2009

Finding the Optimal Pre-set Boundaries for Pairs Trading Strategy Based on Cointegration Technique

H. Puspaningrum
University of Wollongong, hp261@uow.edu.au

Y. X. Lin
University of Wollongong, yanxia@uow.edu.au

C. Gulati
University of Wollongong, cmg@uow.edu.au

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

21-09

Finding the Optimal Pre-Set Boundaries for Pairs Trading Strategy Based on Cointegration Technique

Heni Puspaningrum, Yan-Xia Lin, Chandra Gulati

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Finding the Optimal Pre-set Boundaries for Pairs Trading Strategy Based on Cointegration Technique

Heni Puspaningrum[1], Yan−Xia Lin[2] , Chandra Gulati[3]
School of Mathematics and Applied Statistics
University of Wollongong, Australia

Abstract

Pairs trading is one of the arbitrage strategies that can be used in trading stocks on the stock market. It incorporates the use of a standard statistical model to exploit the stocks that are out of equilibrium for short-term time. In determining which two stocks can be a pair, Banerjee *et al.* (1993) shows that the cointegration technique is more effective than correlation criterion for extracting profit potential in temporary pricing anomalies for share prices driven by common underlying factors. This paper explores the ways in which the pre-set boundaries chosen to open a trade can influence the minimum total profit over a specified trading horizon. The minimum total profit relates to the pre-set minimum profit per trade and the number of trades during the trading horizon. The higher the pre-set boundaries for opening trades, the higher the profit per trade but the lower the trade numbers. The opposite applies for lowering the boundary values. The number of trades over a specified trading horizon is determined jointly by the average trade duration and the average inter-trade interval. For any pre-set boundaries, both of these values are estimated by making an analogy to the mean first-passage time. The aims of this paper are to develop numerical algorithm to estimate the average trade duration, the average inter-trade interval, and the average number of trades and then use them to find the optimal pre-set boundaries that would maximize the minimum total profit for cointegration error following an AR(1) process.

Keywords: pairs trading, cointegration, integral equation, the mean first-passage time.

1 Introduction

Pairs trading was first discovered in the early 1980s by the quantitative analyst Nunzio Tartaglia and a team of physicists, computer scientists and mathematicians, who did not have a background in finance. Their idea was to develop statistical

[1]email: hp261@uow.edu.au
[2]email: yanxia@uow.edu.au
[3]email: cmg@uow.edu.au

rules to find ways to perform arbitrage trades, and take the 'skill' out of trading (Gatev *et al.* 1999, 2006).

Pairs trading works by taking the arbitrage opportunity of temporary anomalies between related stocks which have long-run equilibrium. When such an event occurs, one stock will be overvalued relative to the other stock. We can then invest in a two-stock portfolio (a pair) where the overvalued stock is sold (short position) and the undervalued stock is bought (long position). The trade is closed out by taking the opposite position of these stocks after the stocks have settled back into their long-run relationship. The profit is captured from this short-term discrepancies in the two stock prices. Since the profit is not depend on the movement of the market, pairs trading is a market-neutral investment strategy.

According to (Gatev *et al.*,1999, 2006), it appears that the growing popularity of the pairs trading strategy may also pose a problem because the opportunities to trade become much smaller, as many other arbitrageurs are aware of the strategy and may choose to enter at an earlier point of deviation from the equilibrium. The profit from the pairs trading strategy in recent times is less than the profit before the pairs trading strategy is found . However, Gillespie and Ulph (2001), Habak (2002), and Hong and Susmel (2003) show that significant returns could still be made in more recent times with the strategy. An extensive discussion of pairs trading can be found in Gatev *et al* (1999, 2006), Vidyamurthy (2004), Whistler (2004)and Ehrman (2006).

In determining which two stocks can be a pair, people commonly choose two stocks that are highly correlated (see Stone (http://www.investopedia.com), Avery-Wright (http://compareshares.com.au), Goodboy (http://biz.yahoo.com) and Ehr man (2006)). However, Banerjee *et al.* (1993) shows that the cointegration technique is more effective than correlation for extracting profit potential as the cointegration relationship guarantees that the two stocks have a long-run stationary relationship. Gillespie and Ulph (2001), Hong and Susmel (2003), Vidyamurthy (2004) and Herlemont (www.yats.com) also suggest this technique. However, no one has developed pairs trading strategy based on cointegration by quantitatively estimating the average trade duration, the average inter-trade interval, the average number of trades, the minimum total profit, and then finding the optimal pre-set boundaries (thresholds) to open the pair trades. The following paragraphs will briefly explain about these terms and pairs trading base on cointegration.

