International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2015): 6.391

Implementation of Statistical Arbitrage to Retail Sector Using a Two-Stage Correlation and Cointegration Approach

Johannes Tshepiso Tsoku[1], Ntebogang Dinah Moroke[2]

[1]North West University, South Africa, Corner Dr Albert Lithuli and University Drive, Private Bag X 2046, Mmabatho, 2735

[2]North West University, South Africa

Abstract: ***The study implemented the statistical arbitrage technique to retail sector using a two-stage correlation and cointegration approach. The study used a daily time series data ranging from 04 January 2010 to 31 December 2015.The analysis of the study was computed using Eviews 9. The results of correlation analysis revealed that only six pairs of stocks were found to be highly correlated. Of the six preselected pairs, only three pairs were found to be cointegrated. Statistical arbitrage was then implemented to the cointegrated pairs. This strategy revealed a single trade possible for LWHL – LBTI and LWHL-LSAB pair. Therefore, an investor had an opportunity of making profit from the two traded pairs in the retail sector. Recommendations for further studies were formulated from the results of the study.***

Keywords: Correlation, Cointegration, Statistical arbitrage strategy, financial data

1. Introduction

The first working paper on statistical arbitrage was written by Gatev, Goetzmann and Rouwenhorst (1999) and it was published seven years later in the Review of Financial Studies. However, statistical arbitrage had been known many years before 1999 on Wall Street. Hogan, Jarrow, Teo and Warachka (2004) defined statistical arbitrage as "a zero initial cost, self-financing trading strategy with cumulative discounted value." The concept of statistical arbitrage is a generalization of the traditional "riskless" interpretation. Statistical arbitrage identifies mispricing based on deviations from common stochastic trends and relies on a predictive modelling framework that attempts to exploit consistent regularities in the movements of asset prices, unlike the traditional interpretation (Alsayed, 2014).It is a pairs trading strategy that has been used by hedge fund managers, professional traders and institutional speculators. This method exploits market inefficiencies by taking into account two highly correlated pairs of stocks (Perlin, 2009).Elliott, van der Hoek and Malcolm (2005) and Gatev et al. (2006) all consider statistical arbitrage as pairs trading strategy. The main idea behind pairs trading is to take advantage of the temporary mispricing of two indexes that have a common historical movement (Xie and Wu, 2013).

Vidyamurthy (2004) opined that the conception of pairs trading is the simultaneous buying and selling two securities that are historically correlated. Pairs trading strategy requires buying the under-valued stock (long position) while short selling (short position) the over-valued stock, consequently keeping market neutrality. Vidyamurthy (2004) further stated that "pairs trading is said to be a market neutral strategy in its most primitive form which eliminates the effect of market movements when using just two securities, consisting of a long position in one security and a short position in the other, in a predetermined ratio." According to Schmidt (2008), in order to make profit from this relative mispricing, a long position in the stock is opened when its value falls adequately below its long run equilibrium and is closed out once the value of the stock reverts to its expected value. In the same way, investors may earn profit when a stock is trading sufficiently above its equilibrium value by withholding the stock until it reverts to its expected value. The idea of pairs trading is to go long on the underperforming stock $(\mu - \Delta)$ while simultaneously going short on the over performing stock$(\mu + \Delta)$. The trade is closed once the position reverts back to its central point.

The more risky types of arbitrage, such as statistical and volatility arbitrage, have received more attention from an international perspective but this attention is not evident in the context of South Africa. The implementation of statistical arbitrage in the South African context is very limited, therefore this study paves a way in the application of the said technique. Hence, this study implements an innovative strategy on the retail sector trading on Johannesburg Stock Exchange (JSE) to try fill a gap in literature and also to set a platform for other scholars who are interested in this area. Specifically, the study fills a gap by implementing statistical arbitrage pairs trading strategy to stock prices that are highly correlated and cointegrated in South African context. The study strives to benefit scholars who are doing research in the area for the application of statistical arbitrage to index option.

The study is structured as follows. Section 2 presents the theoretical framework. Section 3 briefly outlines the methodological applied in the study. Section 4 presents the results and discussions. Section 5 gives the conclusion.

2. Theoretical Framework

This section presents the theoretical framework, constructs and domain definitions of statistical arbitrage and analysis of financial data. The aspects that form part of the theoretical framework relevant for this study are historical volatility,

correlation analysis, time series stationarity, cointegration technique and statistical arbitrage.

2.1. Historical volatility

Volatility is defined as a statistical measure of dispersion around a mean value; or as the changeability or randomness of the underlying asset (Schwert, 1990). This measure is usually viewed from three different perspectives, namely, historical (backward looking), implied (reflected in an option market price) or future/actual (forward looking). Volatility is further defined by Sewell (2011) as "the standard deviation of the change in value of a financial instrument and is considered a proxy for risk." Therefore, historical volatility basically involves computing the variance or standard deviation of stock returns in the usual way over some historical period.

Volatility can be modelled using the ordinary least squares (OLS), autoregressive conditional heteroscedasticity (ARCH) and generalised autoregressive conditional heteroscedasticity (GARCH)specifications. Chatfield (2004) mentioned classical time series models such as ARCH, ARIMA, GARCH and many other extensions and variations as being frequently used to explore the data and make predictions.The classical time series methods are normally used to describe the dynamic volatility on the financial time series but they do not explore the mean reversion of the paired stock prices. Therefore the use of cointegration through the OLS approach helps in identifying the pair of stocks that move together in the long run and have a characteristic mean reversion.

Historical volatility of the actively traded options on a certain stock can be calculated by analysts and statisticians and the volatility is used to calculate the price of a less actively traded option on the same stock. In historical volatility, only closing prices of stock option are required. In such a scenario, historical volatility is defined as the standard deviation of the daily price index return for a period of time.

