Comparison of Statistical Arbitrage in Developed and Emerging Markets

Gabriel Visagie and Alwyn Hoffman

Abstract—**Statistical arbitrage covers a variety of trading strategies that are based on statistical modelling and are usually characterized by a near market-neutral trading book. Statistical arbitrage strategies are generally used by hedge funds and retail traders with the objective of achieving excess returns regardless of the market regime. In this study we examine the performance of two statistical arbitrage systems in developed markets (Germany, Japan and the United States) and emerging markets (China and South Africa) over the period of 2006 to 2016. We investigate whether the performance of these systems are affected by different market regimes and whether excess returns have been altered since the financial crisis in 2008-2009. We also compare the performance of these systems in developed and emerging markets. Our results indicate that statistical arbitrage systems produce higher excess returns during non-trending markets and that these systems generally provide higher returns in emerging markets. The overall performance of the statistical arbitrage systems have weakened since the financial crisis, but a noticeable improvement in excess returns produced is apparent in the last three years of our study.**

Index Terms—**Financial modelling, statistical arbitrage, algorithmic trading, pairs trading.**

I. Introduction

Background

In finance, arbitrage is the practice of exploiting a difference in price between two or more markets. The execution of arbitrage relies on the placement of a combination of trades that capitalizes on the difference between market prices. An arbitrage can be considered a transaction that does not involve negative cash flow at any temporal or probabilistic state and a positive cash flow in at least one of these states. Arbitrage in theory is very intriguing as it can provide a risk-free profit at zero cost.

In finance, statistical arbitrage covers a variety of trading strategies that are based on statistical modelling. These strategies are characterized by a degree of market neutrality such that an investment portfolio based on statistical arbitrage is only slightly affected by movements in the overall financial market [1]. These strategies thus aim to provide excess returns above the market index return (alpha) while simultaneously achieving these profits at lower volatility than the market index (beta). Many statistical arbitrage strategies are based on the concept of mean-reversion in the price series of financial

securities. Typical forms of statistical arbitrage include pairs trading and general long/short strategies. Statistical arbitrage is often employed by hedge funds to produce low volatility investment strategies that inherently take advantage of diversification across assets.

There is no single definition of high frequency trading (HFT), but it is generally described as a form of algorithmic and quantitative trading that is characterized by short holding periods and high order-to-trade ratios. This type of trading is made possible with the availability of high frequency financial data and generally relies on the use of sophisticated mathematical models and computing methods. High frequency traders move in and out of short-term positions at high volumes to capture small profits which may exist because of inefficiencies in the financial markets. HFT strategies generally do not allow for holding positions over night. The potential of achieving low volatility and high excess returns offered by this type of trading, makes it an attractive approach to traders and investors [2]. Some literature however argues that HFT and electronic trading pose new challenges to the financial system and that these methods lead to an increase in market volatility [3].

In this paper we investigate a novel form of statistical arbitrage that combines the formation of artificially created mean reverting instruments, based on baskets of securities selected through clustering techniques, with dynamic thresholds for entering trades.

II. Literature Review

A. Efficient Market Hypothesis

The efficient market hypothesis (EMH) was introduced by Fama in the 1970's [4]. Fama hypothesized that markets are efficient as news spreads quickly, without delay, to be reflected in the prices of financial securities. If the efficient market hypothesis holds true, an investor cannot – using any techniques – pick certain securities that would allow for greater returns than would be achieved by randomly selecting a portfolio of individual securities that exhibits comparable risk. The efficient market hypothesis is consequently associated with the concept of a random walk model. Random walk models are used to describe price series where each subsequent price change represents a random change from the preceding price change.

Some studies suggest that the prices of securities are at least partially predictable. Malkiel [5] argues that markets cannot be completely efficient as the collective judgment of investors are fundamentally bound to make mistakes. Malkiel further states that it can be expected for some market participants to

act irrationally at times. The study by Malkiel suggests that markets are not entirely efficient, but that efficiency has improved noticeably over time.

Grossman and Stiglitz [6] argue that if markets are perfectly efficient, there will be no incentive for professionals to uncover information that gets rapidly reflected in stock market prices.

A large number of empirical studies conclude that security prices contradict the efficient market hypothesis. Jegadeesh and Titman [7] examined a trading strategy in 1993 that buys well-performing stocks and sells poor-performing stocks. The results conclude that the strategy was able to achieve excess returns of up to 12% above the returns provided by the standard capital asset pricing model (CAPM).