Substantial literature (see, for example, Fama and French, 1988; Liu *et al.*, 1997; Narayan, 2005; and references cited therein) confirm that stock prices are characterized by a unit root which means the stock prices are I(1) non-stationary time series. Sometimes an appropriate linear combination of two I(1) non-stationary time series could form a stationary time series. If this happens, we say these two I(1) series are cointegrated. [4]

In order to determine whether cointegration exists between two time series there are two techniques that are generally used: the Engle-Granger two-step approach, developed by Engle and Granger (1987), and the technique developed by Johansen (1988). The Engle-Granger approach uses OLS (Ordinary Least Squares) to estimate the long-run steady-state relationship between the variables in the model, and then test whether the residual from the equation is stationary or not. Even though it is quite easy to use, there are some criticisms of this approach, e.g.: (1) this test for cointegration is likely to have lower power than the alternative tests; (2) its finite sample estimates of long-run relationships are potentially biased; and (3) inferences cannot be drawn using standard t-statistics about the significance of the parameters of the static long-run model (Harris, 1995). To overcome the problems found in the Engle-Granger approach, the Johansen's approach uses a vector error-correction model (VECM) so that all variables can be endogenous. More discussion about these two methods can be found in Harris (1995). One more advantage of Johansen's (1988) technique is that it has become available in a user-friendly software, namely, *PcFiml* (version 8), which has been used for running the cointegration analysis in this paper.

The pairs trading strategy, using a cointegration technique, is briefly introduced below :

Consider two shares $S1$ and $S2$ whose prices are I(1). If the share prices $P_{S1,t}$ and $P_{S2,t}$ are cointegrated, there exist cointegration coefficients 1 and β corresponding to $P_{S1,t}$ and $P_{S2,t}$ respectively, such that a cointegration relationship can be constructed as follows:

$$P_{S1,t} - \beta P_{S2,t} = \epsilon_t^*, \tag{1}$$

where ϵ_t^* (the actual cointegration error) is a stationary time series.

Define ϵ_t (the adjusted cointegration error) is as follows:

$$\epsilon_t = \epsilon_t^* - E(\epsilon_t^*), \tag{2}$$

where ϵ_t is also a stationary time series and $E(.)$ means the expectation. The actual cointegration error ϵ_t^* is adjusted so that the mean of the adjusted cointegration error $E(\epsilon_t)$ is zero in order to simplify subsequent analysis.

We have to set an upper-bound $U(U > 0)$ and a lower-bound $L(L < 0)$ before we apply the pairs trading. The function of these boundaries act as a threshold to open a trade. Let N_{S1} and N_{S2} denote the number of shares $S1$ and $S2$ respectively. Two type of trades, U-trades and L-trades, are considered. For a U-trade, a trade is

[4] I(1) means the time series is non-stationary but the first difference is stationary.

opened when the adjusted cointegration error is higher than or equal to the pre-set upper-bound U by selling N_{S1} of $S1$ shares and buying N_{S2} of $S2$ shares and then closing the trade when the adjusted cointegration error is less than or equal to zero. This is done by buying N_{S1} of $S1$ shares and selling N_{S2} of $S2$ shares. The opposite happens for the L-trade, where a trade is opened when the adjusted cointegration error is less than or equal to the pre-set lower-bound L by buying N_{S1} of $S1$ shares and selling N_{S2} of $S2$ shares. The trade is closed when the adjusted cointegration error is higher than or equal to zero by selling N_{S1} of $S1$ shares and buying back N_{S2} of $S2$ shares. It is assumed that the actual cointegration error (ϵ_t^*) as well as the adjusted cointegration error (ϵ_t) are stationary processes and have symmetric distributions, so the lengths from the upper-bound U to the mean and from the lower-bound L to the mean are the same. As a result, the expected number of U-trades and L-trades are the same. For details, see Lin *et al.* (2003, 2006).