2.2. Correlation Analysis

The correlation analysis is used as a preselection criteria. The preselected indexed should be highly correlated. The criteria for preselection of indexes is based on the assumption that all the selected indexes should be highly correlated. The correlation analysis is used to examine the strength of the relationship between two indexes that have price trends. The correlation coefficient value ranges between -1 and +1. The paired indexes are preselected when the correlation coefficient is at least 0.90. The sample Pearson's correlation coefficient(r) of the paired stock indexes (X and Y) is computed using the following equation:

$$r_{XY} = \frac{\sum_i^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{[\sum_i^N (X_i - \bar{X})^2 \sum_i^N (Y_i - \bar{Y})^2]}} \quad (1)$$

where $\bar{X}$ and $\bar{Y}$ are mean prices of stock X and Y respectively.

2.3. Time series stationarity

It is widely known in econometrics that most of the time series data, especially financial time series data, are nonstationary (i.e. the time series contain unit root). As a result, using nonstationary time series could result in bias in the regression results as stated by Yule (1926).The two commonly used types of stationarity are strict stationarity and weak stationarity. A strict stationarity is obtained if a joint and conditional distribution of a process are unchanged if displaced in time. If $Y_t;\ t \in \mathbb{Z}$ is a time series, then a time series is a strict stationary if the distribution of $(Y_{t1}, \dots, Y_{tk})'$ and $(Y_{t1+h}, \dots, Y_{tk+h})'$ are the equal for all k and all $t_1, \dots, k; h \in \mathbb{Z}$. In short, the narration is written as;

$$(Y_{t1}, \dots, Y_{tk})' \overset{\text{def}}{=} (Y_{t1+h}, \dots, Y_{tk+h})'; \quad (2)$$

where $\overset{\text{def}}{=}$ means "equal in distribution".A series is said to be weak stationary if

- $Var(Y_t) < \infty \quad forallt \in \mathbb{Z}$ (3)
- $\mu_Y(t) = \mu forallt \in \mathbb{Z}$ (4)
- $Cov(Y_t, Y_s) = \gamma_Y(r,s) = \gamma_Y(r+t, s+t) forallr, s, t \in \mathbb{Z}$(5)

If the mean (3), variance (4) and covariance (5) of a time series are independent of time, then the series is said to be weakly stationary. The classical examples of formal tests for the presence of unit root include the Phillips-Perron test (Phillips and Perron, 1988), Augmented Dickey-Fuller test (Dickey and Fuller, 1981), the Kwiatkowski-Phillips-Schmidt-Shin test (1992), the Elliot-Rothenberg-Stock point-optimal test (Elliott et al., 1996) and the Ng-Perron test (Ng and Perron, 1995). The study by Song and Chang (2003) revealed that overlooking stationarity resulted in model coefficients being inflated and model overestimation due to the presence of serial correlation.

Wei (2006) highlighted that the problem of unit root can be solved by introducing the logarithms and differencing the series. If a time series becomes stationary after d times of differencing, the process is referred to as an I(d) series.Unit root tests are generally examined for the null hypothesis of unit root. However, in order to model the dynamics of utmost financial time series, Augmented Dickey-Fuller (ADF) (1981) formulated a test that accommodates the more general autoregressive moving average form. In ADF test, it is assumed that error terms are homoscedastic and there is no presence of serial correlation. The Phillips-Perron (PP) (1988) test is a nonparametric test for stationarity. The null hypothesis tested under the PP test is that the series is non-stationary.

Ng-Perron (1995) came up with a test that allows the lag length to be more flexible so that it can changed based on the sample size on condition that error terms follow the general autoregressive moving average process. The method proposed by Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (1992) differs from the ADF and PP tests. The KPSS test assumes stationarity of the series in the null hypothesis. Once the variables have been tested for stationarity, a cointegration technique may be applied. The study test the presence of unit root using ADF test and KPSS test.

2.3.1. Test for stationarity

The study tested each of the preselected index time series for stationarity. This was done not only to allow the use of statistical arbitrage method, but also to establish the appropriateness of cointegration which also forms part of this study. The Augmented Dickey–Fuller (ADF) by (Dickey & Fuller, 1979, 1981) and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) by (Kwiatkowski et al., 1992) tests were applied to the series to determine their order of integration. The null and alternative hypotheses for ADF test are stated as $H_0: \delta = 0$ versus $H_1: \delta < 0$ while the null and alternative hypotheses for KPSS test is given by$H_0: \sigma^2 = 0$ versus $H_1: \sigma^2 > 0$. The ADF test statistic and KPSS test statistic are given by the following equations respectively

$$ADF = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (6)$$

$$LM = \frac{\sum_{i=1}^{T} S_t^2}{\hat{\sigma}_t^2} \quad (7)$$

where $\hat{\gamma}$ and $SE(\hat{\gamma})$ are the cointegrating coefficient and standard errors of the OLS estimate respectively. In equation (7),$\hat{\sigma}_t^2$ represents the long run error variance estimated from the regression of y_t on a constant and time t and S_t is the partial sum of the residuals $\hat{\varepsilon}_t$ from the same regression.The ADF test is a lower-tailed test, so if the p-value of the test statistic is less than the 5% level of significance, then the null hypothesis of unit root is rejected and it can be concluded is that the series does not have a unit root and it is nonstationary.If the p-value of the KPSS test statistic is greater than the 5% level of significant, the null hypothesis of stationarity is rejected in favour of the alternative. After determining the order of integration, the next step is to test whether the preselected indexes share a long run relationship using the Engle and Granger (1987) cointegration technique.

2.4. Cointegration technique

According to Alexander (1999), in the application of statistical arbitrage technique, correlation analysis and cointegration technique are closely related, even though they address different notions. Alexander and Dimitriu (2015) highlighted that the application of cointegration technique to financial econometrics has increased over the years and the technique has been proven to be significantly effective. Highly correlated pairs of stock does not imply that stock prices are cointegrated in a long run. Correlation simply depicts the co-movements in stock prices. On the other hand, cointegration models the long run relationship in stock prices even when the stock prices are not strongly correlated.The concept of cointegration was pioneered by Granger (1981) and Engle and Granger (1987) and it is widely used in stock markets. The cointegration applicability to stock markets was initiated by Lucas (1997) and Alexander (1999).