A study by Chan, Jegadeesh and Lakonishok [8] in 1996 investigated the predictability of future returns by examining past returns. The study suggests that there is little evidence of subsequent reversals in the returns of stocks that exhibit high price and earnings momentum. Chan, Jegadeesh and Lakonishok argue that markets only gradually respond to new information.

Dunis and Ho [9] examined long-short market neutral strategies on the Dow Jones EUROStoxx 50 index from Januray 2002 to June 2003. The results suggest that market neutral strategies can generate steady returns under adverse market circumstances.

Nobrega and Oliveira [10] investigated the effect of various machine learning models on statistical arbitrage in 2013. The results suggest that some of the models examined were significantly profitable such as the extreme learning machine which was able to achieve an average annual return of 23.58% in out-of-sample data.

B. Established Models for Statistical Arbitrage

Due to the nature of trading strategies, it can be assumed that many statistical arbitrage techniques are not in the public domain. A number of approaches have however been documented in academic literature, some of which will now be discussed.

C. Minimum Distance Method

Gatev, Goetzmann and Rouwenhorst [11] examined and documented the minimum distance method. The application of this method consists of two distinguishable periods: a pair formation period and a trading period. The length of these periods are essentially free parameters and Gatev, Goetzmann and Rouwenhorst mention that the decision to choose a 12 months formation period and a 6 months trading period in their study, was arbitrary.

During the formation period, pairs of securities are chosen such that the sum of squared deviations (Euclidean distance) between the two normalized price series are minimized. Gatev, Goetzmann and Rouwenhorst placed the restriction on pairs that securities belonging to a pair must be from the same stock market sector. Once pairs have been created from the given security universe, a selection of the pairs with the minimum squared distance is chosen. The standard deviation of the spread of each of the selected pairs is calculated over the formation period.

During the trading period, a simple standard deviation-based trading strategy is executed. A long position is taken in the lower-priced security and a short position is taken in the higher priced security when the spread of the normalized prices have diverged by more than two standard deviations. The long and short positions are opened with equal monetary value. The positions are closed when the prices cross. When the trading period ends, all positions are closed regardless of whether prices have converged or not.

D. Arbitrage Pricing Theory

Vidyamuthy [12] suggested a method of detecting potential pairs by making use of arbitrage pricing theory (APT) which was first suggested by Ross [13] to determine asset prices. By making the assumptions that the law of one price holds and market participants have homogenous expectations, the return on any financial security is linearly related to a set of risk factors:

$$r_i = r_f + \sum_{j=1}^{k} \beta_{i,j} r_j^* + \varepsilon_i \tag{1}$$

where $\beta_{i,j}$ is the risk exposure of asset i to risk factor j, r_j^* is the return contribution of risk factor j and r_f is the risk-free return. The residual, ε_i, can be interpreted as the return component arising from the idiosyncratic or specific risk of asset i. The expected value of ε_i should be zero. This model has two constraints:

$$E[\varepsilon_i \varepsilon_h] = 0 \forall i, h; i \neq h$$

$E[\varepsilon_i (r_j^* - E[r_j^*])] = 0$ for all assets and risk factors

The paper by Ross [13] can be consulted for more details regarding arbitrage pricing theory. With a focus on statistical arbitrage, the assumption is made that securities with virtually identical risk exposures should yield approximately equal returns. Thus it may be possible to detect tradable pairs by investigating the risk exposure of the securities to particular risk factors.

It is up to the practitioner to determine the risk factors that should be considered and how a trading strategy based on this concept should be executed as Vidyamurthy [12] does not provide this information. This model based on APT has the clear objective of finding mispricings in pairs of relatable securities. It is built on the notion of a well-known pricing theory that is supported by fundamental economic reasoning.

E. Cointegration Method

In contrast to the minimum distance method which made no model assumptions, the cointegration method for statistical arbitrage is a model-based parametric approach. The notion of this approach is that if two financial securities follow a common stochastic trend, the spread between the price series of the securities may be weakly stationary. More formally, if two securities' price series are both integrated of order d and there exists is a linear combination of the two price series that will be integrated of order $d - b$ ($b > 0$), then the two price series are considered cointegrated (CI(d, b)).

With the objective of performing statistical arbitrage,

interest is placed on the situations where $d - b = 0$ such that there exists a stationary time series for the spread. Since a significant number of price series are integrated of order one, I(1), focus can be placed more specifically on the situations where $b = d = 1$.