In our discussion, the following terms will be required.

- *Trade duration* is the time between opening and closing a U-trade (an L-trade).
- *Inter-trade interval* is the time between two consecutive U-trades (L-trades) or the time between closing a U-trade(an L-trade) and then opening the next U-trade(L-trade). We assume that there is no open trade (neither U-trade nor L-trade) if the previous trade has not been closed yet.
- *Period* is the sum of the trade duration and the inter-trade interval for U-trades (L-trades).

To simplify the discussion in this paper, we subsequently focus mainly on the U-trade case unless stated otherwise. The expected trade durations and the expected inter-trade intervals are estimated to determine the expected number of U-trades over a specified trading horizon. As the expected numbers of U-trades and L-trades are the same, the expected number of U-trades can be doubled to obtain the expected number of trades.

Figure 1 shows two cointegrated shares, i.e. Transonic Travel Ltd (TNS) and Travel.com.au (TVL), and their adjusted cointegration error denoted by *eps* . Both are travel companies listed on the Australian Stock Exchange. In this case, TNS is $S1$ and TVL is $S2$. Further description of the cointegration relationship of these two shares can be found in Section 5. At time t = 5, the adjusted cointegration error of the two stocks (*eps*) is higher than the upper-bound U, so a trade is opened by selling TNS and buying TVL. At t = 14, *eps* is less than the *eps* mean 0 , so the trade is closed by taking the opposite position. Figure 1 also illustrates an example of trade duration, inter-trade interval and period.

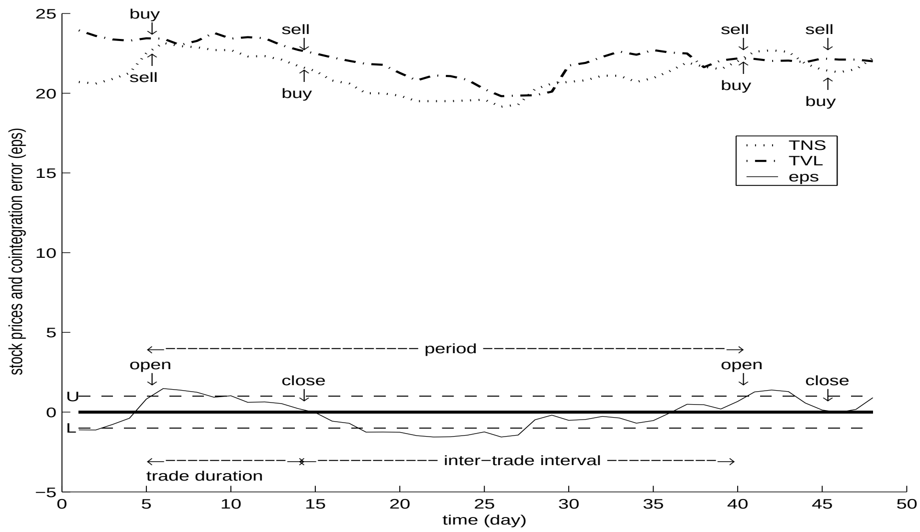

Figure 1: Example of two cointegrated shares (TNS and TVL) with E(eps)=0

Lin *et al.* (2003, 2006) develop a pairs trading strategy based on a cointegration technique called the cointegration coefficients weighted (CCW) rule. The CCW rule works by trading the number of $S1$ and $S2$ shares as a proportion of cointegration coefficients to achieve a pre-set minimum profit per trade. The pre-set minimum profit per trade corresponds to the pre-set boundaries U and L chosen to open trades. However, they did not discuss the optimality issue on the pre-set boundaries. Developing a numerical algorithm to calculate the optimal pre-set boundary values will be the main target of this paper.

We determine the optimality of the pre-set boundary values by maximizing the minimum total profit (MTP) over a specified trading horizon. The MTP corresponds to the pre-set minimum profit per trade and the number of trades during the trading horizon. As the derivation of the pre-set minimum profit per trade is already provided in Lin *et al.* (2003, 2006), this paper will provide the estimated number of trades. The number of trades is also influenced by the distance of the pre-set boundaries from the long-run cointegration equilibrium. The higher the pre-set boundaries for opening trades, the higher the minimum profit per trade but the lower the trade numbers. The opposite applies for lowering the boundary values.