Engle and Granger cointegration techniques are used to test the statistical relationship between two time series data that are integrated to same order d, $I(d)$, to produce a single time series which is integrated to order $d - b$, where $b > 0$. The Engle and Granger (1987) test incorporates the long run equilibrium theory into the model so that any disequilibrium can be corrected. The Engle and Granger (1987) test is a residual based test used to examine the presence of unit roots on residuals of single regression equation models in order to test the null hypothesis of the absence of cointegration. Another method for cointegration testing was developed by Johansen (1988). The author developed the maximum likelihood test, which is an alternative method of establishing the long run cointegration between variables. This method is based on the vector autoregressive process.Since the study focuses on pairs trading, the Engle and Granger cointegration technique is preferred over Johansen cointegration technique.

2.4.1. Engle and Granger cointegration technique

The Engle and Granger cointegration technique is used to check if the preselected stock pairs share any long term equilibrium relationship. The idea of cointegration assumes that the two stock prices follow a common stochastic trend. The spread between these variables may be weakly stationary. The Engle and Granger (1987) cointegration is computed by employing the ADF test. The long run relationship between two stock prices (P_t^A, P_t^B) is estimated using the following OLS equation:

$$A_t = \beta_0 + \beta_1 B_t + \varepsilon_t \quad (8)$$

where β_0 is a constant and β_1 is the coefficient of the cointegration. The ADF stationarity test is used on the regression residual ε_t to determine whether it has a unit root. The residual series is estimated by the following:

$$\hat{\varepsilon}_t = A_t - \beta_0 + \beta_1 B_t \quad (9)$$

The ADF test result in equation (6) will be compared with the critical value of the ADF test. If the test statistic is less than the critical value or the p-value of the test statistic is less the 1%, 5% and 10% level of significance then the null hypothesis will be rejected. This means that there is no existence of unit root and the paired stock indexes are cointegrated.

2.5. Statistical arbitrage

Statistical arbitrage is a pairs trading that has been used by hedge fund managers, professional traders and institutional speculators. This method exploits market inefficiencies by taking into account two highly correlated pairs of stocks

According to Ehrman (2006), statistical arbitrage is based purely on historical statistical financial data that is utilized in very short term for numerous small positions and it is almost purely model and computer driven when any single trade has very little human analysis.Statistical arbitrage is therefore convincingly described by Lo (2010) as “a highly technical short-term mean-reversion strategy which involves large number of securities, short holding periods, substantial computational models, and trading.” Statistical arbitrage technique is strongly related to the cointegration technique since it depends on mean reversion (Meki, 2012 and Perlin, 2009).

The discovery of pairs trading was in the early 1980s by the quantitative analyst Tartaglia and the team of physicists, computer scientists and mathematicians who had no background in finance. Their discovery was to develop statistical rules that could help in performing arbitrage trades (Gatev et al., 2006). Pairs trading is defined by Ehrman (2006) as “a non-directional, relative-value investment strategy that seeks to identify two companies with similar

characteristics (a pair) whose equity securities are currently trading at a price relationship that is outside their historical trading range." The pairs trading is often referred to as statistical arbitrage strategy or a market neutral trading strategy. The idea behind pairs trading strategy is to identify a pair of stocks prices that exhibit historical co-movement. A position is opened when there is a significant deviation from the historical relationship and close position when there is a convergence in the stocks (Gatev et al., 2006).

The statistical arbitrage strategy was designed to exploit short term deviations from a long run equilibrium between pairs of stocks. The pairs trading strategy used in this study is based on the assumption that a linear combination of stock prices reverts to a long run equilibrium (cointegration) and a trading rule can be constructed to take advantage of the temporary deviations (Chan, 2011).An entry is estimated by using the standard deviation (SD) of the paired stock price ratio. Using the historical stock prices, the entry points are set at ± 2 SD from the one-year moving average. Meaning that when the price ratio is above (below) 2 SD, there are profitable opportunities to trade when going short (long) on the numerator and long (short) on the denominator. An investor may buy (long) the spread at a certain time period t when the relative mispricing reaches -2 SD and sell (short) when spread reaches 2 SD. The position is closed when the spread reaches the central point ($\hat{\mu}$).Figure 1 by Yakop (2011) summarises the implementation of pairs trading strategy. The x-axis represents period of trading and y-axis represents the units of the historical relationship between paired stock prices.

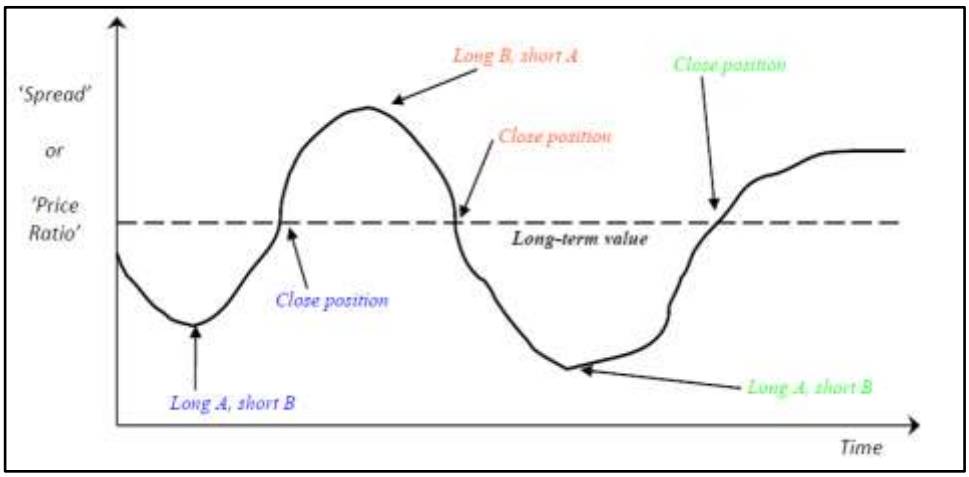

Figure 1: Graphical representation of pairs trading strategy

3. Data

The current study used daily time series data ranging from 04 January 2010 to 31 December 2015 consisting of 1500 observations. One year is approximated to have 250 trading days. The data did not include weekends and public holidays since the stock markets do not operate on the given days, therefore the sample period exclude these days. The data used in this study comprises of closing prices of option stock for retail sector traded on JSE.The stock market data together with the stock code provided by JSE is presented in Table1.