An advantage of cointegrated price series is that the series can be represented by an error correction model (ECM). In an ECM, the dynamics of a time series at a particular point in time is a correction of the preceding period's deviation from the equilibrium with the addition of possible lag dynamics. Harlacher [14] expresses the cointegrated relation between two time series X_t and Y_t as:

$$\Delta y_t = \Psi_0 - \gamma_y(y_{t-1} - \alpha - \beta x_{t-1})$$
$$+ \sum_{i=1}^{K} \Psi_{x,i}\Delta x_{t-i} + \sum_{i=1}^{L} \Psi_{y,i}\Delta y_{t-i} + \varepsilon_{y,t} \tag{2}$$

and similarly

$$\Delta x_t = \xi_0 + \gamma_x(y_{t-1} - \alpha - \beta x_{t-1})$$
$$+ \sum_{i=1}^{K} \xi_{y,i}\Delta y_{t-i} + \sum_{i=1}^{L} \xi_{x,i}\Delta x_{t-i} + \varepsilon_{x,t} \tag{3}$$

where $\varepsilon_{y,t}$ and $\varepsilon_{x,t}$ depict white noise, Δy_t and Δx_t depict one period differences in y_t and x_t respectively and the terms α, β and γ represent the OLS estimates of the cointegrating regression that relates the time series. In the case that no deterministic trend is present in the series, the constants Ψ and ξ are zero. The main advantage of the ECM representation of this system is that active forecasts can be done by simply using past information. From equations (2) and (3) it can be seen that the part that represents the deviation from the long run equilibrium is $y_{t-1} - \alpha - \beta x_{t-1}$. This term must be weakly stationary and the two coefficients γ_y and γ_x must have opposite algebraic signs. If these conditions did not hold true, no error-correcting behaviour would be present. The presence of this phenomenon and of cointegration in general can be tested for by using the procedure set out by Engle and Granger [15] or the more generalized approach of Johansen [16].

From a trading perspective, it is not necessary that a spread series is weakly stationary. The spread only needs to be mean-reverting in nature for an effective trading rule to be implementable. This model is not limited to trading only pairs of securities, but can be generalized to more securities where a possible mean-reverting equilibrium can exist.

F. Previous Documented Performance of Statistical Arbitrage Models

Gatev, Goetzmann and Rouwenhorst [11] tested the minimum distance method (referred to as pairs trading) on the US stock market using daily data over the period of 1962-2002. They documented average annualized returns of up to 11% for self-financing portfolios of stock pairs. They argue that the robustness of the excess returns indicate that the minimum distance approach to statistical arbitrage profits from temporary mispricing of close substitutes.

Caldeira and Moura [17] investigated the selection of a portfolio of pairs based on the cointegration approach to statistical arbitrage on the São Paulo Stock Exchange. Their study used daily data from the exchange over the period of January 2005 to October 2012. Empirical analysis showed that their approach delivered excess returns of up to 16.38% per year, having a Sharpe ratio of 1.34 and low correlation to the market. Their study considered transaction costs of 0.5% of each trade value (one-way).

Avellaneda and Lee [18] studied model-driven statistical arbitrage strategies in the US equity market over the period of 1997-2007. Their models consisted of an approach based on principle component analysis (PCA) and sector ETFs. Their market-neutral PCA-based strategy achieved an annual Sharpe ratio of 1.44 over the entire period. They also document that the strategy performance degraded over the time frame, having had an average Sharpe ratio of 0.9 over the period of 2003-2007.

Cummins [19] tested the minimum distance method for statistical arbitrage on equity markets in the United States, Japan, Hong Kong and China by using daily data over the period of January 2004 to December 2009. The results conclude that the algorithm does not yield consistent positive excess returns throughout time nor in different markets. The results were different for each of the markets examined in terms of profitability and portfolio volatility. According to Cummins the algorithm did prove to be extremely profitable during periods of stock market turmoil.

