracking error ε_t on the full historical dataset of the portfolio of stocks chosen on the basis of their Sharpe ratios

3.2.2 Back-test analysis: Testing the length of the training period

The first back-test analysis is performed on the portfolio chosen on the basis of the market capitalisation of the stocks. The analysis is performed using a one-year and a three-year training period respectively to test whether the cointegration relationship is more stable when using a longer training period and to address the issue with the trends observed over certain sub-periods (please refer to Figures 4 and 5).

It is assumed that the investor follows a long-only trading strategy. The regression coefficients are estimated using the optimisation routine specified in Equation (2) with the added constraint that $0 \le \beta_i^* \le 1$ to ensure that only long positions are captured.

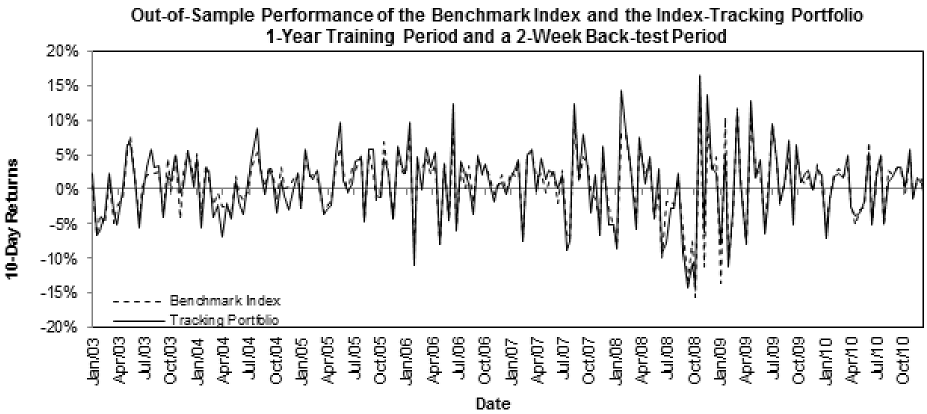


Figure 6: A comparison of the performance of the index-replicating portfolio (chosen on the basis of the market cap of the stocks) and the benchmark index over the two-week back-test period using a one-year training period

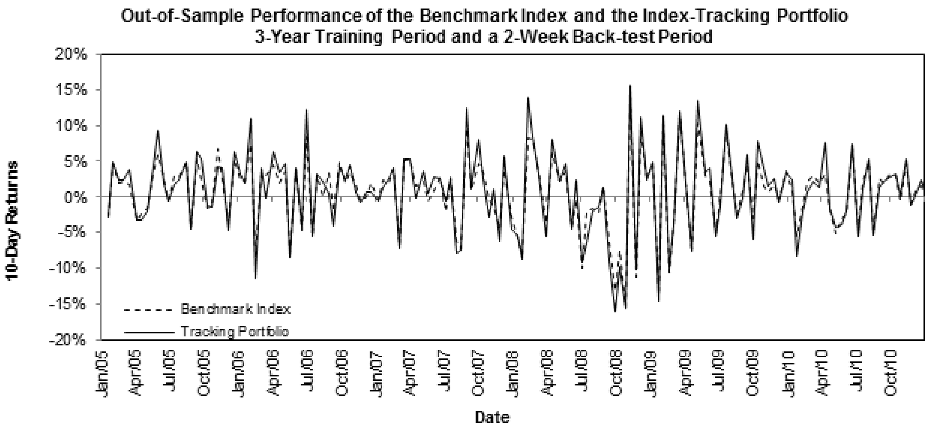


Figure 7: A comparison of the performance of the index-replicating portfolio (chosen on the basis of the market cap of the stocks) and the benchmark index over the two-week back-test period using a three-year training period

Figure 6 shows a comparison of the performance of the index-replicating portfolio and the benchmark index over each of the two-week back-test periods when using the one-year training period. The performance of the portfolio is very similar to that of the benchmark index. The returns of the benchmark index and the index-replicating portfolio also show similar annualised volatility of 23% and 26% respectively. Similar results are indicated in Figure 7 where the index-replicating portfolio is derived using a three-year training period. The two-week returns are relatively stationary except over the 2008 stress period where more negative returns are observed.

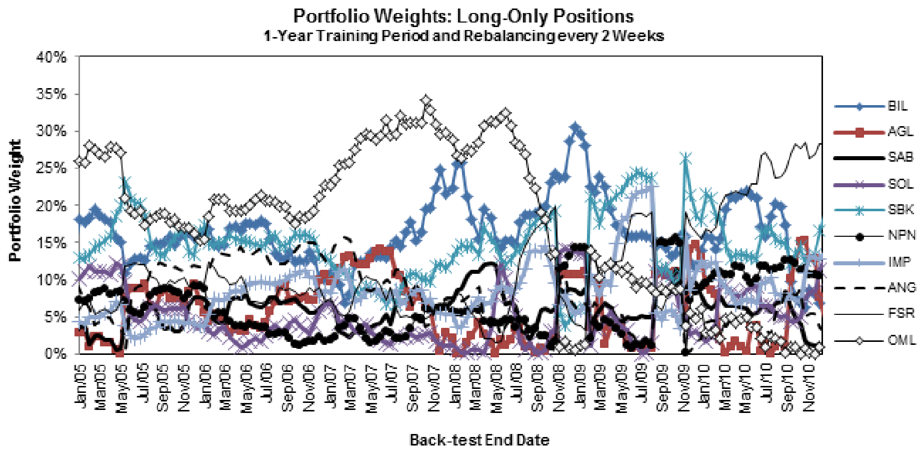


Figure 8: Normalised regression weights for each back-test period using the stocks chosen on the basis of their market capitalisation based on the one-year training period

Figure 8 shows the portfolio weights estimated for each of the back-test periods when using the one-year training period. The OML stock has the highest weight in the portfolio up to 2009 where the weight then decreases substantially so that by the end of 2010 it has the lowest weight in the portfolio. The stocks in most cases do not carry a weight of more than a third of the portfolio. The stock BIL has a relatively high weight over the full back-test period. The stock SOL has one of the lowest weights in the portfolio over time.