Table 1: Variables in the retail sector

Variables
The Foschini Group (TFG)
Aspen Pharmacare Holdings (APN)
British American Tobacco PLC (BTI)
The Spar Group (TSG)
Woolworths Holdings (WHL)
Truworths International (TRU)
Tiger Brands (TBS)
Shoprite (SHP)
SABMiller (SAB)
Pick'n Pay Stores (PIK)

4. Characteristics of the financial time series data

The preliminary data analysis will be computed to describe the characteristics of the time series data used in the study. Financial time series is characterised by stylised facts. The most common stylised facts are no presence of autocorrelation (also referred to as serial correlation), flat tails and gain/loss asymmetry. Serial correlation is tested using the following equations:

$$Q = n(n+2)\sum_{k=1}^{z}\frac{\hat{\rho}_k^2}{n-k} \qquad (10)$$

where Q is the Ljung-Box test statistic,n denotes the sample size, $\hat{\rho}_k$ is the serial correlation at lag k, and z is the number of lags being tested. The asymmetric distribution is tested using the following equationThe Jarque-Bera (JB) (1980) normality test:

$$JB = n\left[\frac{s^2}{6}+\frac{(k-3)^2}{24}\right] \qquad (11)$$

where

$$s = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{3/2}} \qquad (12)$$

and

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} \quad (13)$$

Under the assumption for normality, the coefficient of skewness and excess kurtosis are expected to be 0 and 3 respectively. The null hypothesis is rejected if the p-value is less than 5% level of significance.According to Sewell (2011), the serial correlation of log returns/stock prices is generally not significant.Kat (2003), Ling (2006) and Ang and Chen (2002) are of the view that most of the stocks or assets tend to have a negative skewness or excess kurtosis.

5. Results and Discussion

This section presents data analysis and provides interpretation of results. The analysis is summarised using tables and graphs.

5.1. Preliminary data analysis results

Descriptive measures gives clear understanding of the characteristics of the data used in the study. Stylised facts of the study are derived from the descriptive measures presented in Table 2.

Table 2: Descriptive statistics for log transformed data

Index	LAPN	LBTI	LPIK	LSAB	LSHP	LTBS	LTFG	LTRU	LTSG	LWHL
Mean	9.720	10.691	8.455	10.576	9.509	10.146	9.271	8.954	9.371	8.551
Median	9.697	10.729	8.414	10.597	9.610	10.221	9.301	8.971	9.365	8.767
Max	10.699	11.377	8.836	11.478	9.935	10.595	9.898	9.359	9.926	9.277
Min	8.780	10.009	8.169	9.893	8.773	9.729	8.594	8.328	8.835	7.467
Std. Dev.	0.597	0.373	0.150	0.430	0.278	0.228	0.271	0.189	0.266	0.504
Skewness	0.051	-0.208	0.620	-0.008	-0.798	-0.416	-0.216	-0.889	0.325	-0.481
Kurtosis	1.458	1.768	2.523	1.641	2.430	1.894	2.619	3.949	2.588	1.841
JB test	149.157	105.755	110.241	115.387	179.515	119.816	20.752	253.731	37.028	141.810
Prob.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LB (15)	22167	21825	20347	21787	21355	21529	21329	20022	21512	21725
Prob.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Obs	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500

The highest mean value occurs in British American Tobacco PLC (LBTI) (10.691), while Pick'n Pay Stores (LPIK) has the lowest mean value of 8.455. LBTI, SABMiller (LSAB), Shoprite (LSHP), Tiger Brands (LTBS), The Foschini Group (LTFG), Truworths International (LTRU) and Woolworths Holdings (LWHL) are skewed to the left while Aspen Pharmacare Holdings (LAPN), LPIK and The Spar Group (LTSG) are skewed to the right. The kurtosis value for all the variables except LTRU exceeds the threshold of 3, implying that the variables are platykurtic. Leptokurtosis distribution is only observed in LTRU only. The data also reveals that all the variables reject the hypothesis of normality. Therefore, it is concluded that all the variables are not normally distributed. The p-values of the LB test statistics indicate that all the variables have autocorrelation. The results are in line with the views by Kat (2003), Ling (2006) and Ang and Chen (2002).

5.2. Plot of the retail sector

The following Figure2 presents the graphical presentation of the retail sector.

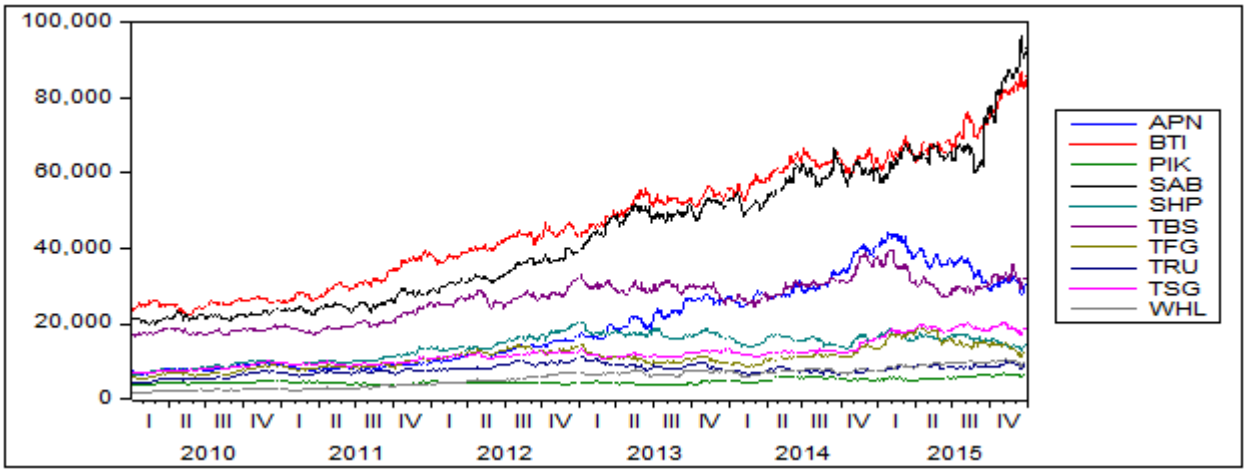

Figure 2: Time series plot historical volatility of retail sector at level

The visual inspection of Figure 2 depicts that the retail sector is not stationary at level. The graph of BTI and SAB seem to be trending upwards from the start to the end of the sample period. The graph of APN and TBS have a similar pattern from fourth quarter of 2013 until fourth quarter of 2015. Other variables seem to be stable throughout the sample period. It is therefore concluded that all the variables are nonstationary. The data as a result conforms to one of the stylised facts of financial time series called lack of stationarity. The log differenced retail sector data is presented in Figure 3.

