

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Statistical Arbitrage in South African Financial Markets

Kieran Govender¹

Abstract

Engle and Granger's (1987) cointegrating framework provides a useful method of analyzing the dynamics of nonstationary data in both the short and long-run. However, despite its popularity in various areas of research, the application of cointegration to financial data has been limited. This paper provides an example of the application of cointegration in a pairs trading strategy to identify mean reverting spreads. The strategy is implemented with an algorithmic trading setup that models the spread in a state-space form. A recursive Bayesian algorithm is used to continuously estimate the time-varying parameters in this model; as in Triantafyllopoulos and Montana (2011). After modelling departures from the long-term relationship as an Ornstein-Uhlenbeck process, an optimal trading strategy is then derived to maximize expected returns subject to the cost of trading. This optimization problem is solved by deriving an exact analytical solution, to minimize computational time; as in Bertram (2010). This strategy is applied to securities on the Johannesburg Stock Exchange and the results suggest that it is possible to derive large arbitrage profits from its implementation.

¹*govender.kieran@gmail.com*

1. Introduction

Pairs trading is a relatively simple investment practice that involves the identification of a common long-term movement in the behaviour of two assets (i.e. where the spread between the prices or returns has remained relatively constant). When the prices of these assets temporarily deviate from the historical relationship (or equally when the spread is temporarily greater than the long-term level), then one would short the overpriced asset and buy the underpriced asset. If the spread was to return to the long-term level, the trader could close the respective positions at a profit.

Despite the simplicity of this strategy, Gatev et al. (2006) suggest that pairs trading strategies have performed relatively well in the past,¹ which has attracted the attention of market participants. Following the greater use of relatively simple strategies, traders were no longer able to derive large returns from these strategies, and were thereafter largely compensated for enforcing the Law of One Price.²

Traditionally, pairs trading strategies have focused on the correlation in the returns of two assets. These types of strategies would ignore the behaviour of changes to the long-term trend and are primarily concerned with a very short trading term (Alexander, 1999). In addition, any changes to the long-term trend would establish significant instability in the relationship, which would necessitate frequent rebalancing.

In this thesis, an investigation is conducted into the merits of a new pairs trading methodology, which is based on a long-term cointegrating relationship that may exist between certain share prices.³ Hence, if it can be shown that two share prices are cointegrated, then they would converge on the long-term trend, following a temporary disturbance. After identifying this disturbance, one would then be able to execute the corresponding buy

¹Indeed, the relative success of these strategies has led to the development of a branch of literature that questions the overall efficiency of the market. Much of this literature is combined with that which considers strategies that profit from market overreaction and momentum effects, as in Jegadeesh and Titman (1995), where short-run sentiment drives the market like a “popularity contest” (Graham and Dodd, 2009).

²The Law of One Price suggests that assets that produce the same payoff during every state, must sell for the same price. If not then arbitragers would short the over-valued asset and purchase the under-valued asset, to produce an arbitrage profit (Danthine and Donaldson, 2005). This would force prices to converge, thus ensuring that the Law of One Price holds.

³An alternative implementation of this type of strategy, for shares quoted in the Standard & Poor’s 100 index, is contained in Alexander et al. (2002).

and sell orders that would eventually lead to an arbitrage profit, after the prices have converged.⁴

The identification of the temporary disturbance that may lead to a profitable trading opportunity, is performed in an algorithmic trading setup that incorporates a state-space model, with state equations for the level and time varying parameters. This structure of this model follows Triantafyllopoulos and Montana (2011) and is estimated recursively, using Bayesian methods; to provide new estimates for the parameters following the introduction of new information from the market. The estimation algorithm also monitors the underlying statistical relationships to ensure that they continue to apply once a position has been opened. This is important, since if it is found that the shares are no longer cointegrated, after a position has been opened, then a stop-loss trade would need to be executed.

To ensure that the execution of trades is limited to those opportunities that will maximize expected returns, an optimization framework is utilized to evaluate the possible trading opportunities that are identified by the state-space model. This procedure provides estimates for the optimal points at which positions should be opened and closed, based on the size of inefficiencies, which is modelled as an Ornstein-Uhlenbeck (OU) process.⁵

This strategy is then applied to daily data for securities on the Johannesburg Stock Exchange (JSE) from June 2002 to August 2008, where the initial in-sample period covers the first 100 days. A running total of the respective profits and losses from the trades is kept and it is noted that on only a few isolated occasions, would this strategy provide a loss. Hence, as long as the trader is able to remain in the market, they would be able to derive substantial profits, that may range up to 140% per annum.

In what follows; section 2 considers the theoretical framework of cointegrated pairs trading strategies and section 3 considers the formulation of the algorithmic trading setup. Practical details that relate to the implementation of the strategy are contained in section 5. The results are presented in section 6, and concluding comments are provided in section 7.

2. Cointegrated Pairs Trading Strategies

When the prices of two financial assets contain a common stochastic trend then they are deemed to be cointegrated (Stock and Watson, 1988).

⁴Despite the interest of South African investment banks in this type of strategy we are not aware of any successful implementation that currently exists.

⁵This section largely follows Bertram (2009) and Bertram (2010)

To describe such a relationship in a formal manner consider a vector, $\mathbf{P}_t = (P_{1t}, P_{2t})'$, that contains the prices of two shares. If each of these share prices is integrated of order d , and there exists a vector, $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, which ensures that the linear combination of $\boldsymbol{\beta}' \mathbf{P}_t$ is integrated of an order that is lower than d , then the share prices are said to be cointegrated (Engle and Granger, 1987).^{6,7}

The short-term relationship that exists between these variables could then be expressed by an error-correction model (ECM), which takes the form,⁸

$$\begin{aligned}\Delta P_{1,t} &= \vartheta_1 \left(P_{1,t-1} - \tilde{\beta} P_{2,t-1} \right) + \xi_{1,t} \\ \Delta P_{2,t} &= \vartheta_2 \left(P_{1,t-1} - \tilde{\beta} P_{2,t-1} \right) + \xi_{2,t}\end{aligned}$$

Hence, if a cointegrating relationship exists between two share prices then a short-term change in price will be influenced by past deviations from the long-term relationship. When the long-term relationship is $CI(1, 1)$, the ECM will be stationary, which would imply that any deviation from the long-term relationship will be temporary.

To derive a profit from this relationship, one would sell short the share that is currently overvalued, and long the share that is currently undervalued.⁹ After the share prices have then converged on the value that is determined by the long-term relationship, the individual would then need to close these positions to realize their profit. The method that is used to identify the respective optimal points in time, when positions should be opened and closed, is determined by algorithmic trading setup, which is described below.

3. Algorithmic trading setup

The algorithmic trading setup contains two parts. The first part considers the identification of instances where the current value of the observed

⁶A variable is integrated of order d , i.e. $I(d)$, if it contains d stochastic trends and needs to be differenced d times before it is stationary.

⁷Where $\beta_1 P_{1t} + \beta_2 P_{2t}$ is integrated of an order $(d - b)$, with $b > 0$ and $\boldsymbol{\beta} \neq \mathbf{0}$, these variables are then cointegrated of order $CI(d, b)$.

⁸The long-run relationship, $\beta_1 P_{1t} + \beta_2 P_{2t}$, may be normalised by dividing through by β_1 , such that $\tilde{\beta} = \beta_2 / \beta_1$.

⁹In this instance a share is deemed to be overvalued when it is trading above the price that would be inferred by the long-term relationship, and vice versa.

spread differs from the 'true' spread¹⁰, which characterises a market inefficiency. The estimate of the true spread is achieved through the use of a Gaussian linear state-space model for the spread, which is continuously estimated in real time. The second part of the setup considers the evaluation of prospective trading opportunities, to ensure that the execution of trades is limited to those that are deemed to be optimal. This involves the formulation of an optimization problem, for which an analytical solution is derived to minimize the time that is taken to evaluate prospective opportunities.

3.1. State-space models for the identification of trading opportunities

State-space models may be used to describe the behaviour of an unobserved component, by defining its relationship with an observed process (Tsay, 2005). We are looking to estimate the true level of the spread,¹¹ through its relationship with the observed spread. In this paper we make use of the Gaussian linear state-space model of the spread and the associated online estimation algorithm as presented in Triantafyllopoulos and Montana (2011). We will give a brief introduction and theoretical overview of the model, for a more rigorous treatment the interested reader may refer to the original paper by Triantafyllopoulos and Montana.

Recall the observed spread, y_t . We will assume that y_t is some noisy realisation of the true spread, call it x_t , which cannot be observed. This assumption will allow us to make use of a Gaussian linear state-space model to estimate the true spread. So we have formulated the observation equation as (Elliott et al., 2005) and (Triantafyllopoulos and Montana, 2011)¹²,

$$y_t = x_t + w_t, \quad w_t \sim N(0, D^2) \quad (1)$$

The true spread, x_t , will depict the true level of the market and as such any difference between the true and observed spread indicate a temporary market inefficiency from which we will seek to profit (Triantafyllopoulos and Montana, 2011). It is clear that we would expect the difference $y_t - x_t = w_t$ to be mean reverting. Since w_t is white noise, it is stationary and deviations from the long-run mean are temporary. When $y_t - x_t$ (or w_t) becomes too large (small) we would expect that the system will move back

¹⁰When we say the true spread we mean the fair value of the spread.

¹¹The true spread will be unobserved, and generally not equal to the observed spread

¹²The equation (1) below describes a market where observed asset prices fluctuate around their fair value. Indicative of many modern day theories of stock market behaviour where prices are driven by sentiment, and exhibit both overreaction and under reaction.

towards equilibrium.¹³ We attempt to profit from the market disequilibrium and subsequent correction by trading accordingly. We will use the online estimation algorithm to obtain real time estimates of the true spread, then we define a new spread equal to w_t , it is this spread that will form our trading signal. In other words when w_t is too large or too small, this will 'signal' an inefficiency in the market and we will trade accordingly.

The model in Triantafyllopoulos and Montana (2011) is an extension of the model presented in Elliott et al. (2005), and as such we shall refer to both papers. Assume that there are two assets whose price process' can be represented as p_{1t} and p_{2t} , we obtain the observed spread y_t through OLS (Ordinary Least Squares) estimation as^{14 15}

$$y_t = p_{1t} - \beta_0 - \beta_1 p_{2t} \quad (2)$$

Elliott et al. (2005) then describes the unobserved true spread, x_t , as

$$x_t - x_{t-1} = a - bx_{t-1} + \varepsilon_t \quad (3)$$

where, ε_t has an independent and identical normal distribution. By defining $A = a$ and $B = 1 - b$, the equation can then be simplified to,

$$x_t = A + Bx_{t-1} + \varepsilon_t \quad (4)$$

To ensure that x_t is stationary in equation (4), it requires, $|B| < 1$. Triantafyllopoulos and Montana (2011) have since extended this model by incorporating time dependence in the parameters. This involves the addition of two state equations to describe the evolution of the level, A , as an autoregressive process, and specification of the coefficient, B , as a time-varying parameter. After substituting $y_t - w_t$ for x_t , the observed spread, y_t , may be modelled as,

$$\begin{aligned} y_t &= A_t + B_t y_{t-1} + \varepsilon_t \\ A_t &= \phi_1 A_{t-1} + \nu_{1t} \\ B_t &= \phi_2 B_{t-1} + \nu_{2t} \end{aligned} \quad (5)$$

¹³That is the observed spread will correct by falling (rising)

¹⁴We obtain estimates of β_0 and β_1 through OLS.