The fact that the weights in general change slowly over time may indicate that there is a proper cointegrated relationship between the benchmark index and the index-tracking portfolio. The weights derived from the three-year training period show similar behaviour. Alexander & Dimitriu (2005) indicate that volatile portfolio weights may indicate a spurious cointegration relationship. They also conclude that a true cointegration relationship should not yield an extreme exposure to any individual stock.

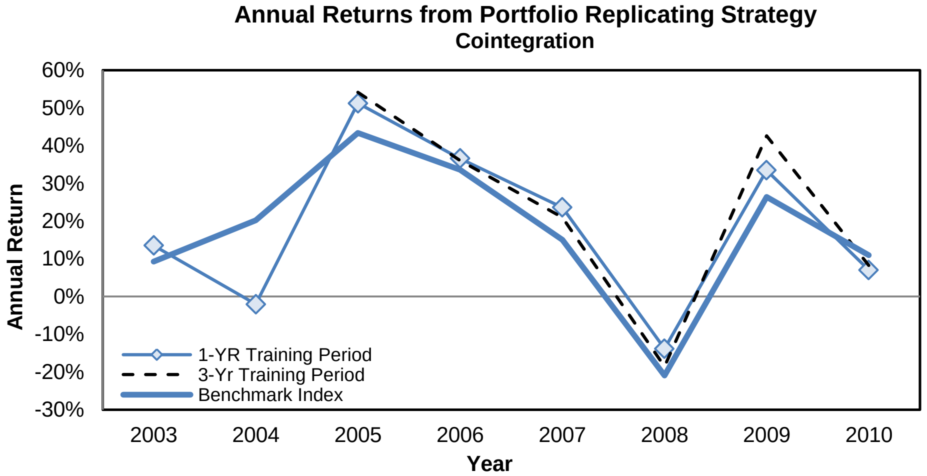


Figure 9: Comparison of the annual performance of the index-replicating portfolio (chosen on the basis of their market capitalisation) using different training periods, but rebalancing the portfolios every two weeks

Figure 9 shows annualised returns for the index-replicating portfolio using the two different training periods and compares them with the performance of the benchmark index over the same periods. In general the index-replicating portfolio shows slightly higher returns than the benchmark index. The performance of the index-tracking portfolios is very similar when using the one-year or the three-year training periods; one does not consistently outperform the other.

3.2.3 Back-test analysis: Testing the different trading strategies

The back-test analysis in this section is performed using the portfolio chosen on the basis of the stocks' Sharpe ratio. The optimisation function specified in Equation (2) is used to estimate the regression coefficients. To allow for long and short positions the constraint on the regression coefficients is $-1 \le \beta_i^* \le 1$ for i = 1,..,n. A three-year training period is used.

Figure 10 shows the estimated portfolio weights over time. In most cases the portfolio weights change slowly over time. Consider for instance how the stock PIK carries an extremely high weight in 2006 to 2008 and then the weight decreases to almost zero from the end of 2008 to 2010. In 2010 the stock PPC carries the highest weight in most of the back-test periods. The stock TRU carries a negative weight, which indicates a short position in the portfolio, in most of the back-test periods. In general the results indicate that the portfolio weights of the long-only trading strategy are much more volatile compared to the portfolio weights when allowing long and short positions.

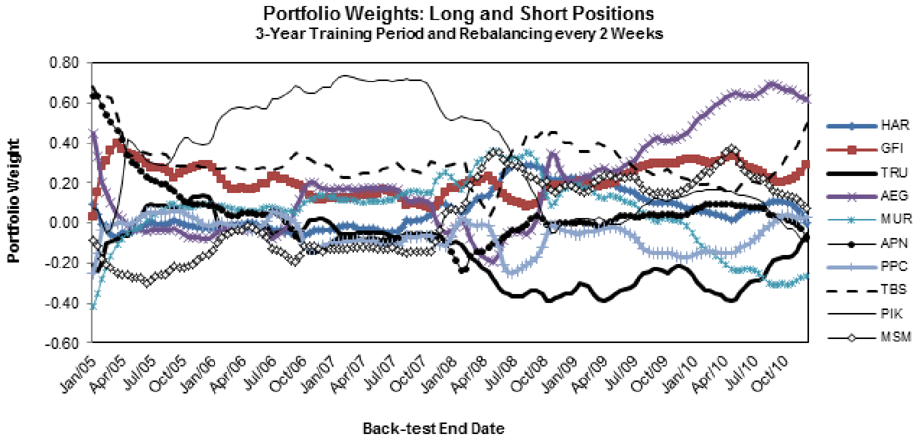


Figure 10: The weights of the index-replicating portfolio with stocks chosen on the basis of the Sharpe ratio and allowing long and short positions