The number of trades over a specified trading horizon is determined jointly by the average trade duration and the average inter-trade interval. For any pre-set boundaries, both of those values are estimated by making an analogy to the mean first-passage times for an AR(1) process. This paper applies an integral equation approach to evaluate the mean first-passage times from Basak and Ho (2004).

The paper is organized as follows. Section 2 gives a brief summary of the trading rules to obtain the pre-set minimum profit per trade. In Section 3, we give a brief description of the mean first-passage time of an AR(1) process using an integral equation approach and apply the concepts to estimate the average trade duration, the average inter-trade interval and then the number of trades in the pairs trading strategy. In Section 4, a numerical algorithm is developed to calculate the optimal pre-set upper-bound, denoted U_o, that would maximize the minimum total profit. Section 5 provides two empirical examples, i.e. BHP-RIO and TNS-TVS and the last section has discussion and a conclusion.

2 Minimum Profit Per Trade

This section will explain how to determine the number of shares of $S1$ and $S2$ needed to get the pre-set minimum profit per trade. Using Eqs.(1) and (2), this paper follows the derivation of the minimum profit per trade as in Lin *et al.* (2003, 2006). Consider the following assumptions.

1. The two share price series are cointegrated over the relevant time period.
2. Long (buy) and short (sell) positions always apply to the same shares in the share-pair.
3. Short sales are permitted or possible through a broker and there is no interest charged for the short sales and no cost for trading.
4. $\beta > 0$

Assumptions 1 and 2 are fairly non-controversial. The others assumptions are applied to simplify the analysis. To support the fourth assumption, we have examined seven share pairs (ANZ-ADB, ABC-HAN, ABC-BLD, CCL-CHB, HAN-RIN, BHP-RIO, and TNS-TVL)[5] from the Australian Stock Exchange using daily data for 2004 (www.finance.yahoo.com.au) and find that the β's for those cointegrated shares were positive.

2.1 U-trades

Consider two cointegrated shares, $S1$ and $S2$ as in Eq.(1). By using Assumption 1, we can conclude that

[5]ANZ Banking Group Ltd (ANZ), Adelaide Bank (ADB), Adelaide Brighton (ABC), Boral Ltd (BLD), BHP Billiton Ltd (BHP), Coca-cola Amatil (CCL), Coca-cola Hellenic (CHB), Hanson Plc (HAN), Rinker Group Ltd (RIN), Rio Tinto Ltd (RIO), Transonic Travel (TNS), Travel.com.au (TVL)

- When $\epsilon_t \geq U$, the price of one unit share $S1$ is higher than or equal to the price of β unit shares $S2$, relative to their equilibrium relationship. In other words, $S1$ is overvalued while $S2$ is undervalued, relative to their equilibrium relationship. A trade is opened at this time. Let t_o represent the time of opening a trade position.

- If $\epsilon_t \leq 0$, the price of one unit share $S1$ is less than or equal the price of β unit shares $S2$, relative to their equilibrium relationship. In other words, $S1$ is under-valued while $S2$ is over-valued according to their equilibrium relationship. The trade is closed at this time. Let t_c represent the time of closing out a trade position.

When the adjusted cointegration error is higher than or equal to the pre-set upper-bound U at time t_o, a trade is opened by selling N_{S1} of $S1$ shares at time t_o for $N_{S1}P_{S1,t_0}$ dollars and buying N_{S2} of $S2$ at time t_o for $N_{S2}P_{S2,t_o}$ dollars.

When the adjusted cointegration error has settled back to its mean at time t_c, the positions are closed out by simultaneously selling the long position shares for $N_{S2}P_{S2,t_c}$ dollars and buying back the N_{S1} of $S1$ shares for $N_{S1}P_{S1,t_c}$ dollars.