III. METHODOLOGY AND DATA

A. Description of Data

The financial data that was available for this study includes several security universes that consist of stocks and exchange-traded funds (ETFs). The data was provided in the form of price bars that include the daily opening, high, low and closing price of each security. The volume that was traded for each security was also provided. Data was not available for financial securities that have been delisted from the respective exchanges; this may cause some survivorship bias in the results. The period that will be examined ranges from 1 January 2006 to 30 June 2016. The security universes included in this study represent all tradable securities from the following exchanges:

1) New York Stock Exchange and NASDAQ (US)
2) Deutsche Börse Xetra (DE)
3) Tokyo Stock Exchange (JP)
4) Johannesburg Stock Exchange (SA)
5) Hong Kong Stock Exchange (CH)

B. Methodology

As a means of studying statistical arbitrage strategies, two trading systems will be studied. The first system is built on the minimum-distance approach that was studied by Gatev, Goetzmann and Rouwenhorst [11], which we will refer to as the pairs trading system. The second system follows a cointegration approach and is referred to as the adaptive statistical arbitrage system. Consult [20] for more details on the adaptive statistical arbitrage system.

The systems were set up using the baseline parameters that were chosen by Gatev, Goetzmann and Rouwenhorst to allow for a direct comparison of the results.

In this section we briefly explain our approach by reviewing the implementation details of the statistical arbitrage systems that will be examined. The default parameters are listed at the end of the system review. We also discuss the details regarding the simulation (or backtesting) environment.

C. Pairs Trading System

Similar to the approach by Gatev, Goetzmann and Rouwenhorst [11], the pairs trading system distinguishes between two periods, namely a formation period and a trading period.

In the formation period the squared error (SE) of all the normalized price series are calculated. The system ranks pairs of securities by prioritizing smaller SE values over larger SE values. This is done since similar securities are searched for by using only price data. A relatively small SE value corresponds to very similar price series. Pairs of securities are selected from the ranked SE values. It should be noted that each security is only paired at its lowest SE value with another security. The spread of each pair is calculated as:

$$s_{xy} = \frac{x_t}{y_t} \qquad (4)$$

where s_{xy} is the spread of the one period returns of the dividend-adjusted closing price series of securities x_t and y_t. The average price of the spread (μ_s) and the standard deviation of the spread (σ_s) are calculated for each of the pair spreads.

During the trading period, the spread value is calculated daily and a z-score (which indicates how many standard deviations an element is from the mean) is calculated with regards to the average price and standard deviation of the spread that was calculated in the formation period. If the z-score is higher or lower than a certain threshold value, a market position is taken such that one security is bought and the other is sold short. The monetary value of the bought and sold securities should be equal for the system to remain market-neutral. When the spread of the returns has converged back to the average value (μ_x), the z-score will have a value of zero and the pair position is closed. When the trading period ends all positions are closed regardless of whether the spreads have converged. The system then reverts back to the formation period.

D. Adaptive Statistical Arbitrage System

The adaptive statistical arbitrage system is built on the notion of the cointegration of security prices. This allows the system to trade baskets of related securities rather than only pairs. Similarly to the pairs trading system, the adaptive system also alternates between two periods. The first period is referred to as the learning period and the second as the trading period.

During the learning period the affinity propagation clustering technique is applied to the security universe. The clustering of the universe allows for groups of related securities to be extracted and thus avoids having to perform statistical tests on all possible combinations of securities. The clustering is performed by constructing a similarity function that quantifies the similarity between securities using correlation and squared error. Clusters that consist of more than twelve securities are split until all clusters contain less than thirteen securities. This is done to allow for the baskets to be tested by the Johansen cointegration method which only allows for up to twelve securities to be tested at one time.

When the Johansen cointegration test is applied to the clusters, it is possible to search for related securities and construct (weakly) stationary series when cointegration is found. The stationary series are obtained by linearly weighting the securities in the basket according to the largest resulting eigenvector of the Johansen method. We limit the creation of these weakly stationary series to instances where the test indicates at least a 95% certainty of cointegration between the securities. Baskets of securities that have passed the test for cointegration are ranked (in ascending order) according to the variance of the stationary series.

A GARCH(1,1) model is constructed using a student-t likelihood function with the objective of modelling the volatility of the weakly stationary series. The objective of the GARCH model is to dynamically update market entry thresholds by increasing or decreasing the threshold when volatility increases or decreases.

During the trading period, the value of each (weakly) stationary series that has been constructed is calculated. Similarly to the pairs trading system, a z-score is calculated for each of the stationary series. This z-score is compared to the threshold for market entry. When the z-score is larger than the positive threshold value, the stationary series is effectively sold short. When the z-score is smaller than the negative threshold value, the stationary series is effectively bought. The baseline parameters for this system are depicted in Table I.