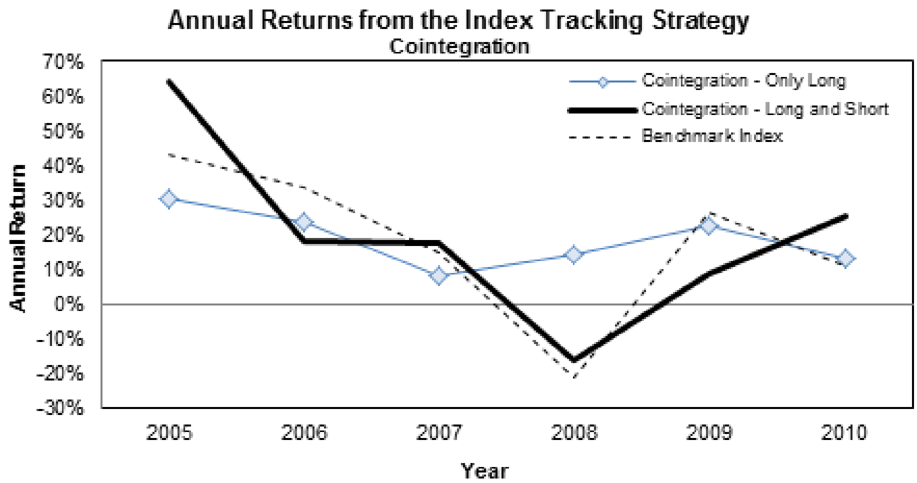


Figure 11: Comparison of the annual performance of the index-replicating portfolio (chosen on the basis of the Sharpe ratio) using a three-year training period and rebalancing every two weeks

Figure 11 compares the performance of the portfolios where only long positions are allowed with the performance of the portfolio where long and short positions are allowed. It is interesting to note that the benchmark index outperforms the long-only portfolio in most years; however, in 2008, where the benchmark index has a negative return, the long-only index-tracking portfolio still shows a relative high positive return.

The portfolio that allows for long and short positions outperforms the benchmark index in four out of the six years. The portfolio also seems to track the performance of the benchmark index more closely, because it also shows the negative return in 2008 that the benchmark index shows. The returns of the long-only portfolio seem to be more stable over time in that the returns fluctuate between 10% and 30%.

3.2.4 Back-test analysis: Comparison of the two sets of stocks

There are some interesting observations when the performance of the portfolio of stocks selected on the basis of their Sharpe ratios is compared with the performance of the portfolio of stocks selected on the basis of their market capitalisations. Figure 12 shows that the performance of the market cap portfolio follows that of the benchmark index more closely (it actually outperforms the benchmark index in almost every period) compared to the portfolio based on the Sharpe ratio, which, as discussed before, shows a more stable return over time. The reason for this may be that the index is weighted by the market capitalisation and that most of the equities included in the market cap portfolio also carry the biggest weights as index constituents.

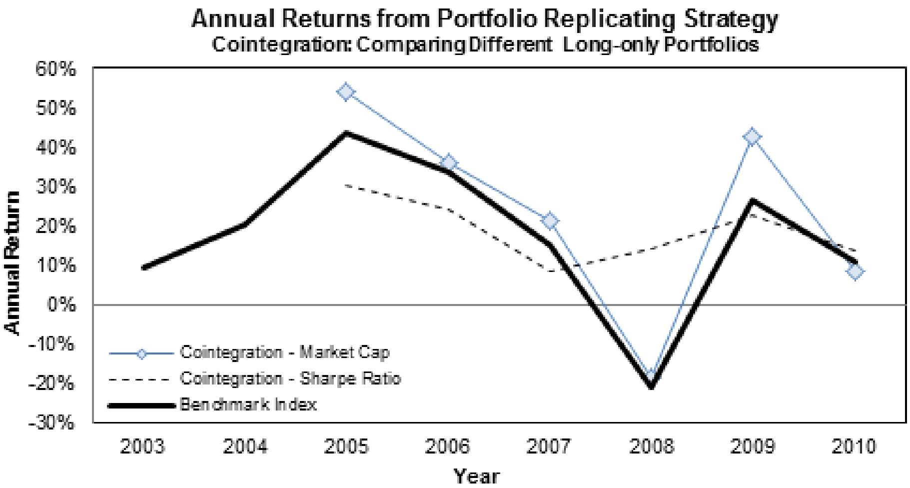


Figure 12: Comparison of the performance of the portfolio of stocks selected on the basis of their Sharpe ratios with the performance of the portfolio of stocks selected based on their market capitalisations.

3.3 Asset allocation using copulas

In this section index-tracking portfolios are constructed using the simulation-based copula approach. Traditional approaches typically use the multivariate Gaussian structure, but there is a lot of research to indicate the inadequacies of that approach (see for instance Rong & Truck, 2010). By using a multivariate copula, it is possible to model the different types of dependence structures more accurately.

3.3.1 Multivariate copula dependence structure

The copula-based asset allocation approach is very computationally intensive, because it entails generating values from a multivariate copula structure for each back-test. A back-test analysis can take a very long time if for every training period the whole copula dependence structure has to be estimated (including choosing which copula fits best) and then to simulate using that dependence structure. To circumvent this issue the historical dependence structures between the various equities were analysed to determine which copulas consistently model the dependence adequately over time.

The first analysis is performed using the portfolio of stocks selected on the basis of the Sharpe ratio of the equities. For each equity the log-return is calculated, and then the Spearman rank correlation matrix is calculated for the returns. The correlation matrix is used to determine the order in which to put the equities through the routine used to derive the multivariate dependence structure. Researchers have shown that the order of the equities is very important and that the fit is optimal when equities that are highly correlated are grouped together (please refer to Heinen & Valdesogo, 2011).