¹⁵Where the asset prices p_{it} for $i = 1, 2$ are not necessarily an individual share and may be a portfolio of assets (Triantafyllopoulos and Montana, 2011)

where $\epsilon_t = w_t - Bw_{t-1} + \varepsilon_t$. The time subscripts on A and B achieve time dependence in the parameters (Triantafyllopoulos and Montana, 2011). We further assume that ϕ_1 and ϕ_2 lie inside the unit circle, so as to ensure that the processes A_t and B_t are weakly stationary (Triantafyllopoulos and Montana, 2011). The equations can be simplified through the use of matrix notation¹⁶ with $\theta_t = (A_t, B_t)'$ and $F_t = (1, y_{t-1})'$ as,

$$\begin{aligned} y_t &= F_t \theta_t + \epsilon_t \\ \theta_t &= \text{diag}(\phi_1, \phi_2) \theta_{t-1} + \nu_t \\ \epsilon_t &\sim N(0, \sigma^2) \end{aligned} \tag{6}$$

Where $\epsilon_t \sim N(0, \sigma^2)$, $\nu_t = (v_{1t}, v_{2t}) \sim N_2(0, \sigma^2 \nu_t)$ and N_2 denotes a bivariate Normal distribution. We assume that the error processes, ϵ_t and ν_t , are uncorrelated with each other and the initial state vector (θ_1) for all t .¹⁷ In addition we make the usual assumption that $E(\epsilon_t \epsilon_s) = 0$, for all $s \neq t$ and similarly for v_t (Triantafyllopoulos and Montana, 2011).¹⁸

The mean reversion of the observed spread, y_t , is of particular interest to us, and since the parameters in the model of the observed spread are time varying we need to revise the conditions which ensure the spread is mean reverting (Triantafyllopoulos and Montana, 2011). We will use part of **Theorem 1** as presented in Triantafyllopoulos and Montana (2011),

“ $\{y_t\}$ will be mean reverting when ϕ_1 and ϕ_2 lie inside the unit circle, ν_t is bounded and $|B_t| < 1$ for all t ”

We use the online estimation algorithm as set out in Triantafyllopoulos and Montana (2011) to obtain estimates of the parameters in model (6). The estimation algorithm will allow us to obtain real time estimates of B_t , which will allow us to monitor mean reversion in real time using **Theorem 1** above.

Estimation takes place within a Bayesian framework which makes use of a recursive algorithm to obtain real time estimates (Triantafyllopoulos and Montana, 2011). Recall,

$$\begin{aligned} y_t &= A_t + B_t y_{t-1} + \epsilon_t \\ \theta_1 &= (A_1, B_1)' \end{aligned} \tag{7}$$

¹⁶This is the model written in state space form

¹⁷This assumption is common to a state space framework see Durbin and Koopman (2001)

¹⁸See Durbin and Koopman (2001)

where $\epsilon_t \sim N(0, \sigma^2)$ and θ_1 is the initial state vector. We make the assumption that m_1, P_1, n_1 and d_1 are known and that¹⁹,

$$\begin{aligned} (\theta_1 | \sigma^2) &\sim N_2(m_1, \sigma^2 P_1) \\ \sigma^2 &\sim IG\left(\frac{n_1}{2}, \frac{d_1}{2}\right) \end{aligned} \quad (8)$$

this assumption will allow us to obtain the posterior densities using the recursive formulae described below (Triantafyllopoulos and Montana, 2011). To see this consider the following, assume that we have the posterior distributions at time $t - 1$,

$$(\theta_{t-1} | \sigma^2, y_{t-1}, y_{t-2}, \dots) \sim N_2(m_{t-1}, \sigma^2 P_{t-1}) \quad (9)$$

$$(\sigma^2 | y_{t-1}, y_{t-2}, \dots) \sim IG\left(\frac{n_{t-1}}{2}, \frac{d_{t-1}}{2}\right) \quad (10)$$

Then an application of Bayes Theorem,

$$P(\theta_t | \sigma^2, y^t) = \frac{P(y_t | \theta_t, \sigma^2) P(\theta_t | \sigma^2, y^{t-1})}{P(y_t | \sigma^2, y^{t-1})},$$

allows us to obtain the posterior distributions at time t as,

$$(\theta_t | \sigma^2, y_t, y_{t-1}, \dots) \sim N_2(m_t, \sigma^2 P_t) \quad (11)$$

$$(\sigma^2 | y_t, y_{t-1}, \dots) \sim IG\left(\frac{n_t}{2}, \frac{d_t}{2}\right) \quad (12)$$

Triantafyllopoulos and Montana (2011) summarize this procedure for parameter estimation by listing a number of updating equations that are to be estimated sequentially, to derive the recursive posterior distributions:²⁰

$$R_t = \Phi P_{t-1} \Phi + V_t, \quad Q_t = F_t' R_t F_t + 1 \quad (13)$$

$$e_t = y_t - F_t' \Phi m_{t-1}, \quad K_t = R_t F_t Q_t^{-1} \quad (14)$$

$$m_t = \Phi m_{t-1} + K_t e_t, \quad P_t = R_t - K_t' K_t Q_t \quad (15)$$

$$r_t = y_t - F_t' m_t, \quad n_t = n_{t-1} + 1 \quad (16)$$

$$d_t = d_{t-1} + r_t e_t, \quad S_t = \frac{d_t}{n_t} \quad (17)$$

¹⁹ IG denotes an inverted gamma distribution, used to reflect that small positive values for the variance are more likely to occur.

²⁰Note that some matrices are defined implicitly by the equations in (13)-(17)

Where $F_t = (1, y_{t-1})'$, m_t is the mean vector of $\theta_t = (A_t, B_t)$, Φ is defined as,

$$\Phi = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix},$$

and V_t is a covariance matrix.

$$V_t = \begin{pmatrix} \delta_1^{-1}(1 - \delta_1)\phi_1^2 p_{11,t-1} & 0 \\ 0 & \delta_2^{-1}(1 - \delta_2)\phi_2^2 p_{22,t-1} \end{pmatrix} \quad (18)$$

In addition we assume the initial values as suggested by Triantafyllopoulos and Montana (2011),

$$(m_1; P_1; n_1; d_1) = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; 1000 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; 3; 1 \right) \quad (19)$$

and estimate the value of the hyper parameters $\phi_1, \phi_2, \delta_1, \delta_2$ through maximum likelihood estimation. The hyper parameters maximise the log-likelihood function,

$$\begin{aligned} \ell(\phi_1, \phi_2, \delta_1, \delta_2) &= \sum_{t=2}^T p(y_t | y_{t-1}, y_{t-2}, \dots) \quad (20) \\ &= \sum_{t=2}^T \log \frac{\Gamma(n_t/2)}{\sqrt{\pi n_{t-1}} \Gamma(n_{t-1}/2)} - \frac{1}{2} \sum_{t=2}^T n_t \log \left(1 + \frac{(y_t - f_t)^2}{n_{t-1} Q_t S_{t-1}} \right) \end{aligned}$$

The ability of the algorithm to monitor the mean reversion of the spread in real time provides the user powerful stop loss tools. Following Triantafyllopoulos and Montana (2011), the trader could implement that all positions are closed when y_t no longer exhibits mean reverting behaviour, which occurs when the characteristic roots of ϕ_1 and ϕ_2 are no longer within the unit circle, ν_t is unbounded, or $|B_t| \geq 1$. As we shall later see the ability to monitor mean reversion in real time will also reduce model risk.

The algorithm, which continuously calculates estimates as new information comes from the market, is used to monitor differences between the observed spread, y_t , and the unobserved true level of the spread, x_t . By monitoring differences between the true and observed spread, y_t , we can identify market inefficiencies. We now consider w_t and how to trade upon w_t in order to maximize expected returns.²¹ To do this we use an Ornstein-Uhlenbeck process to model the inefficiencies, and an optimal trading strategy is derived following the work done by Bertram (2009) and Bertram (2010).

²¹ $\{w_t\}$ is the process of inefficiencies.

3.2. Optimal trading strategies for Ornstein-Uhlenbeck processes

When the true spread differs from the observed spread, that is when $w_t \neq 0$, there is an inefficiency in the market. The problem is that not all inefficiencies can be traded upon profitably after transaction costs are taken into account. A trader would open a long trade when the inefficiency is large enough, say when

$$w_t \leq a \quad \text{for some } a < 0, a \in \mathbb{R} \quad (21)$$

The trader would close the trade when the inefficiency corrects, say when

$$w_t \geq m \quad \text{for some } m \in \mathbb{R} \quad a < m \quad (22)$$

and earn the profit $m - a - c$, where c are the transaction costs incurred from opening and then closing a trade in the spread. We now attempt to answer the question of when the inefficiency is “large enough”. To be more specific we will determine the values of a and m that maximise the expected return given a level of transaction cost, c . It is a well established idea that active traders keep markets efficient, however small inefficiencies often persist even in the presence of active managers. These small inefficiencies must not be mistaken for trading opportunities as transaction costs can wipe out profits on these small inefficiencies, ensuring losses for anyone willing to trade. It is evident the problem of when to open and close trades (the values of a and m) must take into account the level of transaction costs (c).

In addressing the problem we adopt the approach and eventual solutions of Bertram (2009) and Bertram (2010), who expressed the problem of when to open and close trades as a function of two random variables; namely the return per trade and the frequency at which trades take place (Bertram, 2010). When we think of the process of pairs trading there are two possible dangers with regard to the open and close signals, namely open and close signals that are ‘too small’ and of course having open and close signals that are ‘too large’.²² These ideas lead to the formulation by Bertram (2010). We derive open and close signals which maximise expected return, and in order to do this we need an equation for determining the expected return per trade taking into account transaction costs.

²²When the open and close signals are ‘too small’ the result is frequent trading, and lots of small profits and losses which are accompanied by high levels of transaction costs. On the other hand when the open and close signals are ‘too large’ we miss out on profitable trading opportunities. Clearly neither situation is optimal for maximising expected returns.

Consider a process X_t , where in particular X_t can be modelled as an Ornstein-Uhlenbeck (OU) process (with Wiener process, W_t)

$$dX_t = -\alpha X_t dt + \eta dW_t \quad \text{for } \alpha > 0 \quad \eta > 0 \quad (23)$$

The OU process is often used to model mean reverting processes, in particular it is common practice for the observed spread to be modelled as a mean reverting OU process (Bertram, 2009) and (Ekstrom et al., 2009). In (23) when $\alpha > 0$, the process is mean reverting and α is the rate of reversion (Bertram, 2009) and (Herlemont, 2003). It is also important to note that the use of an OU process to model, X_t , imposes a Normal distribution on X_t . In particular (23) has a long term distribution with zero mean and constant variance.²³ This has often been a problem for financial data since empirically financial data have shown many non Normal characteristics, and the validity of any such model is questionable in practice (Bertram, 2009) and (Bertram, 2004).

We postulate that we can use the same models used to model the observed spread from a general pairs trading strategy, to model w_t . In particular we use the model of Bertram (2009) and Bertram (2010) to model the process of inefficiencies, w_t , for use in a pairs trading strategy.²⁴ It is a good time to make two important points, firstly the use of (23) to model w_t will impose a Normal distribution on w_t which will in the long run have zero mean and constant variance. Now recall that this is consistent with our state space model and it is what we assumed about w_t in equation (1). Secondly note that the use of an OU process to model financial processes introduces

²³The long term distribution is obtained by letting $t \rightarrow \infty$ in the distribution of $X_t \sim N\left(X_0 e^{-\alpha t}, \frac{\eta^2}{2\alpha}(1 - e^{-2\alpha t})\right)$

²⁴Recall that the process of inefficiencies, w_t , is both Normal and mean reverting. The OU model thus fits the process of inefficiencies. It is a subtle but important point that the level of w_t gives us an indication of future movements of the spread. Furthermore we assume that the observed spread moves toward the true spread, so that loosely speaking the size and direction of movements in w_t give a reasonable estimate of the size and direction of movements in y_t . These last two points make the use of models in Bertram (2009) and Bertram (2010) appropriate. It is also worth noting that the processes $\{y_t\}$ and $\{w_t\}$ hold the same information. They are both indicators of a relative mispricing. Consider that the observed spread y_t , identifies inefficiencies in the market since any deviation of y_t from its equilibrium indicates a mispricing. As such we see that y_t and w_t deliver us the same information. Note that y_t in this discussion is simply a mean reverting linear combination of assets in absence of a state space model. Where y_t is simply modelled as mean reverting to some long run mean, and not the unobserved state x_t . w_t comes from the state space framework discussed above.

model risk, in the sense that (23) describes a mean reverting process and if the financial process is at some stage no longer mean reverting the model is invalid (Ekstrom et al., 2009). The online estimation algorithm monitors mean reversion of y_t to x_t in real time, and thus in theory the model monitors the mean reversion of w_t in real time thus reducing model risk.