TABLE I: BASELINE PARAMETERS OF SYSTEMS

Parameter	Pairs trading system	Adaptive statistical arbitrage system
Formation/learning period	1 year	1 year
Trading period	1 semester	1 semester
Max amount of pairs/baskets to trade	5 pairs	5 baskets
Entry threshold	±2.0 standard deviations	Initially 1.5 standard deviations (GARCH updated thereafter)

E. Backtest Setup

The initial capital that will be considered for each backtest is 100,000 currency units. The currency unit is reliant on the market that is being examined. A default transaction fee of 0.3% plus an additional 0.1% of the trade value is taken into account to allow for rental costs of short positions. Orders are placed (and executed) at the next market opening price after having generated a trading signal. Only 85% of the capital is made available for all active trades to allow for possible margin requirements.

IV. HYPOTHESES

This study tests three hypotheses concerning statistical

arbitrage performance on different markets during different market regimes. The hypotheses are defined as follows:

1) Hypothesis 1: The returns produced by statistical arbitrage models is *larger* for emerging markets than for developed markets.
2) Hypothesis 2: Statistical arbitrage excess returns have *decreased* since the financial crisis of 2008-2009.
3) Hypothesis 3: Statistical arbitrage strategies produce *larger* excess returns during non-trending/bearish market regimes than during bullish market regimes.

V. EMPIRICAL RESULTS

In order to test hypothesis 1, the two statistical arbitrage models were backtested on the different security universes. The main statistic that is of concern for testing this hypothesis is the compound annual growth rate (CAGR). The CAGR of the statistical arbitrage systems that were obtained from the backtest results of the developed markets are compared to the CAGR of the systems tested on the emerging markets.

Table II contains the CAGR of each statistical arbitrage system with regards to the different security universes that were examined. For markets that have more than 200 exchange traded funds (ETFs) available, the system results are shown for both a stock-only universe and an ETF-only universe.

TABLE II: PERFORMANCE EVALUATION OF SYSTEMS (2006-2016)

| Market	Pairs trading system (CAGR)	Adaptive statistical arbitrage system (CAGR)
Deutsche Borse Xetra stocks (DE)	5.4427%	9.0051%
Tokyo Stock Exchange ETFs (JP)	3.1954%	3.7382%
Tokyo Stock Exchange stocks (JP)	3.5416%	8.7773%
NYSE/NASDAQ ETFs (US)	5.4197%	7.8518%
NYSE/NASDAQ stocks (US)	2.4285%	7.1178%
Johannesburg Stock Exchange (SA)	9.3950%	4.9076%
Hong Kong Stock Exchange stocks (CH)	17.3566%	18.1910%
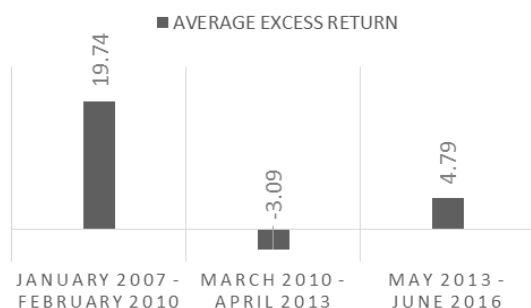
Fig. 1. Average excess return produced by systems.

From the results in Table II, we can compute the average CAGR of the statistical arbitrage systems for developed markets and emerging markets. The average CAGR of the systems on the developed markets is 5.65% while the average CAGR of the systems, as applied on the emerging markets, is 12.46%. We can hereby conclude that hypothesis 1 holds true over the period of January 2007 – June 2016. Note that the

systems have a training period where no trading takes place from January 2006 to January 2007.

We test for hypothesis 2, by dividing the period of January 2007 to June 2016 into three periods of equal length such that each period contains 37 months. The statistic that is of importance in testing this hypothesis is excess returns (or alpha). Once again the systems were evaluated over these three time frames and the excess returns were measured. The average excess returns are shown in Fig. 1.

From the results shown in Fig. 1, it is clear that the period of January 2007 to February 2010 provided considerably more exploitable opportunities for statistical arbitrage than the succeeding two periods. The second period (when financial markets recovered from the financial crisis) did not provide fertile grounds for statistical arbitrage systems. The excess returns improved in the third period, but are much less than the results from the first period. Hypothesis 2 thus holds true as excess returns for statistical arbitrage systems have declined since the financial crisis in 2008-2009.