Table 3: Spearman rank correlation derived on the index and equity returns using data from January 2002 to December 2010

	ALSI	HAR	GFI	TRU	AEG	MUR	APN	PPC	TBS	PIK	MSM
ALSI	100%										
HAR	39%	100%									
GFI	37%	76%	100%								
TRU	30%	5%	5%	100%							
AEG	33%	9%	7%	23%	100%						
MUR	33%	11%	11%	24%	45%	100%					
APN	26%	6%	6%	22%	19%	16%	100%				
PPC	33%	10%	7%	22%	28%	23%	19%	100%			
TBS	38%	9%	8%	24%	22%	18%	22%	26%	100%		
PIK	34%	7%	6%	28%	23%	18%	19%	24%	29%	100%	
MSM	26%	5%	4%	33%	22%	21%	20%	24%	22%	28%	100%

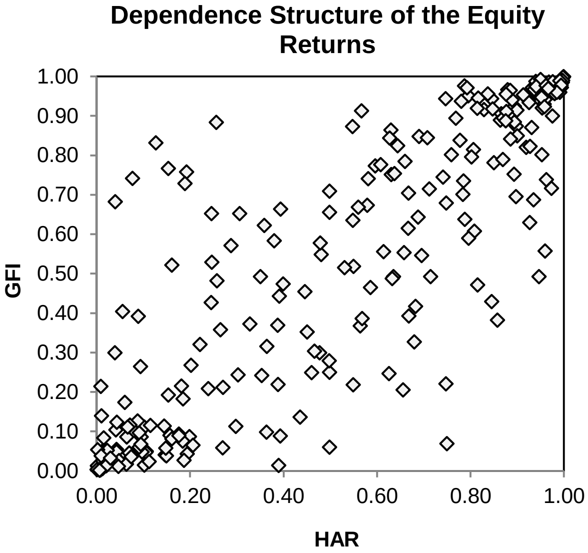
Table 3 shows the long-run correlation estimates derived from the index and equity returns using data from January 2002 to December 2010. All the equities have a long-run correlation of around 25%-40% with the benchmark index. Typically, the highly correlated equities exhibit upper- and lower-tail dependence, whereas the tail dependence is not observable for the other equity pairs that are not as highly correlated.

Figure 13 shows the dependence structure between two pairs of selected equities using data from January to December 2002. The dependence structure of the GFI-HAR pair shows a strong upper- and lower-tail dependence, which indicates that the N14 copula could be considered to model the dependence. The dependence structure of the MUR-AEG pair does not exhibit any dependence over the period. Where the equity pairs do exhibit some dependence, but they do not exhibit tail dependence, the Frank copula can be used. Similar results are obtained over time, which means that the copula to fit can be fixed in the simulation analysis; only the copula parameters have to be updated for each back-test.

Correlations change quite substantially over time depending on market conditions. Figure 14 shows the dependence structures of the GFI-HAR and MUR-AEG equity pairs for the period September 2008 to September 2009. This is generally considered a stress period in the South African market, which means that the correlation between the equities over that period increased artificially. It is interesting to note that the MUR-AEG pair which typically does not exhibit tail dependence, does exhibit tail dependence over the stress period.

The analyses indicate that the same copula can be used to model each of the equity pairs over time, except over a stress period where the increased correlations are observed. It is noteworthy that equities that are relatively highly correlated exhibit lower- and upper-tail dependence, but for the pairs that are not that strongly correlated the tail dependence is not observable.

Increased correlation in stress periods have been discussed by various authors. Consider for instance the work by Longin & Solnik (2001) and Ang & Chen (2002) who find that correlation increases in bear markets but not in bull markets.



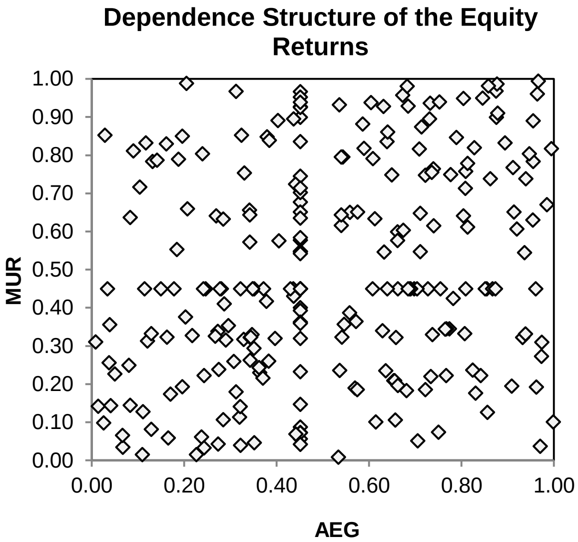
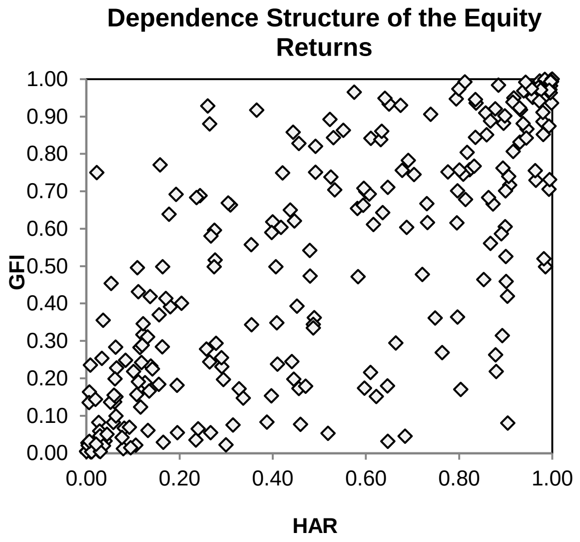


Figure 13: Correlation structures derived from the equity returns for selected equities using data from January to December 2002

Figure 15 shows an example of the d-vine dependence structure fitted to the index and equity returns of the portfolio chosen based on the Sharpe ratio using data from January 2010 to December 2010. The nodes indicate the equity pair to which the copula is fitted, and the estimated parameters are shown under each of the nodes. The p-values of the Kolmogorov-Smirnov test statistic indicate that the copulas at each of the nodes show an acceptable fit (please refer to Table 4).