The estimation of the Ornstein-Uhlenbeck process in (23) is carried out using the process and formulae given in (Jurek and Yang, 2007).²⁵ We assume that we can model inefficiencies as an Ornstein-Uhlenbeck process. That is we assume that the OU process, X_t , is the process of inefficiencies, w_t , in the derivation that follows. Furthermore in the derivation that follows assume that we open a long position when $X_t \leq a$ and close the position when $X_t \geq m$, $a < m$ ²⁶. We continue as in Bertram (2009) by considering the frequency with which we trade. The trading frequency for a strategy is defined for a unit of time. It is simply the number of trades that occur in a single unit of time (Bertram, 2009). To make this more formal we define the total trade length. Recall that we open trades at $X_t \leq a$ and close trades at $X_t \geq m$ and $a < m$, then the total trade length is the time it takes for X_t to move from a through m and back to a . When this occurs we say that we have completed a trade cycle. Define τ_{exit} as the time it takes to move from trade enter to trade exit (to move from a to m), τ_{enter} as the time to move from trade exit to trade enter (to move from m to a) and τ_{total} as the time it takes to complete a trade cycle. It is clear that,

$$\tau_{total} = \tau_{exit} + \tau_{enter} \quad (24)$$

The frequency at which trades take place is defined as $\frac{1}{\tau_{total}}$ (Bertram, 2009). In addition τ_{total} is a random variable because X_t is a random variable²⁷, and as such we can calculate the expectation and variance of τ_{total} (Rice, 1995). Note that τ_{exit} and τ_{enter} are first passage times, furthermore they are independent due to the independent increments property of Brownian motion and have finite expectation and variance²⁸ (Bertram, 2010), (Riccardi and Sato, 1988) and (Darling and Siegert, 1953). So we may write,

$$E(\tau_{total}) = E(\tau_{exit}) + E(\tau_{enter}) \quad (25)$$

²⁵For completeness the derivation and formulae as presented in Jurek and Yang (2007) are contained in the appendix.

²⁶In words we open trades when inefficiencies are large enough, and close trades after the subsequent correction, $a < 0$.

²⁷The source of randomness is the Wiener process, W_t

²⁸The first-passage times of an Ornstein-Uhlenbeck process have a finite mean and variance

$$Var(\tau_{total}) = Var(\tau_{exit}) + Var(\tau_{enter}) \quad (26)$$

Define N_t as the number of trades over a time interval of length t ²⁹ (Bertram, 2010). In addition from the definition of both N_t and the trade frequency we must have,

$$\frac{N_t}{t} = \frac{1}{\tau_{total}} \quad (27)$$

Note the trivial but important point that (27) implies that if we multiply the trade frequency over an interval by the length of that interval we obtain the number of trades over the interval, N_t . In Bertram (2010) it is noted that technically N_t is a counting process for a renewal process. The inter-arrival times for this renewal process are given by τ_{total} , the trade cycle length Bertram (2010). The trade cycle lengths for different trade cycles are independent and identically distributed, this characteristic allows us to invoke the renewal density theorem along with the central limit theorem to obtain,

$$\lim_{t \rightarrow \infty} E(N_t) = \frac{t}{E(\tau_{total})} \quad (28)$$

(Bertram, 2010). Since we are using the natural logarithm of prices we may define the continuously compounded return per trade as (another advantage of using log prices),

$$r(a, m, c) = m - a - c \quad (29)$$

Where c is the transaction cost (a and m are open and close signals as before) (Bertram, 2010). The expected return per unit time $\mu(a, m, c, t)$ is (Bertram, 2010),

$$\mu(a, m, c, t) = r(a, m, c) \frac{E(N_t)}{t} \quad (30)$$

Using (28), we can write³⁰,

$$\mu(a, m, c) \approx \frac{r(a, m, c)}{E(\tau_{total})} \quad (31)$$

We have a formula for $r(a, m, c)$ so what remains is to find an expression for $E(\tau_{total})$, the first moment of the first-passage time. The moments of first-passage for the Ornstein-Uhlenbeck process are well documented for a dimensionless system. As in Bertram (2010) we transform (23) into a

²⁹Where t are the number of units of time in the interval

³⁰Where we have dropped the t to simplify notation

dimensionless system by setting $Y_t = \frac{X_t\sqrt{2\alpha}}{\eta}$, using a time dilation $\tau = \alpha t$ and a simple application of Ito's formula.³¹

Proof.

$$\begin{aligned} dY_t &= dX_t \frac{\sqrt{2\alpha}}{\eta} \\ &= (-\alpha X_t dt + \eta dW_t) \frac{\sqrt{2\alpha}}{\eta} \\ &= -\alpha \frac{Y_t \eta}{\sqrt{2\alpha}} \frac{\sqrt{2\alpha}}{\eta} \frac{1}{\alpha} d\tau + \sqrt{2\alpha} dW_t \end{aligned}$$

Now recall that $W_t \sim N(0, t)$ and that $\tau = \alpha t$, so $\frac{1}{\sqrt{\alpha}} dW_\tau \sim N(0, t)$ and we may write the dimensionless system as (Bertram, 2010),

$$dY_t = dY_t = -Y_t d\tau + \sqrt{2} dW_\tau \quad (32)$$

□

With this transformation we must be aware that the trade entry level is now $\bar{a} = \frac{a\sqrt{2\alpha}}{\eta}$, the trade exit level is $\bar{m} = \frac{(m\sqrt{2\alpha})}{\eta}$ and transaction cost $\bar{c} = \frac{(c\sqrt{2\alpha})}{\eta}$, where we have used the relationship between the processes, $Y_t = \frac{(X_t\sqrt{2\alpha})}{\eta}$ (Bertram, 2010). Similarly the trade cycle length has been scaled to $T_{total} = \alpha\tau_{total}$, this will be important when obtaining the trade frequency (Bertram, 2010). As in Bertram (2010) and Riccardi and Sato (1988) let $\{Y(t), t \geq 0\}$ be an OU process, consider the system beginning at $y = y_0$ and hitting a barrier $y = b$ for the first time, we may define the first passage time, T_{b,y_0} as,

$$T_{b,y_0} = \inf \{t \geq 0 : Y_t > b | Y_0 = y_0\} \quad (33)$$

(Bertram, 2010). We now make use of the moments as calculated in Riccardi and Sato (1988) and presented in Bertram (2010), in particular

$$E(T_{b,y_0}) = \begin{cases} \phi_1(b) - \phi_1(y) & \text{if } y_0 < b, \\ \phi_1(-b) - \phi_1(-y_0) & \text{if } y_0 > b. \end{cases}$$

³¹Ito's formula can be obtained in Shreve (2004).

Where

$$\phi_1(z) = \frac{1}{2} \sum_{k=1}^{\infty} \frac{\Gamma(\frac{k}{2})(\sqrt{2}z)^k}{k!} \quad (34)$$

And $\Gamma(\cdot)$ is the gamma function, defined as (Rice, 1995),

$$\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u} du$$

Note that the process (32), is symmetric about 0. This allows us to write (using the notation introduced in (33)),

$$T_{enter} = T_{\bar{a}, \bar{m}} = T_{-\bar{a}, -\bar{m}} \quad (35)$$

$$T_{exit} = T_{\bar{m}, \bar{a}} \quad (36)$$

So,

$$E(T_{total}) = E(T_{exit}) + E(T_{enter}) \quad (37)$$

$$= E(T_{\bar{m}, \bar{a}}) + E(T_{-\bar{a}, -\bar{m}}) \quad (38)$$

$$= [\phi_1(\bar{m}) - \phi_1(\bar{a})] + [\phi_1(-\bar{a}) - \phi_1(-\bar{m})] \quad (39)$$

$$\begin{aligned} &= \frac{1}{2} \sum_{k=1}^{\infty} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{m})^k}{k!} - \frac{1}{2} \sum_{k=1}^{\infty} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{a})^k}{k!} \\ &+ \frac{1}{2} \sum_{k=1}^{\infty} \frac{\Gamma(\frac{k}{2})(\sqrt{2}(-\bar{a}))^k}{k!} - \frac{1}{2} \sum_{k=1}^{\infty} \frac{\Gamma(\frac{k}{2})(\sqrt{2}(-\bar{m}))^k}{k!} \end{aligned} \quad (40)$$

Define κ as the set of even integers and ζ as the set of odd integers, and note that when k is even we have that,

$$\frac{1}{2} \sum_{k \in \kappa} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{m})^k}{k!} - \frac{1}{2} \sum_{k \in \kappa} \frac{\Gamma(\frac{k}{2})(\sqrt{2}(-\bar{m}))^k}{k!} = 0 \quad (41)$$

$$\frac{1}{2} \sum_{k \in \kappa} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{a})^k}{k!} - \frac{1}{2} \sum_{k \in \kappa} \frac{\Gamma(\frac{k}{2})(\sqrt{2}(-\bar{a}))^k}{k!} = 0 \quad (42)$$

Similarly when k is odd we have that,

$$\frac{1}{2} \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{m})^k}{k!} - \frac{1}{2} \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2}(-\bar{m}))^k}{k!} = \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{m})^k}{k!} \quad (43)$$

$$\frac{1}{2} \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{a})^k}{k!} - \frac{1}{2} \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2}(-\bar{a}))^k}{k!} = \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2}\bar{a})^k}{k!} \quad (44)$$

Applying this to (40) we get,

$$E(T_{total}) = \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2\bar{m}})^k}{k!} - \sum_{k \in \zeta} \frac{\Gamma(\frac{k}{2})(\sqrt{2\bar{a}})^k}{k!} \quad (45)$$

Now recall that when $k \in \zeta$ we have that (Rice, 1995),

$$\Gamma\left(\frac{k}{2}\right) = \frac{\sqrt{\pi}(k-1)!}{2^{k-1}\left(\frac{n-1}{2}\right)!} \quad (46)$$

Using (45) and (46) we can write $E(T_{total})$ as,

$$E(T_{total}) = \left[\frac{\sqrt{\pi}(\sqrt{2\bar{m}})}{1!} + \frac{\sqrt{\pi}2!(\sqrt{2\bar{m}})^3}{2^23!} + \frac{\sqrt{\pi}4!(\sqrt{2\bar{m}})^5}{2^42!5!} + \dots \right] \quad (47)$$

$$- \left[\frac{\sqrt{\pi}(\sqrt{2\bar{a}})}{1!} + \frac{\sqrt{\pi}2!(\sqrt{2\bar{a}})^3}{2^23!} + \frac{\sqrt{\pi}4!(\sqrt{2\bar{a}})^5}{2^42!5!} + \dots \right]$$

$$= \pi \left[\frac{2}{\sqrt{\pi}} \left(\frac{\bar{m}}{\sqrt{\pi}}\right) + \frac{2}{3\sqrt{\pi}} \left(\frac{\bar{m}}{\sqrt{\pi}}\right)^3 + \frac{1}{5\sqrt{\pi}} \left(\frac{\bar{m}}{\sqrt{\pi}}\right)^5 + \dots \right] \quad (48)$$

$$- \pi \left[\frac{2}{\sqrt{\pi}} \left(\frac{\bar{a}}{\sqrt{\pi}}\right) + \frac{2}{3\sqrt{\pi}} \left(\frac{\bar{a}}{\sqrt{\pi}}\right)^3 + \frac{1}{5\sqrt{\pi}} \left(\frac{\bar{a}}{\sqrt{\pi}}\right)^5 + \dots \right]$$