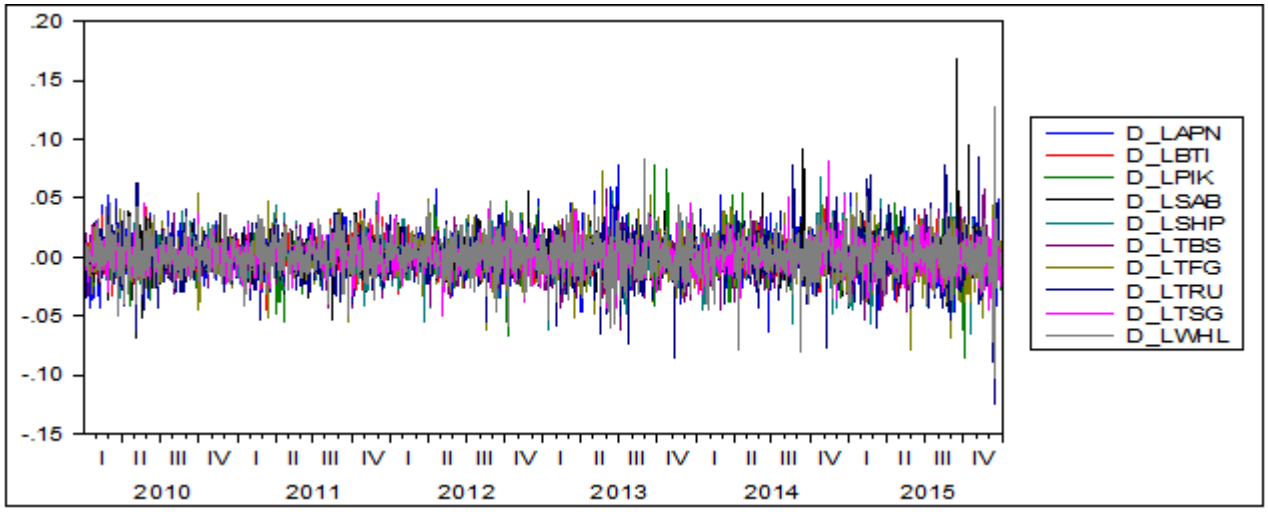

Figure 3: Time series plot of retail sector at logarithm difference

The first log differenced retail sector presented in Figure 3 depicts that all the variables under retail sector are stationary. The formal tests for stationarity will be computed to confirm the graphical presentation. The following section present the correlation analysis prior to the unit root test.

5.3. Correlation analysis results

The following Table 3 presents the results of the correlation analysis of the variables in the retail sector.

Table 3: Correlation analysis of the retail sector

Correlation	APN	BTI	PIK	SAB	SHP	TBS	TFG	TRU	TSG	WHL
APN	1.000									
BTI	0.918	1.000								
PIK	0.707	0.752	1.000							
SAB	0.914	0.988	0.770	1.000						
SHP	0.677	0.737	0.315	0.687	1.000					
TBS	0.846	0.869	0.523	0.842	0.873	1.000				
TFG	0.729	0.744	0.548	0.705	0.745	0.797	1.000			
TRU	0.385	0.540	0.214	0.501	0.833	0.669	0.741	1.000		
TSG	0.854	0.895	0.761	0.884	0.659	0.779	0.887	0.589	1.000	
WHL	0.898	0.969	0.690	0.954	0.834	0.893	0.811	0.663	0.909	1.000

From Table 3, only six pairs of stocks were selected and used for further analyses. The pairs are BTI – APN, SAB – APN, SAB – BTI, WHL – BTI, WHL – SAB and WHL – TSG.The pairs were selected using the cut-off point of 0.9. The preselected pairs were further tested for stationarity and the results are presented in Table 4.

5.4. Unit root test results

The five variables making up the six preselected paired were examined for the presence of unit root. The results are summarised in Table 4.

Table 4: Unit root tests results of the five preselected variables in the retail sector

Index	ADF test statistic	P-value	KPSS test statistic	P-value
APN	-0.745	0.833	4.443	0.000
LAPN	-0.906	0.787	4.673	0.000
ΔLAPN	**-38.865**	**0.000**	**0.162**	0.021
BTI	1.129	0.998	4.722	0.000
LBTI	-0.254	0.929	4.727	0.000
ΔLBTI	**-28.942**	**0.000**	**0.034**	0.003
SAB	1.789	1.000	4.625	0.000
LSAB	0.370	0.982	4.761	0.000
ΔLSAB	**-30.199**	**0.000**	**0.122**	0.009
TSG	-0.674	0.851	3.863	0.000
LTSG	-0.995	0.757	4.083	0.000
ΔLTSG	**-21.721**	**0.000**	**0.052**	**0.065**
WHL	-0.682	0.849	4.593	0.000
LWHL	-1.699	0.432	4.515	0.000
ΔLWHL	**-23.876**	**0.000**	**0.148**	0.010

The results in Table 4 depicts that the original data and the log transformed data is nonstationary.The p-values of ADF test are all less than 5% level of significance. Therefore, it is concluded that all the variables are stationary after first log difference. The p-values for KPSS test also revealed that only ΔLTSG is stationary at log difference while all the other variables are nonstationary. Since there is a contradiction between two tests of unit root, the conclusion of the study is based on at least one test which is the ADF test and it is concluded that all the variables are stationary at first log difference. This means that all the variables are integrated to order 1, *I(1)*.