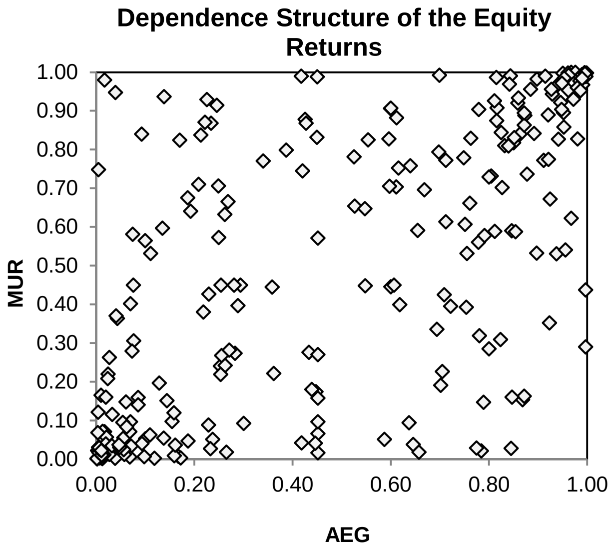


Figure 14: Correlation structures derived from the equity returns for selected equities using data from September 2008 to September 2009

3.3.2 Back-test analysis: Testing the different trading strategies

Figure 16 shows a comparison of the annual performance of the two trading strategies where the annual performance is derived using Equation (6). A training period of one year is used. The figure indicates that the performances of the two strategies are very similar over time. This is somewhat in contrast to the results in Patton (2004) that reported limited gains for the long-only trading strategy.

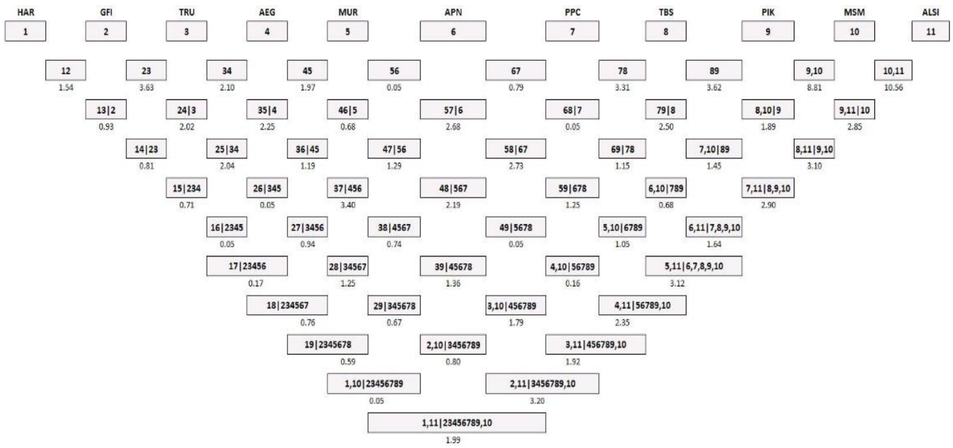


Figure 15: D-vine dependence structure fitted on the benchmark index and equity returns of the portfolio chosen based on the Sharpe ratio using data from January 2010 to December 2010

Table 4: The p-value of the Kolmogorov-Smirnov test statistic used to test whether the copula at each node fits the index and equity returns adequately

Level	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8	Node 9	Node 10
1	0,88	0,93	0,93	0,97	0,99	0,93	0,45	0,93	0,97	0,99
2	0,67	0,97	0,93	0,97	0,97	0,75	0,97	0,88	0,82	
3	0,88	0,93	0,75	0,60	0,97	0,88	1,00	0,75		
4	0,97	0,52	0,75	0,93	0,97	0,99	0,93			
5	0,45	0,99	0,93	0,97	1,00	0,97				
6	1,00	0,93	0,52	0,97	0,93					
7	1,00	0,93	0,99	0,88						
8	0,99	0,97	0,82							
9	1,00	1,00								
10	0,93									

3.3.3 Back-test analysis: Testing the length of the training period

Figure 17 compares the annual performances when using the long-only strategy but training periods of different lengths. In most cases using a shorter training period leads to a better annual performance.

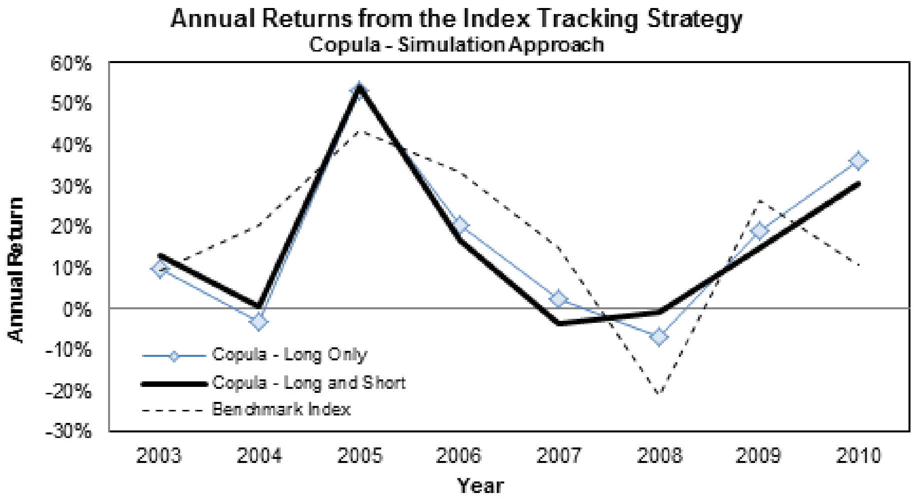


Figure 16: Comparison of the annual performance of the index-replicating portfolio (chosen on the basis of the Sharpe ratio) derived using the copula simulation approach with one-year training periods and rebalancing the portfolios every two weeks