To simplify things further we must now consider the complex error function. The complex error function is defined in appendix A. Now note that the Taylor expansion of the complex error function (or erfi) around zero is,

$$erfi(x) = \sum_{n=0}^{\infty} u(n)x^n \quad (49)$$

Where the $u(n)$ function is defined recursively as,

$$u(0) = 0 \quad (50)$$

$$u(1) = \frac{2}{\sqrt{\pi}} \quad (51)$$

$$-2nu(n) + (n^2 + 3n + 2)u(n+2) = 0 \quad (52)$$

To make things a little clearer we give the first few terms of the Taylor expansion,

$$erfi(x) = \frac{2}{\sqrt{\pi}}x + \frac{2}{3\sqrt{\pi}}x^3 + \frac{1}{5\sqrt{\pi}}x^5 + \dots \quad (53)$$

It is now easy to see that we may simplify $E(T_{total})$ using (53) and write,

$$E(T_{total}) = \pi \left[\operatorname{erfi} \left(\frac{\bar{m}}{\sqrt{2}} \right) - \operatorname{erfi} \left(\frac{\bar{a}}{\sqrt{2}} \right) \right] \quad (54)$$

Using $\bar{a} = \frac{a\sqrt{2\alpha}}{\eta}$, $\bar{m} = \frac{m\sqrt{2\alpha}}{\eta}$ and $T_{total} = \alpha\tau_{total}$ and substituting this into (54) we obtain the expression for $E(\tau_{total})$ as presented in Bertram (2010),

$$E(\tau_{total}) = \frac{\pi}{\alpha} \left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\sqrt{\eta}} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right] \quad (55)$$

Using equations (29) and (56), and substituting this into (31) we obtain a formula for the expected return,

$$\mu(a, m, c, t) = \frac{m - a - c}{\frac{\pi}{\alpha} \left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\sqrt{\eta}} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right]} \quad (56)$$

This is what we set out to do, we may now solve for the open and close signals which maximise the expected return. To do this we take the derivatives with respect to both a and m , set the derivatives equal to zero and solve for the open and close signals. To obtain said derivatives we use $\frac{d}{dx} \operatorname{erfi}(x) = \frac{2}{\sqrt{\pi}} e^{x^2}$, along with a fairly trivial application of the chain and quotient or product rule. A little manipulation of the terms then yields the desired result,

$$\frac{\partial \mu}{\partial m} = \frac{\alpha}{\pi} \frac{1}{\left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\eta} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right]} \quad (57)$$

$$\begin{aligned} & - \frac{\alpha}{\pi} \frac{1}{\left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\eta} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right]^2} \frac{2}{\sqrt{\pi}} e^{\frac{m^2\alpha}{\eta^2}} \frac{\sqrt{\alpha}}{\eta} \\ & = \frac{\alpha}{\pi} \left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\eta} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right] \\ & - \frac{\alpha}{\pi} \frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha}}{\eta} e^{\frac{m^2\alpha}{\eta^2}} (m - a - c) \end{aligned} \quad (58)$$

Set (58) equal to zero to obtain,

$$\frac{\partial \mu}{\partial m} = \frac{\alpha}{\pi} \left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\eta} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right] \quad (59)$$

$$- \frac{\alpha}{\pi} \frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha}}{\eta} e^{\frac{m^2\alpha}{\eta^2}} (m - a - c) \\ = \frac{\pi}{\alpha} \left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\eta} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right] \quad (60)$$

$$- \frac{\pi}{\alpha} \frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha}}{\eta} e^{\frac{m^2\alpha}{\eta^2}} (m - a - c) \\ = \sqrt{\frac{4\pi}{\alpha\eta^2}} e^{\frac{m^2\alpha}{\eta^2}} (m - a - c) - \frac{\pi}{\alpha} \left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\eta} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right] \quad (61) \\ = 0$$

Similarly we obtain,

$$\frac{\partial \mu}{\partial a} = \sqrt{\frac{4\pi}{\alpha\eta^2}} e^{\frac{a^2\alpha}{\eta^2}} (m - a - c) - \frac{\pi}{\alpha} \left[\operatorname{erfi} \left(\frac{m\sqrt{\alpha}}{\eta} \right) - \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right] \quad (62) \\ = 0$$

Equations (61) and (62) together with the assumption that $a < m$ imply that to achieve $\frac{\partial \mu}{\partial m} = \frac{\partial \mu}{\partial a} = 0$ we must have $a^2 = m^2$ and $m = -a$ (Bertram, 2010). Substituting $m = -a$, into (56) and (61) we obtain a new formula for the optimum expected return,

$$\mu^*(a, c) = \frac{\alpha(2a + c)}{2\pi \operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right)} \quad (63)$$

We obtain the optimal value of a by solving the equation,

$$e^{\frac{a^2\alpha}{\eta^2}} (2a + c) - \sqrt{\frac{\pi}{\alpha}} \eta \left[\operatorname{erfi} \left(\frac{a\sqrt{\alpha}}{\eta} \right) \right] = 0 \quad (64)$$

Although equation (64) may be solved using numerical methods, an approximate analytic solution may also be derived, as in Bertram (2010). Such a solution would speed up the time that is taken to find the optimal values of the parameters. To derive this solution, take a third order Taylor expansion of equation (64), around $a = 0$. The optimal value of a is then given by,

$$a = -\frac{c}{4} - \frac{c^2\alpha}{4(c^3\alpha^3 + 24c\alpha^2\eta^2 - 4\sqrt{3c^4\alpha^5\eta^2 + 36c^2\alpha^4\eta^4})^{\frac{1}{3}}} \\ - \frac{(c^3\alpha^3 + 24c\alpha^2\eta^2 - 4\sqrt{3c^4\alpha^5\eta^2 + 36c^2\alpha^4\eta^4})^{\frac{1}{3}}}{4\alpha} \quad (65)$$

This form of the solution was utilized in the algorithmic trading setup, that was implemented in the following application.

4. Strategy

Following this explanation, we are able to apply the strategy to the natural logarithm of daily price data over the period beginning in 2002/05/30 and ending 2007/08/01, using the above estimation technique to identify respective mispricings. The data is separated into trading periods of 250 days. The first 100 days prior to each trading period (call this the estimation period) are used to estimate all the parameters necessary to execute the strategy over the trading period. During each trading period we behave as if we were actually trading with the strategy, in particular at each stage we behave as if we have no knowledge of the future. In order to run the strategy for a year³² we require 350 days of price data (estimation and trading period data). During every period we use the same shares; to be more precise each 350 day period we use only the shares with both price data and margin requirements available for the whole period 2002/05/30-2007/08/01 as our share universe.

Initially we ensure that all the share prices are $I(1)$. We then run an Augmented Dickey Fuller (ADF) test on each pair. The cointegrating regression is,

$$y_t = \ln p_{1t} - \beta \ln p_{2t} \quad (66)$$

If the p-value for the ADF test is less than 1 per cent, then we may proceed. If the p-value is greater than 1 per cent we conclude that the spread, y_t , is not mean reverting and we cannot use the models developed, in particular the Gaussian linear state-space is only valid for mean reverting spreads (Triantafyllopoulos and Montana, 2011).³³ It is an appropriate time to comment on the use of cointegration to determine the existence of mean reverting spreads. The use of cointegration could result in missed trading opportunities since it is a strict statistical test. Cointegration may reject some spreads which are only weakly mean reverting (Cummins, 2010). In addition to this we have not used cointegration with the usual 5 or 10 per cent level of significance, but have used a 1 per cent level of significance. The use of a p-value cut off of 1 per cent, results in us excluding even more

³²We assume 250 trading days in a year.

³³Note again that the algorithm monitors mean reversion in real time, reducing model risk and allowing to use cointegrated pairs that are not from the same industry.

possible trading opportunities. It could be argued that the use of some less stringent determinants may be more appropriate, however we have not imposed the usual restriction of confining our search for mean reverting pairs to shares in the same industry. This as we have explained introduces the additional risk that the relationship is only temporary. With that in mind the use of strict determinants may be more appropriate than is at first apparent.

When the p-value of the ADF test on (66) is less than 1 per cent, we have identified a mean-reverting observed spread y_t . We proceed by modelling y_t using the Gaussian linear state-space model described above. We estimate the hyperparameters by maximising the log-likelihood given by (20) by making use of the initial values suggested in Triantafyllopoulos and Montana (2011), and given by (19)³⁴. The use of the estimates of the hyperparameters and the online estimation algorithm, will allow us to obtain estimates for the true spread, x_t over the 350 days. Recall that along with the observed spread, estimates of the true spread will allow us to identify inefficiencies in the market. We know that when the difference $y_t - x_t = w_t$ is large enough, the observed spread will correct and move toward the true spread (as long as the cointegration relationship holds). In order to determine when the difference $y_t - x_t = w_t$ is large enough, we model w_t as an Ornstein-Uhlenbeck process. We then use the open and close signals that maximize expected returns as given by Bertram (2010) and (65). So when $w_t < a$, the observed spread must correct and move towards the true spread. The observed spread will rise. To do this we assume that p_{1t} will rise and p_{2t} will fall. We would then invest in the spread given by (67). Thus the position of w_t will act as a signal as to the expected future movements of the assets.

$$\ln p_{1t} - \ln p_{2t} \tag{67}$$

When $w_{t_1} < a$, we open a trade and invest in $\ln p_{1t_1} - \ln p_{2t_1} = a^*$, and when $w_{t_2} > m$ we close the trade by selling $\ln p_{1t_2} - \ln p_{2t_2} = m^*$. The continuously compounded return for the trade will then be³⁵ (Bertram, 2010),

$$m^* - a^* - c \tag{68}$$

Where c is the level of transaction cost as a percentage for a full trade cycle and $t_1 < t_2$. When setting the level of c we follow a similar process to that in Bernardi and Gnoatto (2010). Assume that transaction costs are

³⁴The maximisation is done using the data available in the estimation period, that is we use the 100 days of estimation period data to carry out the estimation.

³⁵Recall that we are working with the natural logarithm of prices

a percentage of trade. That is when we trade in an asset with price p_t we pay transaction cost³⁶ of $c_1 \times p_t$, where $0 < c_1 < 1$. When we trade in the spread we go long one asset and short another, taking transaction costs into account this will cost,

$$p_{1t}(1 + c_1) - p_{2t}(1 - c_1) \quad (69)$$

It is important to note that c is the transaction cost from a full trade cycle in the spread of the natural logarithm of the prices. So we must take natural logarithms of (69) and then by manipulating the terms a little we obtain,

$$\ln p_{1t} - \ln p_{2t} + \ln \left(\frac{1 - c_1}{1 + c_1} \right) \quad (70)$$

Note that $\ln \left(\frac{1 - c_1}{1 + c_1} \right)$ is always negative, and thus must be added. Again recall that the way c is constructed, it is the transaction costs incurred from a full trade cycle³⁷. It is fairly easy to show that the transaction cost incurred when we close the trade in the spread is again, $\ln \left(\frac{1 - c_1}{1 + c_1} \right)$, thus we have that

$$c = -2 \times \ln \left(\frac{1 - c_1}{1 + c_1} \right) \quad (71)$$

Where the negative has come about since c is constructed as the absolute value of transaction costs in the model.