Profit per trade will be

$$P = N_{S2}(P_{S2,t_c} - P_{S2,t_o}) + N_{S1}(P_{S1,t_o} - P_{S1,t_c}). \tag{3}$$

According to the CCW rule as in Lin *et al.* (2003, 2006), if the weight of N_{S2} and N_{S1} are chosen as a proportion of the cointegration coefficients, i.e. $N_{S1} = 1$ and $N_{S2} = \beta$, the minimum profit per trade can be determined as follows: [6]

$$\begin{aligned} P &= N_{S2}(P_{S2,t_c} - P_{S2,t_o}) + N_{S1}(P_{S1,t_o} - P_{S1,t_c}) \\ &= \beta[P_{S2,t_c} - P_{S2,t_o}] + [P_{S1,t_o} - P_{S1,t_c}] \\ &= \beta[P_{S2,t_c} - P_{S2,t_o}] + [(\epsilon_{t_o} + E(\epsilon_t^*) + \beta P_{S2,t_o}) - (\epsilon_{t_c} + E(\epsilon_t^*) + \beta P_{S2,t_c})] \\ &= (\epsilon_{t_o} - \epsilon_{t_c}) \geq U. \end{aligned} \tag{4}$$

Thus, by trading the shares with the weight as a proportion of the cointegration coefficients, the profit per trade is at least U dollars.

2.2 L-trades

For an L-trade, the pre-set lower-bound L can be set to be $-U$. So, a trade is opened when $\epsilon_t \leq -U$ by selling $S2$ and buying $S1$.

[6]For simplicity, fractional share holdings are permitted

Profit per trade will be:

$$P = N_{S2}(P_{S2,t_o} - P_{S2,t_c}) + N_{S1}(P_{S1,t_c} - P_{S1,t_o}). \tag{5}$$

Analogous to the derivation of minimum profit per trade for an U-trade, let $N_{S2} = \beta$ and $N_{S1} = 1$. Thus,

$$\begin{aligned} P &= \beta[P_{S2,t_o} - P_{S2,t_c}] + [P_{S1,t_c} - P_{S1,t_o}] \\ &= \beta[P_{S2,t_o} - P_{S2,t_c}] + [(\epsilon_{t_c} + E(\epsilon_t^*) + \beta P_{S2,t_c}) - (\epsilon_{t_o} + E(\epsilon_t^*) + \beta P_{S2,t_o})] \\ &= (\epsilon_{t_c} - \epsilon_{t_o}) \geq U. \end{aligned} \tag{6}$$

So, trading 1 unit share $S1$ and β unit shares $S2$, either in U-trades or L-trades would make a minimum profit per trade as much as U. However, for L-trades we need to borrow some money because with ϵ_t negative at opening time means that the income from the short sales (selling β unit shares $S2$) is insufficient to buy 1 unit share $S1$.

3 Mean First-passage Time of an AR(1) Process and Pairs Trading

As a stationary process, the actual cointegration error (ϵ_t^*) as well as the adjusted cointegration error (ϵ_t) may follow linear stationary processes (e.g.: White noise, Autoregressive, Moving average, and Autoregressive-Moving Average processes), non-linear stationary processes or other stationary processes. We have examined seven share pairs (ANZ-ADB, ABC-HAN, ABC-BLD, CCL-CHB, HAN-RIN, BHP-RIO, and TNS-TVL) from the Australian Stock Exchange using daily data for 2004 (www.finance.yahoo.com.au). All of these share pairs produce cointegration error with AR(1) processes. Elliott (2005) and Herlemont (www.yats.com) also suggested AR(1) processes for modeling pairs trading, but they used the Ornstein-Uhlenbeck process which is the continuous-time counterpart of an AR(1) process to estimate the optimal boundaries. However, due to the complexity of stochastic analysis in the Ornstein-Uhlenbeck process, their results are difficult to be applied in practical situation. Therefore, in this paper we focus on an AR(1) process and use an integral equation approach from Basak and Ho (2004) which is more practicable than the Ornstein-Uhlenbeck process .

This section will provide steps to obtain an estimation of the number of trades over a specified trading horizon. Firstly, we will give a brief summary of the mean first-passage time of AR(1) process using an integral equation approach from Basak and Ho (2004). Secondly, a numerical scheme is provided to calculate the mean first-passage time of an AR(1) process using an integral equation approach. Thirdly, the average trade duration and the average inter-trades interval are estimated using an analogy of the mean first-passage time. Fourthly, the number of trades over a specified trading horizon is approximated using the average trade duration and the average inter-trade interval.