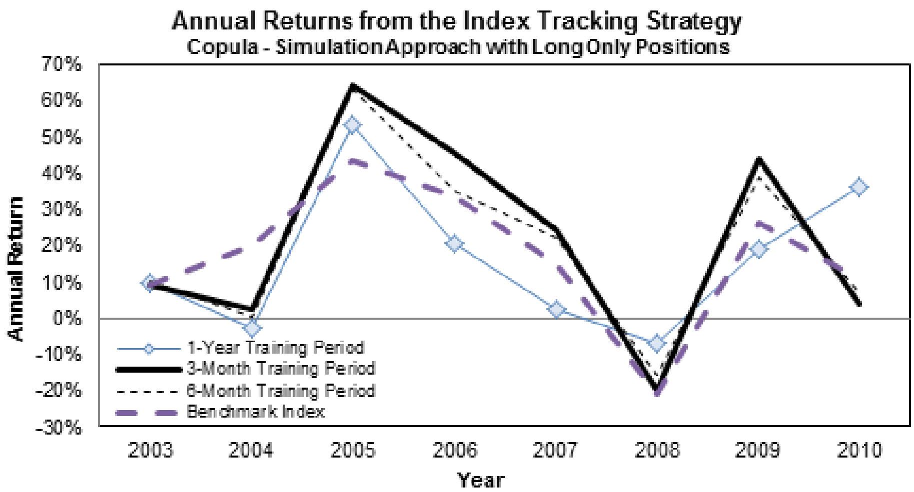


Figure 17: Comparison of the annual performance of the index-replicating portfolio (chosen on the basis of the Sharpe ratio) derived using the copula simulation approach using different training periods and rebalancing every two weeks

3.4 Comparison of the different approaches

There are significant differences in the motivation for the cointegration and copula-based techniques (please refer to Table 5 for a summary). Cointegration techniques focus on modelling the long-term relationship between the index and stock prices. Typically, a longer training period is necessary to estimate the portfolio weights to ensure that the cointegration relationship is captured adequately. The copula technique, on the other hand, makes use of the short-term correlations between the index and stock returns. The portfolio weights are estimated by simulating from the multivariate dependence structure, which means that it is possible to generate sets of returns that have not yet been observed historically, but that still have the correct dependence structures. This presumably improves the estimate of the portfolio weights, as more possible future scenarios are allowed for.

Table 5: Comparison of the philosophy behind cointegration and copula techniques

	Cointegration	Copula
Hypothesis	• The spread between a benchmark index and the stocks in an index-tracking portfolio is mean-reverting over the long-run, which is why the portfolio is expected to track the index more closely over time. • An approach that makes use of long-term relationships.	• Modelling the multivariate dependence structure taking the short-term correlation into account ensures that the time-dependence of the correlation is picked up timeously and should lead to a portfolio whose returns are highly correlated with that of the index. • An approach that make use of short-term market correlations.
Correlation	• A cointegrated relationship does not necessarily imply that the variables are correlated.	• The correlations between the stocks and the index are modelled explicitly.
Training Period	• It is necessary to use longer training periods to ensure that the benchmark index and the index-tracking portfolio are cointegrated. Use a training period of at least three years	• Use shorter training periods to ensure that changes in the dependence structure are picked up sooner in the portfolio weights. The results showed that using a three-month training period produces better results than a one-year training period.
Portfolio Constituents	• The portfolio of stocks must be sufficiently large to ensure a valid cointegrated relationship with the benchmark index.	• The multivariate dependence structure has to be fitted to the portfolio of stocks. At this stage this is a very computationally-intensive process. It should be more optimal instead to use a smaller portfolio of stocks. • The correlations between the stocks were not considered when choosing the portfolio constituents. This resulted in including stocks that are relatively highly correlated with each other, which made the use of a multivariate dependence structure appropriate. • It is not clear how useful copula-based approaches will be when modelling a portfolio where all the stocks are only weakly correlated (to address the issue of multicollinearity).
Portfolio Weights	• Estimated using the log of the benchmark index and equity prices by minimizing the variance and maximizing the stationarity of the tracking error. • Only make use of historically observed sets of prices.	• Minimising the differences between the index returns and the index-tracking portfolio returns. • Values are generated from a multivariate dependence structure. This implies that it is possible to obtain sets of returns that have the correct dependence structure, but have not been historically observed.

A comparison of the annual performances of the portfolio based on the Sharpe ratios and using different asset allocation approaches is shown in Figure 18. The results reflect the long-only trading strategy using a short-term training period of three months for the copula approach and a three year training period for the cointegration approach.

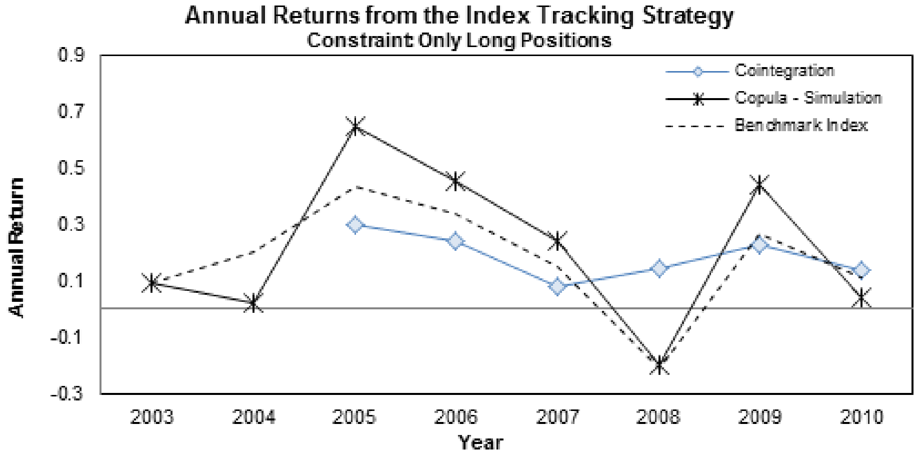


Figure 18: A comparison of the annual performance of the portfolio chosen on the basis of the Sharpe ratio using different asset allocation approaches and allowing long positions only