The strategy will be trading in contracts for difference or CFD's as opposed to a direct investment in the underlying share. A CFD will allow us to achieve the capital gain or loss from an investment in the share without having to purchase the underlying share. We can go both long or short in a CFD, which would correspond to a long or short investment in the underlying share. Trading in CFD's will allow significantly reduced transaction costs and both the benefits and added risks associated with gearing. The gearing comes as a result of only requiring the initial margin (a small percentage of the share price) to open a trade and earn the capital gain (or loss) from that share.

Effectively, we will run a margin account for all trades. When we open a trade we deposit the initial margin in the margin account. The initial margin payable will depend on the volatility and liquidity of the underlying share in question (StandardBank, 2011). The initial margin in the account

³⁶We shall later see that this is exactly what is happening in this case.

³⁷A full trade cycle involves both opening and closing a trade in the spread

earns interest which is paid daily by some institutions and monthly by others (StandardBank, 2011) and (Nedbank, 2011). The account will pay position interest on long positions and earn position interest on short positions. The interest on long and short positions are calculated on the aggregate of long and short positions at the close of every business day (StandardBank, 2011).

We incur transaction costs as a percentage of trade, and this is added to the price of the CFD. This will have an effect on the capital gain or loss, in other words the total capital gain or loss at the end of a long trade will now become $p_{close}(1 - c_1) - p_{open}(1 + c_1)$, where c_1 is the level of transaction costs. As usual all positions are marked-to-market daily. The first and last capital gains or losses will include the transaction costs. To make things a little clearer when we open a trade we would, firstly deposit the initial margin in the margin account. This initial margin would earn interest on each subsequent day. The capital gain³⁸ would be added to the margin account at the end of every day. The capital losses including transaction costs must be paid into the margin account at the end of every day as variation margin³⁹(StandardBank, 2011). We would pay interest on the aggregate long positions at the end of every business day; and would receive interest on the aggregate short positions at the end of every business day.

To run a margin account we need the initial margin requirement, rates at which initial margins and short positions earn interest and rates at which we pay interest on long positions. The initial margin requirements for shares, and other costs are obtained from the Nedbank and Standard Bank websites. The costs used in the strategy are,

- Cost for execution 0.35%
- Overnight long: SAFEX overnight rate + 2%
- Overnight short: SAFEX overnight rate - 3%

We ignore the interest earned by the initial margin for simplicity, but other than that the margin accounts operate as they would in reality⁴⁰. The costs seem more than reasonable since most hedge funds would have access to even lower transaction costs for the same services. A hedge fund can expect to have costs for execution lie somewhere in the range between 0.1% and 0.3%. We must consider the performance of a one day wait version of the strategy as presented in Gatev et al. (2006).

³⁸Including transaction costs on the first day after opening and on the day of closing

³⁹This is done so as to replace the loss which has been removed from the margin account

⁴⁰This is certainly more conservative and can only lower the performance of the strategy

The contrarian nature of a pairs trading strategy⁴¹ mean that the results can benefit from changes in bid and ask prices (Gatev et al., 2006). To understand this we must remember that the most widely available price data is close price data, and the close price quoted is generally the price at which the last trade of the day took place. We also know that in general ask prices are higher than bid prices, and as such we can immediately see some problems. Higher close prices are more likely to be ask prices and lower prices are more likely to be bid prices.

Another problem is that some movement in prices from one data point to the next could be due to the last price being an ask price and the subsequent price being a bid price or visa versa. So in pairs trading we have that when prices diverge the overpriced stock is more likely to be an ask price and the underpriced stock is more likely to be a bid price (Gatev et al., 2006). The problem is that we cannot trade at these prices since we must sell at bid prices and not at ask prices, and we must buy at ask prices not at bid prices.

A similar problem occurs when we close out the trades. There is the possibility that the capital gain on the short end is partly as a result of the price being more likely to be a bid quote, that is it is possible that part of the capital gain in a short trade is movement in the quoted price from an ask price to a bid price. The same result may occur for the capital gain on the long end being partly as a result of the quoted price moving from a bid price to an ask price.

The one day wait strategy will wait one day after the signal occurs to open (close) trades before actually opening (closing) trades. The belief is that by waiting one day prices will be equally likely to be bid or ask prices, and this will account for some of the bid-ask bias (Gatev et al., 2006). The one day wait version of the strategy should in theory account for the bid-ask bias, and initial runs of a one day wait version of the strategy on this occasion indicated that the method was not doing that. The results are thus excluded. There are still more concerns with the use of close price data, but we leave the discussion of these problems to a later stage.

A pairs trading strategy involves both a long and a short investment, and unless the returns from each end are equal we must have that one side earns a larger return than the other. The total strategy return will then lie somewhere between the return of the two ends. It is clear that some theory as to which end will earn a higher return could prove useful. The returns from investment in the long and short end of a pairs trading strategy should

⁴¹That is buy low and sell high

be roughly equal if profits are driven by mean reversion (Gatev et al., 2006). If profits are driven by changes in the default premium or bankruptcy risk, then we would expect the long end to generate a higher return than the short end (Gatev et al., 2006). The profits to a pairs trading strategy are driven by changes in bankruptcy risk, since we have paired a stock (the long end) with temporarily increasing bankruptcy risk⁴² with a stock (the short end) with constant or decreasing bankruptcy risk (Gatev et al., 2006).

When the risk of bankruptcy falls, the default premium falls driving returns down and price up. The increases in price are profitable to the long end only. In the absence of other price forces these changes in the default premium, help the long end make money and the short end lose money. Also consider the argument in Graham and Dodd (2009); here the authors cite the involvement of federal reserves and government in times of trouble in the market. When markets are falling at some stage the fed comes in and lowers interest rates to help the economy and boost asset prices. Recently government has also bailed out large financial organisations, that have got into trouble but were deemed 'too big to fail'. This sort of behaviour has been believed to give speculators the licence to speculate, as the writers put it "... it is the so called Bernanke put " (Graham and Dodd, 2009). The increased speculation and fed involvement can lead to extended periods in which the market is overvalued. If the market as a whole is considered overvalued, then this means that a large proportion of the shares that make up the market are overvalued for extended periods (Graham and Dodd, 2009). This tendency for shares to stay overvalued for long periods can make the short end more risky and less profitable.

We calculate profits from both the long and the short end respectively, but we only invest in the long end as in theory we believe this would achieve greater profits. The short end is generally used to fund the long end when capital is scarce, however we are using CFD's and both the long and short end require only the investment of initial margin. It is important to note that a long trade in the spread involves both long and short trades in individual shares. Similarly a short trade in the spread involves both long and short trades in individual shares. So even though we trade only in long trades in underpriced shares, we consider when the spread is both under- and overpriced.

We are now in a position to present some of the mechanics of the strategy for the pair, ASA-AFR. This will demonstrate what inefficiencies look like,

⁴²If the stock survives the risk of bankruptcy will begin to fall

and how the open and close signals translate into trades. We will demonstrate where we opened and closed trades for this particular pair. We will then go on to analyse the performance of the strategy as a whole over the 5 trading periods. First we will give a brief explanation of how returns are calculated.

5. Calculation of strategy returns

The calculation of strategy returns has its challenges. We first need to calculate daily returns for each pair, and then we need daily returns for the strategy as a whole⁴³.

We run a margin account for the long end of each pair. We will of course mark-to-market daily and calculate daily excess returns on each margin account as,

$$r_{i,t} = \frac{\text{MarginAccount}(i,t) - \text{CashInputs}(i,t) - \text{MarginAccount}(i,t-1)}{\text{MarginAccount}(i,t-1)}$$

Where i is the pair and t is the time period in days. Once we have the daily returns on each of the margin accounts, $r_{i,t}$, we can calculate the daily returns for the total strategy. The calculation of the total strategy returns are calculated following the procedure in Gatev et al. (2006). To be more specific total strategy excess returns at time t , $R_{P,t}$, are calculated as value weighted returns using the following formulae,

$$R_{P,t} = \frac{\sum_{i \in P} w_{i,t} r_{i,t}}{\sum_{i \in P} w_{i,t}} \quad (72)$$

$$w_{i,t} = w_{i,t-1}(1 + r_{i,t-1}) = (1 + r_{i,1}) \dots (1 + r_{i,t-1}) \quad (73)$$

Where the $w_{i,t}$ are the weights on pair i returns at time t and $w_{i,1} = 1$ for $\forall i$ (Yuksel et al., 2010). We use the more conservative measure of return, which simply has $w_{i,1} = 1$ for $\forall i$ even if pair i does not open a trade during the period under consideration (Gatev et al., 2006). Is is the so called return on committed capital, since the method takes into account that the fund will have to commit capital to a pair even if a pair does not trade (Gatev et al., 2006). It is important to note that a more realistic measure of return is the return on actual capital employed, where weights are non-zero only for pairs

⁴³The entire strategy contains an investment in all the pairs

that trade (Gatev et al., 2006). We have decided to stick with the more conservative measure throughout.

6. Results

We begin with a brief exposition of the signalling process and trades for a single pair, ASA-AFR. Table 1 shows the hyperparameters obtained for the pair by maximising (20). The estimates of ϕ_1 and ϕ_2 meet the conditions for mean reversion set out in Theorem 1, that is we have $|\phi_1| < 1$ and $|\phi_2| < 1$. Table 2 indicates the estimates of the parameters of the OU process, w_t , obtained from data in the estimation period. These estimates are needed for use in (65) to determine the optimal open and close signals for the pair, for use in the trading period. Figure 1 below displays the observed level of the spread along with the estimate of the state process (true level of the spread). The graph seems to indicate how the observed spread moves around its true level, over- and undercorrecting. Indicating at times a market that has characteristics like overreaction and delayed reaction as described by Jegadeesh and Titman (1995). Figure 2 below displays the signalling process, w_t . This is the process of inefficiencies in the market, it is the difference between the two series depicted in figure 1. When the inefficiencies cross either of the outer bands according to our model they become theoretically large enough to be profitably traded upon given the level of transaction costs. The lowest line depicts the position for opening a long trade and closing a short trade in the spread. The uppermost line depicts the position for closing a long trade and opening a short trade in the spread. These lines are a and m adjusted to lie symmetrically about the 20 day moving average. A slight change from the prescribed norm. Technically the Gaussian linear state space model does control for changes in the level of the spread, which means w_t should have a mean of zero. On the occasion where we do not completely capture changes in the level of w_t , this formulation will ensure that we do not miss out on trading opportunities.

Table 1: Estimated values of the hyperparameters

Hyperparameters			
ϕ_1	ϕ_2	δ_1	δ_2
0.8113	0.9985	0.2593	0.9970

Table 2: Estimated values of the Ornstein-Uhlenbeck process

OU parameters	
α	η
506.9104	1.4817

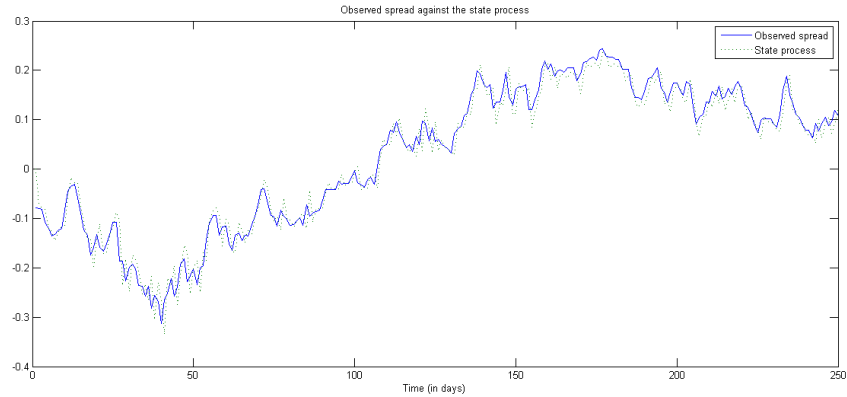


Figure 1: Observed spread against the estimated state process. The estimated state process is the estimation of the true level of the market.