3.1 The mean first-passage time of an AR(1) process using an integral equation approach

Consider an AR(1) process:

$$Y_t = \phi Y_{t-1} + \xi_t, \tag{7}$$

where $-1 < \phi < 1$ and $\xi_t \sim$ i.i.d $N(0, \sigma_\xi^2)$.

The first-passage time $\mathcal{T}_{a,b}(y_0)$ is defined as

$$\mathcal{T}_{a,b}(y_0) = \inf\{t : Y_t > b \text{ or } Y_t < a | a \leq Y_0 = y_0 \leq b\} \tag{8}$$

Particularly,

$$\mathcal{T}_a(y_0) = \mathcal{T}_{a,\infty}(y_0) = \inf\{t : Y_t < a | Y_0 = y_0 \geq a\} \tag{9}$$

and

$$\mathcal{T}_b(y_0) = \mathcal{T}_{-\infty,b}(y_0) = \inf\{t : Y_t > b | b \geq Y_0 = y_0\} \tag{10}$$

$E(\mathcal{T}_{a,b}(y_0))$, $E(\mathcal{T}_a(y_0))$, and $E(\mathcal{T}_b(y_0))$ denote the mean first-passage time of $\mathcal{T}_{a,b}(y_0)$, $\mathcal{T}_a(y_0)$, and $\mathcal{T}_b(y_0)$ respectively. Basak and Ho (2004) derive the mean first-passage time of an AR(1) process using an integral equation approach.

We define a discrete-time real-valued Markov process $\{Y_t\}$ on a probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$ with stationary continuous transition density $f(y|x)$, continuous in both x and y. The term $f(y|x)$ denotes the transition density of reaching y at the next step given that the present state is x. Suppose that $Y_0 = y_0 \in [a, b]$. The mean first-passage time over interval $[a, b]$ of an AR(1) process, starting at initial state $y_0 \in [a, b]$, is given by

$$E(\mathcal{T}_{a,b}(y_0)) = \int_a^b E(\mathcal{T}_{a,b}(u)) f(u|y_0) du + 1. \tag{11}$$

For an AR(1) process in Eq.(7), $f(u|y_0)$ will be a normal distribution with mean ϕy_0 and variance σ_ξ^2. Thus,

$$E(\mathcal{T}_{a,b}(y_0)) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_a^b E(\mathcal{T}_{a,b}(u)) \exp\left(-\frac{(u - \phi y_0)^2}{2\sigma_\xi^2}\right) du + 1. \tag{12}$$

Details of the derivation can be found in Basak and Ho (2004). The integral equation in Eq.(12) is a Fredholm type of the second kind and can be solved numerically using the Nystrom method (Atkinson, 1997) as in the next subsection.

3.2 Numerical scheme

If we want to calculate $E(\mathcal{T}_b(y_0))$, that is, the mean first-passage time over a given level b of an AR(1) process starting at initial state y_0, it can be computed by adding a lower boundary a first. Since $E(\mathcal{T}_{a,b}(y_0))$ converges monotonically to $E(\mathcal{T}_b(y_0))$ as $a \to -\infty$, the approximation of $E(\mathcal{T}_b(y_0))$ can be obtained by evaluating $E(\mathcal{T}_{a,b}(y_0))$ as $a \to -\infty$ instead.

Consider $E(\mathcal{T}_{a,b}(y_0))$ as in Eq.(12). Now, define $h = (b-a)/n$, where n is the number of partitions in $[a,b]$ and h is the length of each partition.