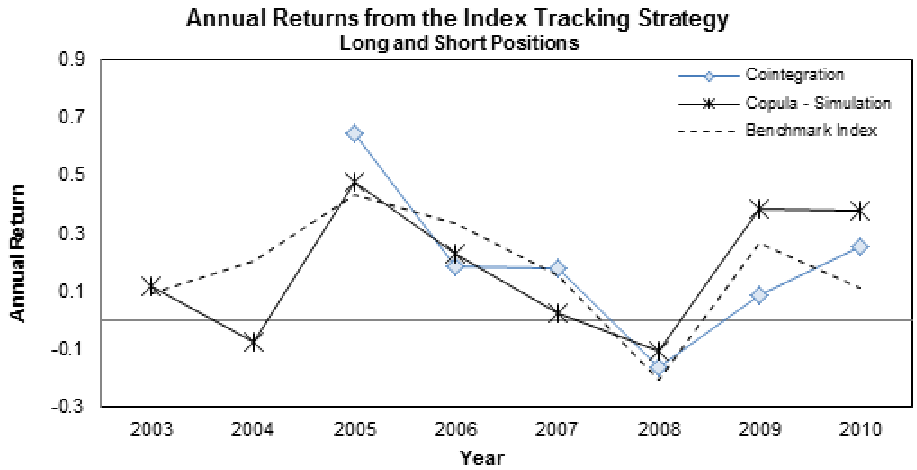


Figure 19: A comparison of the annual performance of the portfolio chosen on the basis of the Sharpe ratio using different asset allocation approaches and allowing long and short positions

The copula approach outperforms the cointegration approach as well as the benchmark index in most of the years. One of the advantages of the cointegration approach is that the performance seems relatively stable over time and it always leads to positive returns, even during the 2008 financial crisis. The results are somewhat different when allowing long and short positions, as can be seen in

Figure 19. The cointegration approach outperforms the copula approach in two of the six years, but also indicates the negative return in 2008 (as was observed for the benchmark index). The copula simulation approach outperforms the benchmark index in five of the eight years.

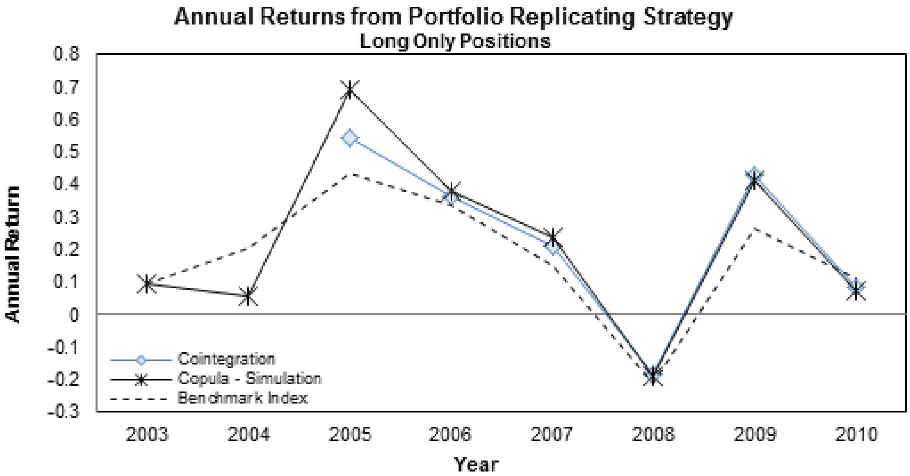


Figure 20: A comparison of the annual performance of the different asset allocation approaches allowing long positions only based on the portfolio chosen on the basis of the market capitalisation of the stocks

A similar analysis is done on the portfolio selected on market capitalisation. Figure 20 shows a comparison of the annual performances of the asset allocation approaches using the long-only trading strategy. A training period of three months was used for the copula-based approach and a three-year training period for the cointegration approach. In most instances the index-tracking portfolios outperform the benchmark index; however, the cointegration and copula strategies show very similar performances since 2006. It is only in 2005 that the copula approach shows a significantly better return.

4. Concluding remarks

The differences between cointegration and copula approaches to asset allocation were explored. It can be argued that the cointegration techniques focus on modelling the long-term relationship between the index and stock prices, whereas copulas make use of the short-term correlations between the index and stock returns.

Two different portfolios were constructed and analysed over time. The first portfolio is based on stocks with the highest market capitalisation. This portfolio was expected to be highly cointegrated and correlated with the benchmark index, because these stocks typically also have the highest weight in the benchmark index. The second portfolio is based on stocks with the highest Sharpe ratios. These stocks are expected to produce the highest returns per unit of risk. In the analyses none of the portfolios consistently outperformed the other.

Two trading strategies were considered. The first trading strategy imposes the constraint that only long positions are allowed, in other words the investor is only allowed to buy stocks and sell stocks already owned. The second trading strategy relaxes this assumption and allows long and short positions. A noteworthy result from the two trading strategies is that the portfolio weights were more stable over time when allowing long and short positions as opposed to long positions only. The implication is that where the portfolio weights are more stable, the transaction costs associated with rebalancing the portfolio are expected to be lower.