In order to test our third hypothesis, that statistical arbitrage systems have larger excess returns during non-trending / bearish markets, two different market regimes were examined for each universe. We characterized the markets as upward trending (bullish) or non-trending / downward trending (bearish) by examining the price action of the indices that track the stock markets. We defined bullish markets as periods where a market index experienced a CAGR of more than 10%. Bearish markets were defined as periods where the CAGR of a market index was less than 2%. Our choices for market index trackers and the periods found to be either bullish or bearish are depicted in Table III.

TABLE III: INDEX TRACKERS AND MARKET REGIME PERIODS

Index tracker (country)	Bearish period	Bullish period
DAX index (DE)	Jul 2007 – Jul 2011	Oct 2011 – Apr 2015
Nikkei 225 index (JP)	Jan 2009 – Jan 2012	Jan 2013 – Jan 2015
S&P 500 index (US)	Jan 2008 – Jan 2102	Jan 2012 – Jan 2015
MSCI EZA (SA)	Jan 2011 – Jan 2015	Jan 2009 – Jan 2011
HSI index (CH)	Jan 2010 – Jan 2012	Jan 2007 – Jan 2008

The annualized growth rate of the index trackers for the bearish and bullish periods that were chosen for each of the financial markets that we examine are shown in Table IV to support our decisions for choosing these periods.

TABLE IV: ANNUALIZED RETURNS OF INDEX TRACKERS

Index tracker (country)	Bearish period	Bullish period
DAX index (DE)	-1.46%	23.78%
Nikkei 225 index (JP)	1.07%	28.90%
S&P 500 index (US)	-2.79%	15.73%
MSCI EZA (SA)	0.38%	34.47%
HSI index (CH)	-2.55%	32.82%

The statistical arbitrage systems were tested over the specified periods in Table III. The excess returns produced by the systems during these periods were measured. The average excess returns during bearish market regimes was calculated to be 5.99%. The average excess returns during bullish

markets were -15.89%. These results support hypothesis 3 as the excess returns during bearish market regimes were noticeably larger than for the identified bullish market regimes.

VI. CONCLUSIONS

In this study we examined the performance of statistical arbitrage systems over a range of different financial markets including the Deutsche Börse Xetra, Tokyo Stock Exchange, New York Stock Exchange, NASDAQ, Johannesburg Stock Exchange and the Hong Kong Stock Exchange. The study considered the period of January 2006 – June 2016 which included different market regimes and a stock market crash.

We examined the performance of two statistical arbitrage systems by formulating three hypotheses. The first hypothesis investigated whether returns for statistical arbitrage systems are larger in emerging markets than in developed markets. It was found that this was indeed the case over the period that was studied. The average CAGR of the systems in developed markets was 5.65% and in emerging markets was found to be 12.46%. The second hypothesis investigated whether excess returns of the systems have increased or decreased since the financial crisis. It was found that excess returns have noticeably decreased since the stock market crash, but that the excess returns seemed to be slightly increasing again in recent years. The third hypothesis investigated the performance of statistical arbitrage during different market regimes. It was found that these systems generally have positive excess returns during non-trending or downward trending markets. Our results indicated average excess returns of 5.99% for non-trending markets. During upward trending markets, the systems generally had negative excess returns, with an average excess return of -15.89%.

We conclude that statistical arbitrage systems can be a source of excess returns during non-trending market regimes and that these systems can provide returns that are uncorrelated to the overall market returns. The systems that we studied were unable to provide positive excess returns during upward trending markets, indicating that these periods may be better suited for trend-following systems.

In future research the statistical arbitrage systems can be tested on higher frequency data to observe performance on an intraday trading basis. As statistical arbitrage has been demonstrated to perform differently in different market regimes methods will be developed to automatically detect the onset of bearish and bullish market to allow automated switching between market strategies suitable for the current regime.

REFERENCES

[1] O. Bondarenko, "Statistical arbitrage and security prices," *Review of Financial Studies*, vol. 16, no. 3, pp. 875-919, 2003.

[2] I. Aldridge, *High Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*, 2nd ed. New Jersey, U.S.: Wiley, ch. 1, pp. 1-19, 2013.

[3] R. K. Narang, *Inside the Black Box: A Simple Guide to Quantitative and High Frequency Trading*, 2nd ed. New Jersey, U.S.: Wiley, ch. 11, pp. 199-203, 2009.

[4] E. F. Fama, "The behavior of stock-market prices," *Journal of Business*, vol. 38, pp. 34-105, 1965.