5.5. Cointegration results

The Engle and Granger cointegration test was applied to the six preselected pairs. The results are presented in Table 5.

Table 5: Engle and Granger cointegration results for LBTI – LAPN, LSAB – LAPN, LSAB – LBTI, LWHL – LBTI, LWHL – LSAB and LWHL – LTSG

	OLS Results		Residuals from OLS	
Pairs	**β_1 Coefficient**	**t-statistic**	**ADF (t-statistic)**	**P-value**
LBTI – LAPN	0.594	119.811	-1.489	0.539
LSAB – LAPN	0.696	145.103	-1.309	0.627
LSAB – LBTI	1.140	244.843	-3.858	**0.002****
LWHL – LBTI	1.316	166.503	-3.680	**0.005****
LWHL – LSAB	1.125	134.741	-3.257	**0.017****
LWHL – LTSG	1.720	84.323	-1.876	0.344

*Note: *, **, and *** indicates the MacKinnon critical values at 1%, 5% and 10% levels are -3.435, -2.863, and -2.568 respectively. Lag length was selected by SIC.*

The results presented in Table 5 revealed that only three pairs are cointegrated since their p-values of the ADF test statistic are significant at 5% level of significance. The three pairs are used for in the implementation of statistical arbitrage pairs trading strategy.

5.6. The implementation of statistical arbitrage strategy

The first step in implementing pairs trading strategy is to determine whether or not the paired stocks are historically moving together. The idea of pairs trading is to buy the relatively undervalued stock and sell the relatively overvalued stock, then close trade when the ratio goes back to the mean ration (line 0). The cointegrated pairs are presented in Figure 3.

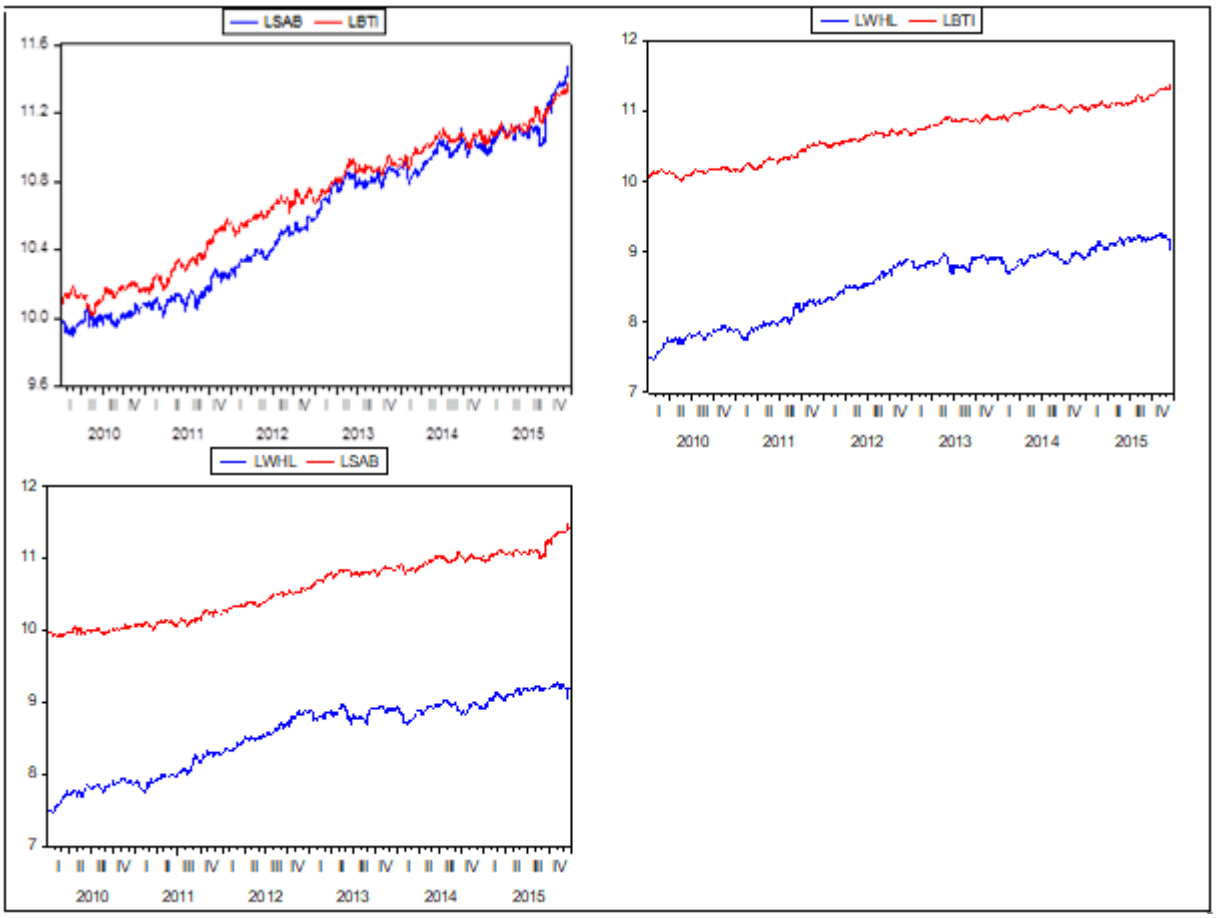

Figure 4: Historical movement of LSAB – LBTI, LWHL – LBTI and LWHL – LSAB pair

Figure 4 depicts the historic price movement of LSAB – LBTI, LWHL – LBTI and LWHL – LSAB pair. The price movement of the pairs seem to have same pattern throughout the sample period. The figure reveals a close relationship between LSAB – LBTI. The historic price movement of LWHL – LBTI and LWHL – LSAB seem to also have a similar pattern throughout the sample period. This historical movement in the pairs is important for the implementation of pairs trading strategy. This implies that the paired variables are related and they can be used in pairs trading strategy.