The empirical results indicate that the portfolio based on the market capitalisation of the stocks shows similar performance when cointegration or the copula-simulation approach is used. However, based on the Sharpe ratio selection criterion the copula simulation generally outperforms that of the cointegration approach.

In conclusion, the empirical results suggest that it is not necessarily true that modelling long-term relationships between stocks by using the cointegration approach lead to portfolios that outperform correlation-based approaches that are more short-term in nature. The copula approach to asset allocation has been shown to achieve long-term results similar to that of cointegration, even though copula parameters are based on short-term correlations.

References

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009): 'Pair-copula constructions of multiple dependence', *Insurance: Mathematics and Economics*, 44(2): 182-198.

Alexander, C. and Dimitriu, A. (2002): 'The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies', Tech. rept. *ISMA Centre Discussion Papers in Finance Series*.

Alexander, C., and Dimitriu, A. (2005): 'Indexing and statistical arbitrage', *The Journal of Portfolio Management*, 31(2): 50-63.

Alexander, C., Giblin, I. & Weddington, W. (2001): 'Cointegration and asset allocation: A new active hedge fund strategy', Tech. rept. *ISMA Centre Discussion Papers in Finance Series*.

Ang A. & Chen, J. (2002): 'Asymmetric correlations of equity portfolios', *Journal of Financial Economics*, 63(3): 443-494.

Caldeira, J.F. and Moura, G.V. (2012): '*Selection of a Portfolio of Pairs Based on Cointegration: The Brazilian Case'*, Department of Economics, Federal University of Rio Grande do Sul, Preprint submitted to the Economics Bulletin.

Cherubini, U., Luciano, E. & Vecchiato, W. (2004): '*Copula Methods in Finance'*. London: John Wiley & Sons Inc.

Chiu, M.C. & Wong, H.Y. (2012): '*Dynamic Cointegrated Pairs Trading: Time-Consistent Mean-Variance Strategies'*, Research Paper, Department of Statistics, The Chinese University of Hong Kong.

Connor, G. & Lasarte, T. (2004): '*An Introduction to Hedge Fund Strategies'*, Research Report, Hedge Fund Research Programme, Financial Markets Group, London School of Economics.

Engle, R.F. & Granger, C.W.J. (1987): 'Co-integration and error correction: representation, estimation, and testing', *Econometrica: Journal of the Econometric Society*, 55(2): 251-276.

Genest, C. & Rémillard, B. (2005): 'Validity of the Parametric Bootstrap for Goodness-of-fit Testing in Semiparametric Models', *Annales de l'Institut Henri Poincaré: Probabilités et statistiques*, 44:1096-1127.

Genest, C., Rémillard, B. & Beaudoin, D. (2007): 'Goodness-of-fit tests for copulas: A review and a power study', *Insurance Mathematics and Economics*, 44(2):199–213.

Granger, C.W.J. (1981): 'Some properties of time series data and their use in econometric model specification', *Journal of Econometrics*, 16(1):121-130.

Heinen, A. & Valdesogo, A. (2011): 'Dynamic d-vine model', In Kurowicka, D. & Joe, H. (eds), *Dependence modeling: Vine copula handbook*. London: World Scientific Publishing Company.

Hennessy, D.A. & Lapan, H.E. (2002): 'The use of Archimedean copulas to model portfolio allocations', *Mathematical Finance*, 12(2):143-154.

Hong, Y., Tu, J. & Zhou, G. (2007): 'Asymmetries in stock returns: Statistical tests and economic evaluation', Review *of Financial Studies*, 20(5): 1547-1581.

Infanger, G. (2011): '*Dynamic Asset Allocation: Using Stochastic Programming and Stochastic Dynamic Techniques'*, Presentation, Stanford University.

Joe, H. (1997): '*Multivariate Models and Dependence Concepts'*, London: Chapman & Hall/CRC.

JSE. (2012): '*Ground Rules for the Management of the FTSE/JSE Africa Index Series v.2.9*', [online] URL: www.jse.co.za.

LeSage, J.P. (1999): '*Applied econometrics using MATLAB'*, Tech. rept. Department of Economics, University of Toronto.

Longin, F. & Solnik, B. (2001): 'Extreme correlation of international equity markets' *The Journal of Finance*, 56(2): 649-676.

MacKinnon, J.G. (2010): '*Critical Values for Cointegration Tests*', Research Paper, Department of Economics, Queen's University, Canada.

Nelsen, R.B. (2006): '*An Introduction to Copulas*', 2nd edition. USA: Springer Science Business Media.

Patton, A.J. (2002): '*Applications of copula theory in financial econometrics'*, Ph.D. thesis, University of California, San Diego.

Patton, A.J. (2004): 'On the out-of-sample importance of skewness and asymmetric dependence for asset allocation', *Journal of Financial Econometrics*, 2(1):130-168.

Patton, A.J. (2012): 'A review of copula models for economic time series', *Journal of Multivariate Analysis*, 110: 4-18.

Rong, N. & Truck, S. (2010): 'Returns of REITS and stock markets: Measuring dependence and risk', *Journal of Property Investment and Finance*, 28(1): 34-57.

Said, S.E. & Dickey, D.A. (1984): 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika*, 71(3): 599-607.

Sklar, A. (1959): 'Fonctions de répartition à n dimensions et leurs marges', *Publ. Inst. Statist*., 8: 229–231.

Van der Hart, J., Slagter, E. & Van Dijk, D. (2001): '*Stock Selection Strategies in Emerging Markets'*, Tinbergen Institute Discussion Paper, TI 2001-009/4.