[5] B. G. Malkiel, "The efficient market hypothesis and its critics," *The Journal of Economic Perspectives*, vol. 17, no. 1, pp. 59-82, 2003.

[6] S. J. Grossman and J. E. Stiglitz, "On the impossibility of informationally efficient markets," *The American Economic Review*, vol. 70, no. 3, pp. 393-408, 1980.

[7] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of Finance*, vol. XLVIII, no. 1, pp. 65-91, 1993.

[8] L. Chan, N. Jegadeesh and J. Lakonishok, "Momentum strategies," *The Journal of Finance*, vol. LI, no. 5, pp. 1681-1713, 1996.

[9] C. L. Dunis and R. Ho, "Cointegration portfolios of european equities for index tracking and market neutral strategies," *Journal of Asset Management*, vol. 6, no. 11, pp. 33-52, 2005.

[10] J. P. Nobrega and A. Oliveira, "Improving the statistical arbitrage strategy in intraday trading by combining extreme learning machine and support vector regression with linear regression models," in *Proc. IEEE 25th International Conference on Tools with Artificial Intelligence*, 2013, pp. 182-188.

[11] E. Gatev, W. Goetzmann and K. Rouwenhorst, "Pairs Trading: Performance of a Relative Arbitrage Rule," *Review of Financial Studies*, vol. 19, no. 3, pp. 797-827, 2006.

[12] G. Vidyamurthy, *Pairs Trading: Quantitative Methods and Analysis*, 1st ed. New Jersey, U.S.: Wiley, ch. 6, pp. 90-92, 2004.

[13] S. A. Ross, "The arbitrage theory of capital asset pricing," *Journal of Economics*, vol. 13, no. 3, pp. 341-360, 1976.

[14] M. Harlacher, "Cointegration based statistical arbitrage," M.S. thesis, Dept. Math., Swiss Federal Institute of Tech., Zurich, Switzerland, 2012.

[15] R. F. Engle and C. Granger, "Co-integration and error correction: Representation, estimation and testing," *Econometrica*, vol. 55, no. 2, pp. 251-276, July 1987.

[16] S. Johansen, *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, 1st ed. Oxford, U.K.: Oxford University Press, 1995.

[17] J. F. Caldeira and G. V. Moura, "Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy," *Rev. Bras. Finanças*, vol. 11, no. 1, pp. 49-80, March 2013.

[18] M. Avellaneda and J. H. Lee, "Statistical arbitrage in the U.S. equities market," *Quantitative Finance*, vol. 10, no. 1, pp. 1-22, July 2010.

[19] R. L. Cummins, "Pairs trading: Testing the consistency of a statistical arbitrage investment strategy," M.S. thesis, Econ. And Buss, Erasmus Uni., Rotterdam, Netherlands, 2010.

[20] G. J. A. Visagie, "An adaptive algorithmic trading system for statistical arbitrage," M.Eng. thesis, Dept. Computer and Elec. Eng., North West Uni., Potchefstroom, South Africa, 2016.

Gabriel Visagie was born on July 10, 1992 in Welkom, South Africa. He was awarded a full scholarship from the National Research Foundation for the M.Eng. in Computer and Electronic Engineering, with focus on intelligent systems and predictive modelling at the North-West University, Potchefstroom in 2015. He received his B.Eng. in Computer and Electronic Engineering from the North West University in 2014 and was awarded the SASOL award (2012), SAIEE award (2013) and SRC prestige awards (2012-2014) for graduating at the top of his class. In 2017 he will have a fulltime position at Allan Gray, the largest privately owned investment manager in Africa with over R230.4bn in assets under management. His research interests are in algorithmic trading systems, quantitative analysis and portfolio management.

Alwyn J. Hoffman was born and raised in Bloemfontein in the Free State. He studied as an Electronics Engineer at the University of Pretoria (UP), where he eventually also obtained Masters in Engineering, Ph. D. and MBA degrees. He worked in the South African high technology electronics industry during the period 1985 to 1994, when he joined North-West University as Director of the School of Electrical and Electronic Engineering for a period of 6 years. Between 2001 and 2008 he again spent time in the high-technology industry, working for Inala Technology Investments in the field of intellectual property development and for IPICO Inc., a company listed on the Toronto Stock Exchange, in the field of automated identification. Since January 2009 he rejoined North-West University's Faculty of Engineering, and is still involved as consultant in the fields of data mining and intelligent systems.