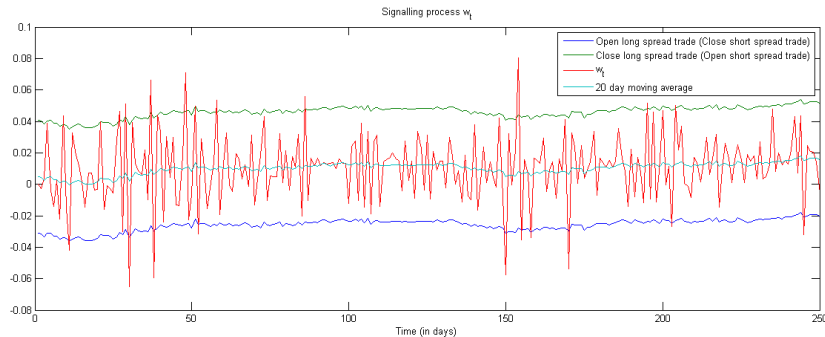


Figure 2: Signalling process

Figure 3 below depicts the open and close positions for long trades in share 1 for the pair ASA-AFR. The first trade was clearly a profit. We see that for the most part trades close at prices higher than they opened, we see some very small changes in price that trigger close signals. At about the 150 day mark there is a good example of the strategy doing what it is supposed to do. The last trade was forced to close and realise losses due to the end of the trading period. The result indicates that a trading rule that does not allow new trades to be opened after a certain period, say 230 days for example, may be useful. The Bertram (2010) model does calculate an expected time to complete a trade cycle for each process, making it possible to optimise this rule for each pair. When considering such a rule, we note that the Bertram (2010) model underestimates the expected time to complete a trade cycle (Cummins, 2010), and that some adjustment for this may be necessary. Figure 4 again depicts the open and close positions but for long trades in share 2. Again most trades are profitable but with the general upward trend in share 2 it would be unlikely that we would get a lot of long trades that lose money. Still the strategy seemed to avoid taking part in any significant falls in the share price. There were many very good trades in this share, and some trades which were in hindsight not losses but less than optimal. The total results over all the shares however indicate that in general there are significantly more good trades than bad.

We will now present the results for the strategy as a whole over the five trading periods. We begin with a fairly close look at the results of the strategy run over the most recent period, period 5. We will then present the results for all five years. The population of investable assets are the 97 listed shares on the JSE for which there was both share price data and

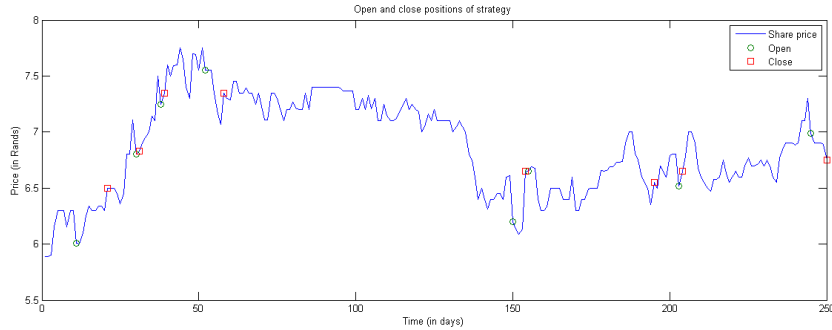


Figure 3: Open and close positions for long trades in share 1

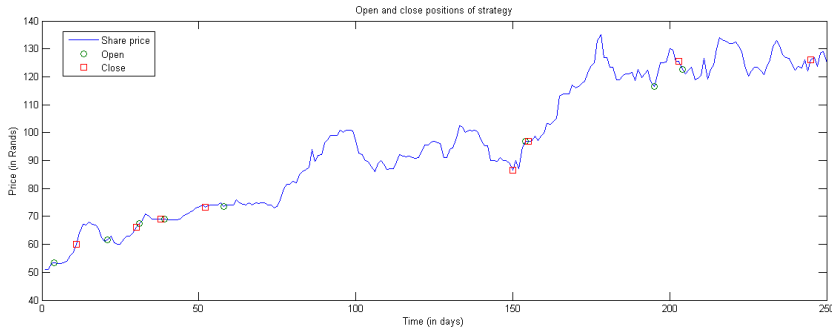


Figure 4: Open and close positions for long trades in share 2

margin requirements available over the entire period. Table 3 below displays the descriptive statistics for the strategy’s daily excess returns and compares these to the descriptive statistics for the daily excess returns of a passive investment in the JSE ALSI in period 5. It is clear that the average return of the strategy is much higher than that of a passive investment in the market. This can also be seen in figure 7 which depicts the cumulative return for the strategy in period 5 against that of the ALSI. The strategy performed exceptionally in this period. Returns in excess of 100% whereas the market yielded a return below 40%. The standard deviation, standard error and sample variance for the strategy are all higher than that of a passive investment in the market. The strategy does involve far greater volatility than a passive investment. Indeed a passive investment does give a better reward for risk (if we take risk to be the variance of return). The

ratio of the mean return to the sample variance give a good indication of this. In period 5 both a passive investment and the strategy returns are negatively skewed, however the strategy returns are somewhat more so. In period 5 the strategy is evidently risky, however on this occasion the additional risks came with higher returns. It is worthwhile to recall that we have derived open and close signals that maximise expected return, and not the reward for risk. Bertram (2010) does provide open and close signals that maximise the Sharpe ratio (or reward for risk) in a similar procedure to that used to maximise expected return.

The strategies performance over all five years seems very similar to that of period 5. In four out of the five years, the strategy beat the market by some distance. In period 1 the strategy did very poorly, losing about 80% of the capital invested. The market did poorly on this occasion as well, but it is again an example of the strategy's risk. Period 1 results showed a negative skewness, again displayed in both the strategy and a passive investment in the market, but again the negative skewness was far greater in the strategy. In Period 2 the strategy performed very well earning returns in excess of 140%, whereas a passive investment in the market would have earned the investor just under 30%. Period 2 strategy daily excess returns had the desirable property of being positively skewed⁴⁴. The strategy was again more volatile, but the positive skewness suggests that a significant portion of the volatility comes from large positive returns. In period 3 the strategy yielded in excess of 130% and the market produced just under 40%. It is noteworthy that in period 3 strategy returns were positively skewed whereas daily excess returns from a passive investment were negatively skewed. In period 3 the strategy does seem to give a slightly lower return for risk as indicated by the mean-variance ratio, the above mentioned positive skewness of the strategy in this period means that the statistic could be misleading. The positive skewness in the strategy returns means that some of the variance is from large positive returns which are not undesirable but are treated as such. Period 4 shows much of the same results, the strategy gave a return just under 60% and the market a return under 40%. The strategy is again more volatile than the market, and both are negatively skewed.

The strategy is clearly risky, and along with the significant use leverage the practical use of the strategy must be carefully considered. At the same time, the strategy can be very profitable. The strategy could be used as a small part of a portfolio, for a hedge fund or some high risk investment

⁴⁴These tables and histograms are contained in the appendices.

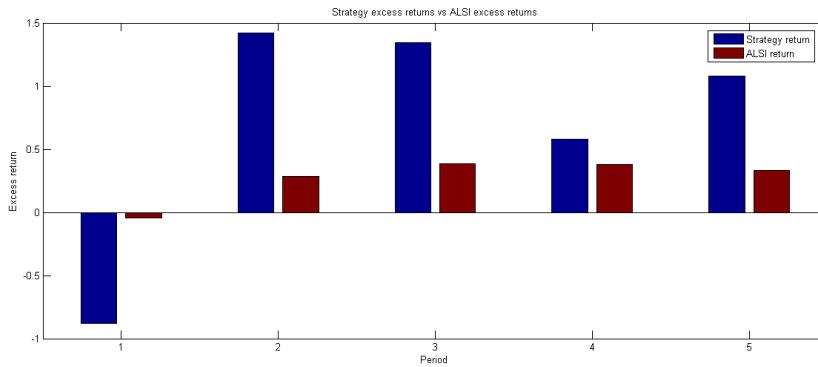


Figure 5: Strategy excess returns versus ALSI excess returns

scheme. The use of leverage and recent developments in banking make it unlikely that the strategy would be used in an investment banking environment. The strategy does seem to do a reasonable job of picking underpriced stocks, and could be used along with other factors (such as financial ratio analysis) to better determine which assets are underpriced.

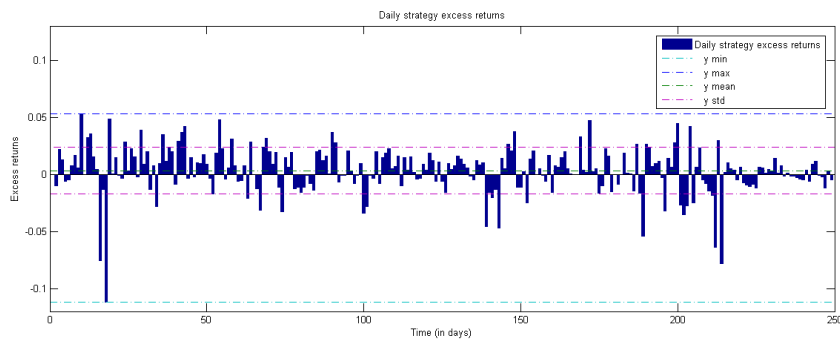


Figure 6: Daily strategy excess returns in period 5

Table 3: Descriptive statistics for the daily returns in period 5

	Strategy	ALSI
Mean	0.00315873	0.001202022
Standard Error	0.00129704	0.000613883
Median	0.003642566	0.002657211
Standard Deviation	0.020466941	0.009686906
Sample Variance	0.000418896	9.38362E-05
Kurtosis	5.409160261	0.77880429
Skewness	-1.252417741	-0.784956907
Range	0.164710951	0.055035379
Minimum	-0.111733967	-0.031702015
Maximum	0.052976985	0.023333364
Mean-Variance Ratio	7.540613246	12.80979108



Figure 7: Daily cumulative excess returns in period 5

7. Comments on results

There are a few things we must keep in mind when considering the results just presented.⁴⁵ An important point is the use of closing prices which are widely available but on the other hand unrealistic. One cannot trade at

⁴⁵We thank Jay Volmacka for his useful comments, discussions and insight into the hedge fund industry and practical use of a pairs trading strategy

these prices, they are often as a result of an auction. This brings additional volatility to the data, and the results will more often than not profit from this. There will thus be a tendency for the results of pairs trading strategies run on close price data to be better than they should be.

Another important point is that of liquidity, we have here considered 97 listed shares on the JSE. A large portion of this investable universe would not be considered as "investable" by a hedge fund for use in a relative pricing strategy. Some hedge funds have a universe of around 60 tradeable shares for the purposes of pairs trading, these are generally determined by some combination of volume traded and market capitalisation. The point is that not all the shares we have considered are liquid enough to be considered by hedge funds for an actual pairs trading portfolio. In reality practitioners avoid the use of anything much more complicated than the simple linear regression. There are concerns of the many (often unrealistic) assumptions involved in complex models. We have here made many assumptions. For many of these assumptions it can be argued that they have not held historically. Most importantly as previously noted the assumption of Normality of financial time series is invalid. The models do give some insight into the workings of the financial market. The results obtained in this paper seem to suggest that there should be more testing and further research into the use of more complicated algorithmic trading strategies for use in the South African market.