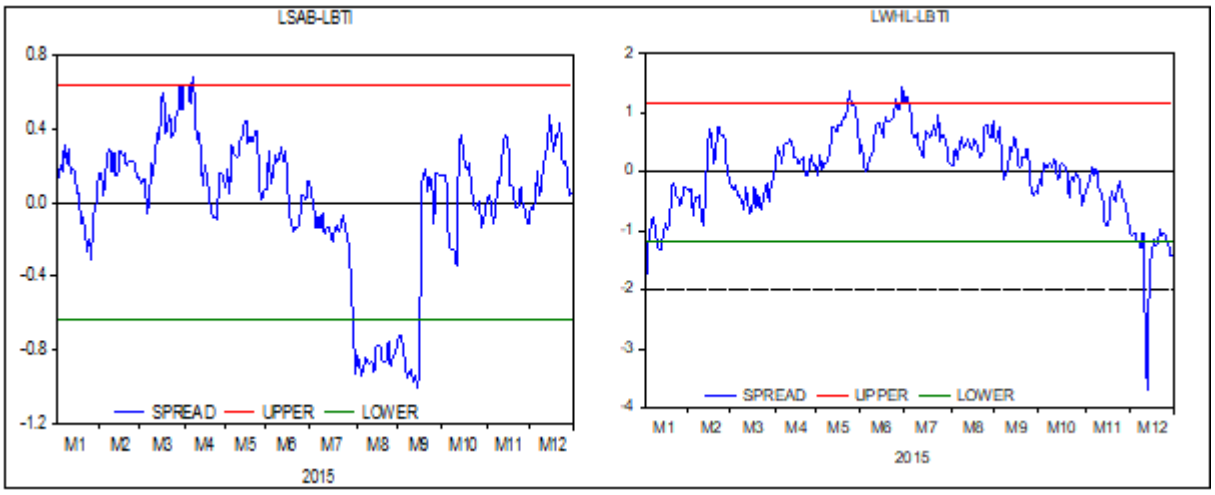

Figure 5: The residual spread series of LSAB – LBTI and LWHL – LBTI pairs with trading thresholds

The spread of LSAB – LBTI implies that trade was not possible for 2015 since the spread of LWHL – LBTI pair ranges between approximately -4.0 and 1.2 standard deviation. Trade was only possible towards the twelfth month in the LWHL – LBTI pair. This means that LWHL was underperforming relative to LBTI. Therefore, an investor would have bought the units of LWHL and sold the units of LBTI. Implicitly, the investor would have chosen to go long on LWHL while simultaneously going short on LBTI. Trade was closed when relationship returns to its statistical normal.

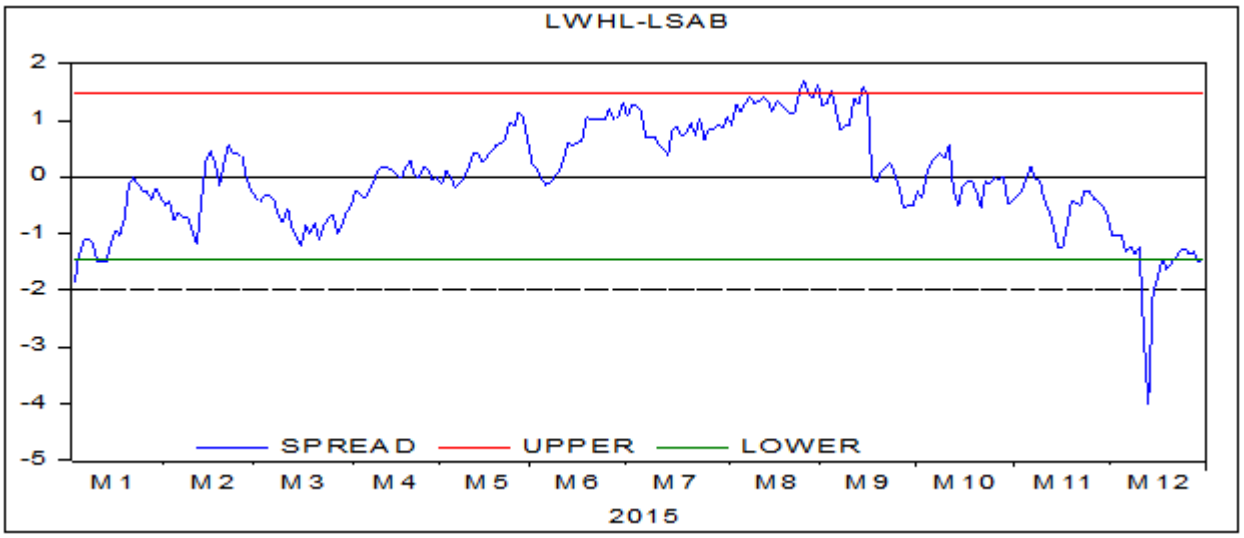

Figure 6: The residual series of LWHL – LSAB pair with trading thresholds

The spread of LWHL – LSAB pair ranges between -4.0 and 1.8 standard deviation, implying that single trade was only possible towards the twelfth month when the ratio reached -2 standard deviation. This also means that LWHL was underperforming relative to LSAB. Therefore, an investor had a good opportunity to buy units of LWHL and sell units of LSAB.

6. Conclusions

The study implemented the statistical arbitrage technique to retail sector using a two-stage correlation and cointegration approach. Daily time series data ranging from 04 January 2010 to 31 December 2015. The potential pairs were selected using correlation analysis and cointegration technique. The study further provided a base for future researchers conducting studies on emerging markets, more specifically in South African context.Out of ten pairs in the retail sector, only six pairs were preselected using Pearson correlation analysis. The preselected pairs were tested for unit root using ADF and KPSS test.Cointegration test revealed that only three pairs were found to be cointegrated and they were used for further analysis.A two-step approach was successfully used to select the pairs within the retail sector.The study used the 2 standard deviation trading rules as previously used by Gatev et al. (2006), Schroder and Smith (2001) and Ruiter (2011).

The implementation of statistical arbitrage revealed that there was only one trade possible for LWHL – LBTI pair. An investor would have chosen to go long on LWHL while simultaneously going short on LBTI. An investor would buy more of LWHL and sell more of LBTI. There was also a single trade possible for LWHL-LSAB pair. This means that an investor had a good opportunity of buying more units of LWHL and selling more units of LSAB. Therefore, there is a possibility of making profit from the traded pairs. The study successfully implemented statistical arbitrage strategy to

retail sector trading on JSE using correlation analysis and cointegration approach.