8. Conclusion

We examined the use of an algorithmic pairs trading strategy on the JSE. The strategy makes use of state space methodology to obtain estimates of the unobserved true value of the spread. The true level of the spread is estimated in real time, and these estimates are compared to the observed level of the spread. This allows us to identify market inefficiencies, these market inefficiencies are then modelled as an OU process. We then derive open and close signals that maximise expected returns, using the properties of an OU process. The online estimation algorithm monitors mean reversion in real time, allowing for a decrease in model risk. The results in this paper indicate that the strategy can be at times very profitable. The strategy is characterised by high risk, with the possibility of very high returns. The strategy outperformed the market on four out of the five periods tested in this paper. The long end of the strategy did earn a higher return on most occasions than the short end, lending support to the notion that profits to a pairs trading strategy are partly driven by changes in bankruptcy risk.

The model does seem to capture some of the relationship between observed market values and the fair market value, and lends support to the many claims of a market that is inefficient in the short run. Despite the use of many assumptions, the model performance indicates that more sophisticated trading algorithms may capture added returns. The work done here indicates that further research into non-Normal methods to identify and model market inefficiencies may be both useful and profitable.

9. Acknowledgements

A special thanks to my supervisor Kevin Kotze for all the help. We again also thank Jay Volmacka for useful discussions.

University of Cape Town

10. References

- Alexander, C., 1999. Optimal hedging using cointegration. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 357 (1758), 2039.
- Alexander, C., Giblin, I., Weddington, W., 2002. Cointegration and asset allocation: A new active hedge fund strategy. ISMA Centre Discussion Papers in Finance Series.
- Bernardi, S., Gnoatto, A., 2010. A pairs trading strategy applied to the european banking sector.
- Bertram, W., 2004. An empirical investigation of australian stock exchange data. *Physica A* (341).
- Bertram, W., 2009. Optimal trading strategies for itô diffusion processes. *Physica A: Statistical Mechanics and its Applications* 388 (14), 2865–2873.
- Bertram, W., 2010. Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and its Applications* 389 (11), 2234–2243.
- Cummins, M., 2010. Empirical analysis of optimal statistical arbitrage trading on the irish stock exchange.
- Danthine, J., Donaldson, J., 2005. *Intermediate Financial Theory* Second Edition. Elsevier, London.
- Darling, D. A., Siegert, A. J. F., 1953. The first passage problem for a continuous markov process. *The Annals of Mathematical Statistics*, 24, 624–639.
- Durbin, J., Koopman, S. J., 2001. *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.
- Ekstrom, E., Lindberg, C., Tysk, T., 2009. Optimal liquidation of a pairs trade, available at <http://www2.math.uu.se/~ekstrom/pairs.pdf>.
- Elliott, R., Van Der Hoek, J., Malcolm, W., 2005. Pairs trading. *Quantitative Finance* 5 (3), 271–276.
- Engle, R. F., Granger, C. W. J., 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55 (2), 251–276.

- Gatev, E., Goetzmann, W., Rouwenhorst, K., 2006. Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies* 19 (3), 797.
- Graham, B., Dodd, D., 2009. *Security Analysis Sixth Edition*. McGraw-Hill, New York.
- Herlemont, D., 2003. Pairs trading, convergence trading, cointegration.
- Jegadeesh, N., Titman, S., 1995. Overreaction, delayed reaction, and contrarian profits. *Review of Financial Studies* 8 (4), 973.
- Jurek, J., Yang, H., 2007. Dynamic portfolio selection in arbitrage. EFA 2006 Meetings paper Paper available at SSRN: <http://ssrn.com/abstract=882536>.
- Nedbank, 2011. Cfd marketinfo, available at http://www.nedbank.co.za/website/content/cfd/marketinfo_CFD_margin.asp.
- Riccardi, L., Sato, S., 1988. First-passage-time density and moments of the ornstein-uhlenbeck process. *Journal of Applied Probability* 25, 43–57.
- Rice, J., 1995. *Mathematical Statistics and Data Analysis Second Edition*. Duxbury Press, Belmont, California.
- Shreve, S., 2004. *Stochastic Calculus for Finance II Continuous Time Models*. Springer Finance, New York, NY.
- StandardBank, 2011. Retail contracts for difference. Available at <https://securities.standardbank.co.za/ost/nsp/FrontOfficePublic/Legacy/onlinereg/contractsForDifference.asp>.
- Stock, J. H., Watson, M. W., 1988. Testing for common trends. *Journal of American Statistical Association* 83 (404), 1097–1107.
- Triantafyllopoulos, K., Montana, G., 2011. Dynamic modeling of mean-reverting spreads for statistical arbitrage. *Computational Management Science* 8 (1-2), 23–49.
- Tsay, R., 2005. *Analysis of Financial Time Series Second Edition*. John Wiley and Sons, New jersey.
- Yuksel, A., Yuksel, A., Muslumov, A., 2010. Pairs trading with turkish stocks. *Middle Eastern Finance and Economics* Issue 7.

Appendix A. Error function

We present the basic definitions of the error function, the complementary error function and the complex error function as in (?). The error function is defined as,

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (\text{A.1})$$

The complementary error function is defined as,

$$erfc(x) = 1 - erf(x) \quad (\text{A.2})$$

$$= \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \quad (\text{A.3})$$

The complex error function is defined as,

$$erfi(x) = e^{-x^2} erfc(-ix) \quad (\text{A.4})$$

Appendix B. Estimation of the Ornstein-Uhlenbeck process

The estimation procedure that follows is as presented in Jurek and Yang (2007). Consider the general OU process,

$$dX_t = \kappa(\mu - X_t)dt + \sigma dW_t \quad (\text{B.1})$$

We can solve this equation using an integrating factor, and assuming that data points occur every Δt time units we can write the solution as Jurek and Yang (2007),

$$X(t_i) = \mu + e^{-\kappa\Delta t}(X(t_{i-1}) - \mu) + \sigma e^{-\kappa\Delta t} \int_{t_{i-1}}^{t_i} e^{\kappa s} dW(s) \quad (\text{B.2})$$

A little manipulation of the terms allows us to write,

$$\frac{X(t_i) - \hat{\mu} - e^{-\hat{\kappa}\Delta t}(X(t_{i-1}) - \hat{\mu})}{\sigma e^{-\kappa\Delta t}} = \int_{t_{i-1}}^{t_i} e^{\kappa s} dW(s) \quad (\text{B.3})$$

Using the properties of brownian motion, and the Ito integral we can write (Jurek and Yang, 2007),

$$\frac{X(t_i) - \hat{\mu} - e^{-\hat{\kappa}\Delta t}(X(t_{i-1}) - \hat{\mu})}{\sigma e^{-\kappa\Delta t}} \sim \sqrt{\int_{t_{i-1}}^{t_i} e^{2\kappa s} ds} \cdot N(0, 1) \quad (\text{B.4})$$

Using (B.4) we can write,

$$\frac{X(t_i) - \hat{\mu} - e^{-\hat{\kappa}\Delta t}(X(t_{i-1}) - \hat{\mu})}{\sigma e^{-\kappa\Delta t}} = \sqrt{\frac{e^{2\kappa\Delta t} - 1}{2\kappa}} \cdot \epsilon \quad (\text{B.5})$$

Where $\epsilon \sim N(0, 1)$. So we can now show that,

$$\frac{X(t_i) - \hat{\mu} - e^{-\hat{\kappa}\Delta t}(X(t_{i-1}) - \hat{\mu})}{\sigma e^{-\kappa\Delta t}} \sqrt{\frac{2\kappa}{e^{2\kappa\Delta t} - 1}} \sim N(0, 1) \quad (\text{B.6})$$

When we combine (B.6) with the independent increments property of brownian motion we have shown that the error terms in (B.1) are independent and indentially distributed. This allows the use of OLS regression of $X(t_i)$ on $X(t_{i-1})$ and a constant to obtain the required estimates as (Jurek and Yang, 2007),

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N X(t_i) \quad (\text{B.7})$$

$$\hat{\kappa} = -\frac{1}{\Delta t} \log \left(\frac{\sum_{i=2}^N (X(t_i) - \hat{\mu})(X(t_{i-1}) - \hat{\mu})}{\sum_{i=2}^N (X(t_i) - \hat{\mu})^2} \right) \quad (\text{B.8})$$

$$\hat{\sigma} = \sqrt{\frac{2\hat{\kappa}}{e^{2\hat{\kappa}\Delta t} - 1} \frac{1}{N-2} \sum_{i=1}^N \left(\frac{X(t_i) - \hat{\mu} - e^{-\hat{\kappa}\Delta t}(X(t_{i-1}) - \hat{\mu})}{e^{-\hat{\kappa}\Delta t}} \right)^2} \quad (\text{B.9})$$

Appendix C. Additional tables and figures

Table C.4: Descriptive statistics for the daily returns in period 1

	Strategy	ALSI
Mean	-0.006963622	-0.000124827
Standard Error	0.003059839	0.000729201
Median	-0.000541126	-0.000103379
Standard Deviation	0.048283439	0.011506592
Sample Variance	0.00233129	0.000132402
Kurtosis	55.56401856	0.986880908
Skewness	-5.794032012	0.423167339
Range	0.665070183	0.079481607
Minimum	-0.516434355	-0.033018345
Maximum	0.148635828	0.046463262
Mean-Variance Ratio	-2.987024646	-0.94278838

Table C.5: Descriptive statistics for the daily returns in period 2

	Strategy	ALSI
Mean	0.003640823	0.001062458
Standard Error	0.000816218	0.000610926
Median	0.001027953	0.000522698
Standard Deviation	0.012879708	0.009640254
Sample Variance	0.000165887	9.29345E-05
Kurtosis	13.01506732	1.038155488
Skewness	2.67113598	0.222041741
Range	0.115019269	0.068072753
Minimum	-0.024344037	-0.03292288
Maximum	0.090675232	0.035149872
Mean-Variance Ratio	21.94762439	11.43232864

Table C.6: Descriptive statistics for the daily returns in period 3

	Strategy	ALSI
Mean	0.003517031	0.001338735
Standard Error	0.000836677	0.000420837
Median	0.002250534	0.001697128
Standard Deviation	0.013202538	0.006640693
Sample Variance	0.000174307	4.40988E-05
Kurtosis	1.417415817	0.604496447
Skewness	0.283478941	-0.409793568
Range	0.088489454	0.042961284
Minimum	-0.039730561	-0.025462671
Maximum	0.048758893	0.017498613
Mean-Variance Ratio	20.17721708	30.35762183

Table C.7: Descriptive statistics for the daily returns in period 4

	Strategy	ALSI
Mean	0.002036672	0.001386256
Standard Error	0.001284266	0.000897207
Median	0.002000471	0.001748551
Standard Deviation	0.020265369	0.014157688
Sample Variance	0.000410685	
Kurtosis	4.150451481	2.807773022
Skewness	-0.360529768	-0.372249901
Range	0.152385878	0.115209963
Minimum	-0.072983874	-0.064808627
Maximum	0.079402004	0.050401336
Mean-Variance Ratio	4.959204853	6.916058346

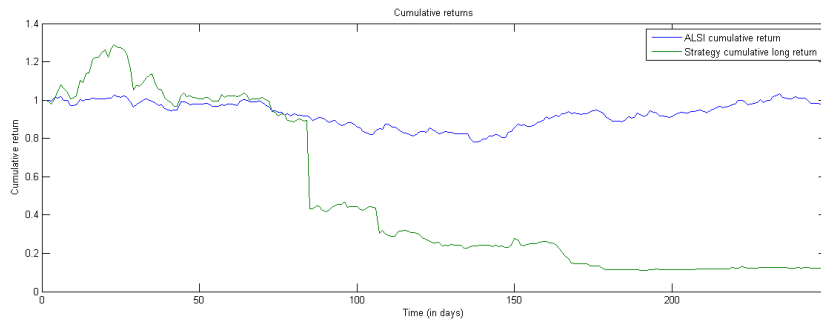


Figure C.8: Daily cumulative excess returns in period 1

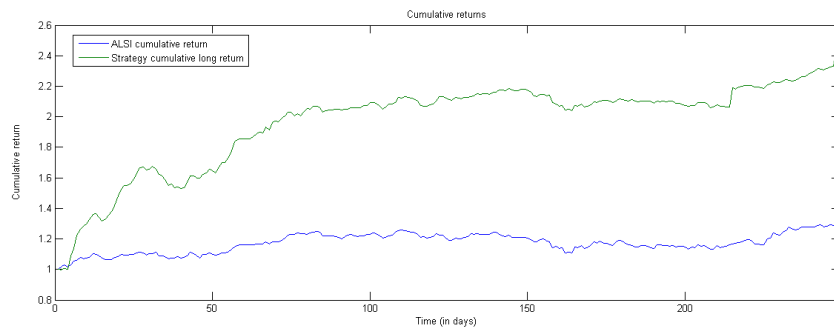


Figure C.9: Daily cumulative excess returns in period 2

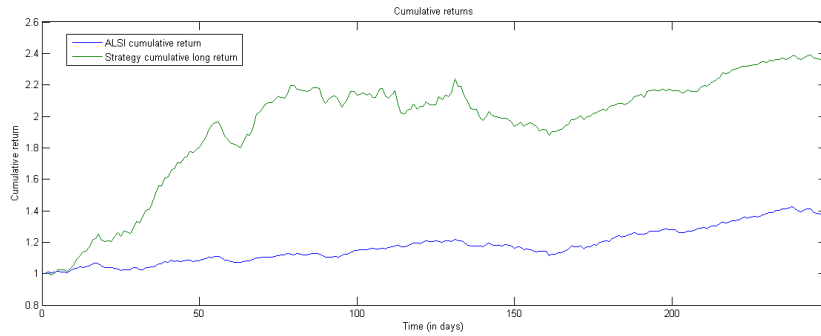


Figure C.10: Daily cumulative excess returns period in 3

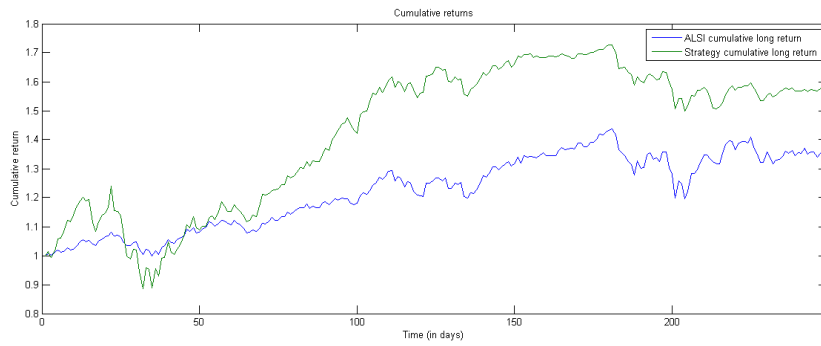


Figure C.11: Daily cumulative excess returns in period 4

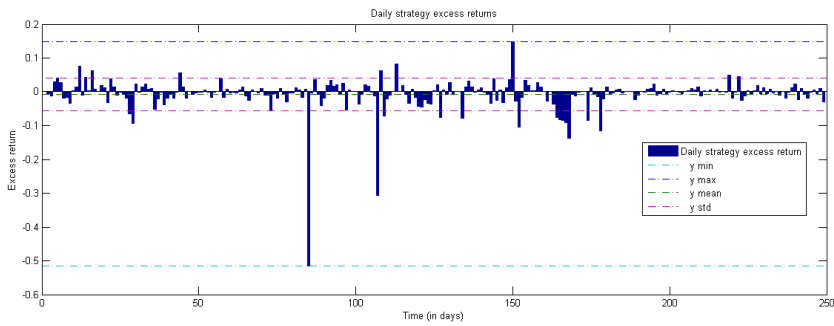


Figure C.12: Daily strategy excess returns in period 1

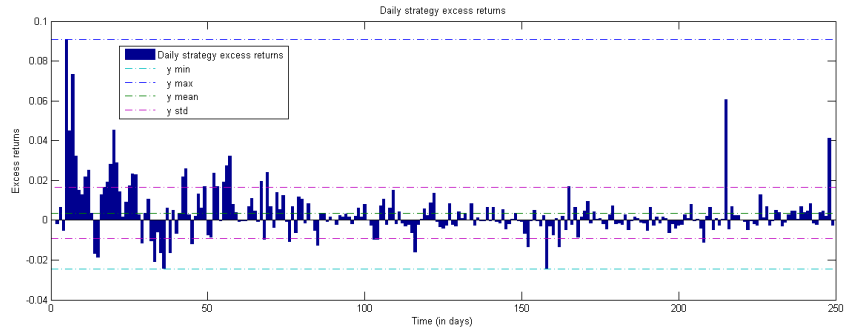


Figure C.13: Daily strategy excess returns in period 2

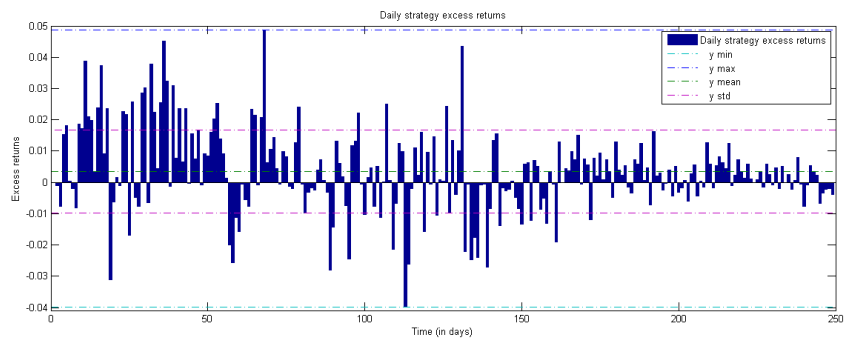


Figure C.14: Daily strategy excess returns in period 3

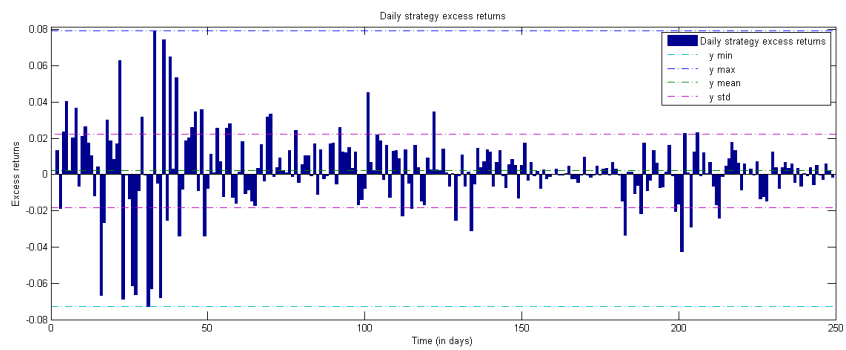


Figure C.15: Daily strategy excess returns in period 4

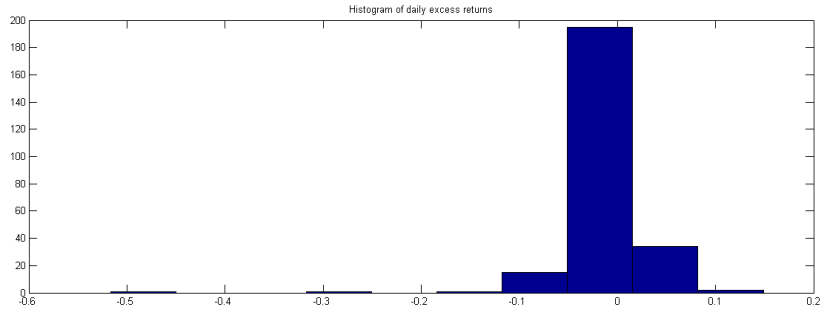


Figure C.16: Histogram of daily strategy excess returns period 1

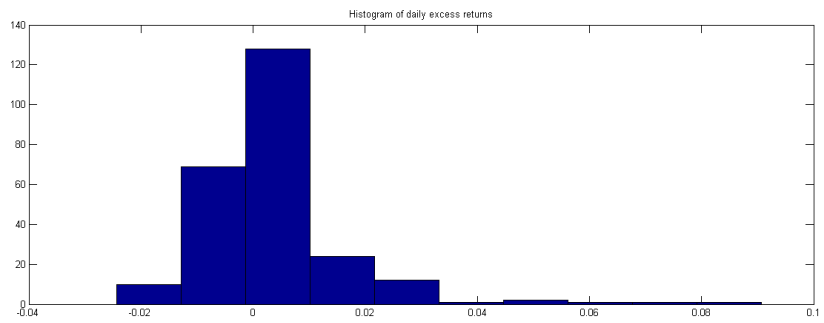


Figure C.17: Histogram of daily strategy excess returns period 2

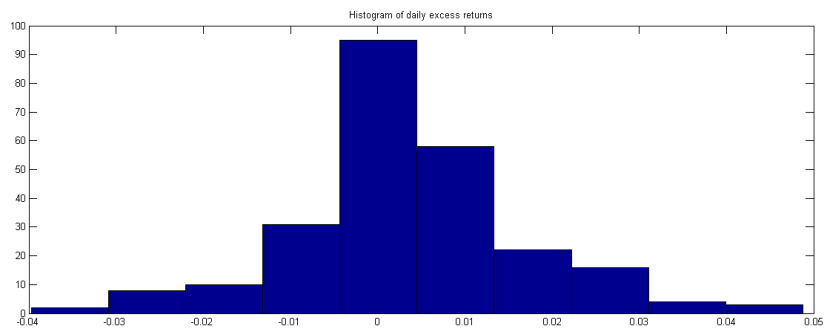


Figure C.18: Histogram of daily strategy excess returns period 3

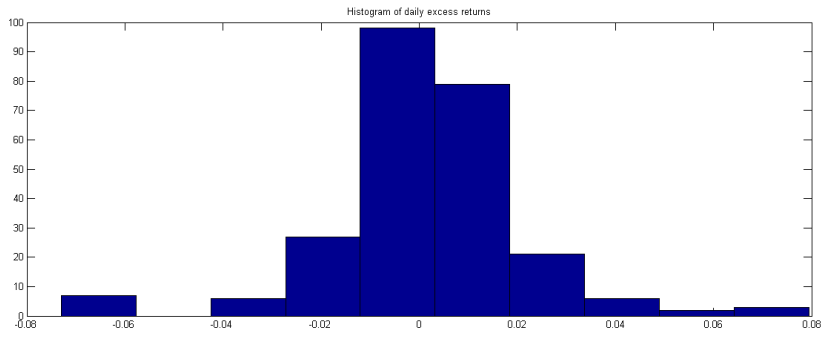


Figure C.19: Histogram of daily strategy excess returns period 4

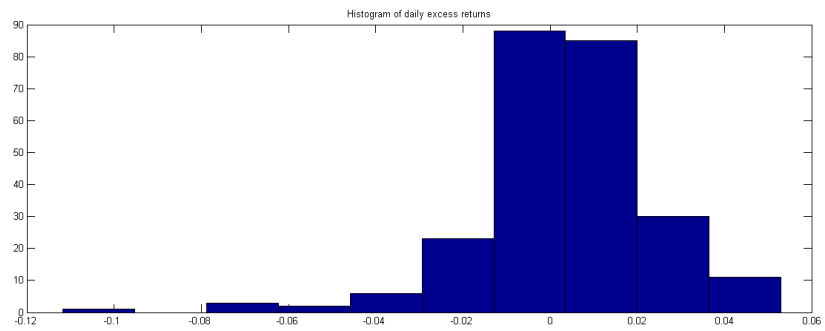


Figure C.20: Histogram of daily strategy excess returns period 5