The study recommends that a similar study could be reproduced using the proposed methodology to high frequency data in order to provide more accurate results. Using high frequency data may help increase possible trade in a given day and may lead to having increased number of trades.Further study could also make use of the copula approach in implementing pairs trading strategy. The strategy may help in examining the dependency between the stock prices in order to increase the possibilities of trade.

References

[1] Alexander, C.O. (1999). Optimal hedging using cointegration. Philosophical Transaction of the Royal Society Series A, Vol. 357, pp. 2039–2058.
[2] Alexander, C.O and Dimitriu, A. (2005). Indexing and statistical arbitrage: Tracking error or cointegration? The Journal of Portfolio Management, pp. 15.
[3] Alsayed. H. (2014). Essay in Statistical Arbitrage. Financial Economics. University of Southampton Research Repository. ePrint Soton. Unpublished thesis.
[4] Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. Review of Financial Studies, Vol. 63, pp. 443-494.
[5] Bondarenko, O. (2003). Statistical arbitrage and security prices. Review of Financial Studies, Vol. 16, pp. 875–919.
[6] Chan, N.H. (2011). Time series: Applications to finance with R and S-plus. John Wiley & Sons.
[7] Chatfield, C. (2004). The analysis of time series: An introduction. 6th edition. New York: Chapman and Hall.
[8] Dickey, D.A. and Fuller, W.A. (1979). Distributions of the Estimators for Autoregressive Time Series with a Unit Root. Journal of American Statistical Association, Vol. 74, pp. 427-481.
[9] Dickey, D.A. and Fuller, W.A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root. Econometrica, Vol. 49, pp. 1057-1072.
[10] Do, B., Faff, R. and Hamza, K. (2006). A new approach to modelling and estimation for estimation for pairs trading. Working paper. Monash University.
[11] Ehrman, D.S. (2006). The handbook of pairs trading: Strategies using equities, options, and futures. Wiley. Vol. 240.
[12] Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. Econometrica, Vol. 64, 4, pp. 813-836.
[13] Elliott, R.J., van der Hoek, J. and Malcolm, W.P. (2005). Pairs trading. Quantitative Finance, Vol. 5, 3, pp. 271–276.
[14] Engle, R.F. and Granger, C.W.J. (1987). Cointegration and error correction: Representation, estimation, and testing. Econometrica, Vol. 55, 2, pp. 251–276.
[15] Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G. (1999). Pairs trading: Performance of a relative value arbitrage rule. Yale School of Management Working Papers ysm3, pp. 1–34.
[16] Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies, Vol. 19, 3, pp. 797–827.
[17] Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. Journal of Econometrics, pp. 121-130.
[18] Hogan, S., Jarrow, R., Teo, M. and Warachka, M. (2004). Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. Journal of Financial Economics, Vol. 73, pp. 525-565.
[19] Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. European Journal of Operational Research, Vol. 196, 2, pp. 819-825.
[20] Huck, N. (2010). Pairs trading and outranking: The multi-step-ahead forecasting case. European Journal of Operational Research, Vol. 207, pp. 1702–1716.
[21] Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. Journal of Economic Dynamics and Control, Vol. 12, pp. 231-254.
[22] Kat, H.M. (2003). The dangers of using correlation to measure dependence. The Journal of Alternative Investments, Vol. 6, 2, pp. 54-58.
[23] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How sure are we that economic time series have a unit root? Journal of Econometrics, Vol. 54, 1-3, pp. 159-178.
[24] Ling, H. (2006). Dependence patterns across financial markets: A mixed Copula approach. Applied Financial Economics, Vol. 16, 10, pp. 717-729.
[25] Lo, A.W. (2010). Hedge funds: An analytic perspective. Princeton, New Jersey: Princeton University Press.
[26] Lucas, A. (1997). Strategic and tactical asset allocation and the effect of long-run equilibrium relations. Series Research Memoranda 0042. VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.
[27] Meki, B. (2012). Examining long-run relationships of the BRICS stock market indices to identify opportunities for implementation of statistical arbitrage strategies. Unpublished thesis.
[28] Ng, S. and Perron, P. (1995). Unit root tests in ARMA Models with data dependent methods for the selection of the truncation lag. Journal of the American Statistical Association, Vol. 90, pp. 268-81.
[29] Perlin, M.S. (2009). Evaluation of Pairs Trading Strategy at the Brazilian Financial Market. Journal of Derivatives & Hedge Funds, Vol. 15, pp. 122-136.
[30] Phillips, P.C. and Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, Vol. 75, 2, pp. 335-346.
[31] Ruiter, H.J. (2011). The Performance of a Pairs Trading Strategy in Asian Markets for 2002 to 2009. Unpublished thesis.
[32] Sewell, M. (2011). Characterization of financial time series. Research note rn/11/01, UCL, Department of Computer Science.
[33] Schmidt, A.D. (2008). Pairs trading: A cointegration approach. University of Sydney. Unpublished thesis.
[34] Schroder, S. and Smith, B. (2001). Pairs Trading: A Way to Profit in a Volatile Equity Market. Equity Research Europe.

[35] Schwert, W.G. (1990). Stock Market Volatility. Financial Analysts Journal, Vol. 46, 3, pp. 23-34.
[36] Song, N. and Chang, S.J. (2003). Nonstationarity and its consequences in modeling the southern timber market. Bugs, budgets, mergers, and fire: Disturbance economics, pp. 241-249.
[37] Vidyamurthy, G. (2004). Pairs Trading, Quantitative Methods and Analysis. Canada: John Wiley & Sons.
[38] Wei, W.S. (2006). Time series analysis: Univariate and multivariate. Boston: Pearson.
[39] Xie, W. and Wu, Y. (2013). Copula-based Pairs Trading Strategy. In Asian Finance Association (AsFA) 2013 Conference.
[40] Yakop, M. (2011). A comparative analysis of pairs trading. University of Amsterdam. Unpublished thesis.
[41] Yule, U.G. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. Journal of the Royal Statistical Society, Vol. 89, 1, pp. 1-63.