Using the trapezoid integration rule (Atkinson, 1997):

$$\int_a^b f(u)du \approx \frac{h}{2}\left[w_0 f(u_0) + w_1 f(u_1) + \cdots + w_{n-1} f(u_{n-1}) + w_n f(u_n)\right], \tag{13}$$

where $u_0 = a, u_i = a + ih, u_n = b, i = 1, \ldots, n$ and the weights w_i for the corresponding nodes are

$$w_i = \begin{cases} 1, & \text{for } i = 0 \text{ and } i = n \\ 2, & \text{for others} \end{cases}$$

Thus, the integral term in Eq.(12) can be approximated by

$$\int_a^b E(\mathcal{T}_{a,b}(u)) \exp\left(-\frac{(u-\phi y_0)^2}{2\sigma_\xi^2}\right) du \approx \frac{h}{2}\sum_{j=0}^{n} w_j E(\mathcal{T}_{a,b}(u_j)) \exp\left(-\frac{(u_j-\phi y_0)^2}{2\sigma_\xi^2}\right), \tag{14}$$

.

Let $E_n(\mathcal{T}_{a,b}(y_0))$ denote the approximation of $E(\mathcal{T}_{a,b}(y_0))$ using n partitions. Thus, the expectation in Eq.(12) using n partitions can be estimated by

$$E_n(\mathcal{T}_{a,b}(y_0)) \approx \frac{h}{2\sqrt{2\pi}\sigma_\xi}\sum_{j=0}^{n} w_j E_n(\mathcal{T}_{a,b}(u_j)) \exp\left(-\frac{(u_j-\phi y_0)^2}{2\sigma_\xi^2}\right) + 1. \tag{15}$$

Set y_0 as u_i for $i = 0, 1, \ldots, n$ and reformulate Eq.(15) as follows

$$E_n(\mathcal{T}_{a,b}(u_i)) - \sum_{j=0}^{n} \frac{h}{2\sqrt{2\pi}\sigma_\xi} w_j E_n(\mathcal{T}_{a,b}(u_j)) \exp\left(-\frac{(u_j-\phi u_i)^2}{2\sigma_\xi^2}\right) = 1, \tag{16}$$

and then solve the following linear equations in (17) to obtain an approximation of $E_n(\mathcal{T}_{a,b}(u_j))$.

$$\begin{pmatrix} 1-K(u_0,u_0) & -K(u_0,u_1) & \ldots & -K(u_0,u_n) \\ -K(u_1,u_0) & 1-K(u_1,u_1) & \ldots & -K(u_1,u_n) \\ \vdots & \vdots & \vdots & \vdots \\ -K(u_n,u_0) & -K(u_n,u_1) & \ldots & 1-K(u_n,u_n) \end{pmatrix} \begin{pmatrix} E_n(\mathcal{T}_{a,b}(u_0)) \\ E_n(\mathcal{T}_{a,b}(u_1)) \\ \vdots \\ E_n(\mathcal{T}_{a,b}(u_n)) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{17}$$

Table 1: Mean first-passage time of level 0, given $y_0 = 1.5$ for $y(t)$ in (7)

σ_ξ^2	ϕ	Integral equation	Simulation
0.49	0.5	3.9181 (b = 5, n= 50)	3.9419
	0	2 (b = 5, n = 50)	1.9918
	-0.5	1.2329 (b=5, n= 50)	1.2341
1	0.5	3.5401 (b = 7, n= 70)	3.5571
	0	2 (b = 7, n = 70)	2.0055
	-0.5	1.3666 (b=7, n= 70)	1.3636
4	0.5	3.0467 (b = 14, n= 140)	3.0512
	0	2 (b = 14, n = 140)	1.9959
	-0.5	1.5626 (b=14, n= 140)	1.5725

where

$$K(u_i, u_j) = \frac{h}{2\sqrt{2\pi}\sigma_\xi} w_j \exp\left(-\frac{(u_j - \phi u_i)^2}{2\sigma_\xi^2}\right).$$

Examples of numerical results for some AR(1) processes are provided in Table 1. We compare the results using an integral equation approach and simulation. For given σ_ξ^2, ϕ, and $y_0 = 1.5$, the mean first-passage time of level zero, using an integral equation approach, is calculated. We use different b and n in order to make the length of partition h the same for each case. The results show that $h = 0.1$ is enough to get results similar to the simulation. For simulation, we generate an AR(1) process as in Eq.(7) for given σ_ξ^2, and ϕ. Using the initial state $y_0 = 1.5$, the time needed for the process to cross zero for the first time is calculated. The simulation is repeated 1000 times and then we calculate the average. The table shows that the simulation results confirm the results from the integral equation approach.

3.3 Trade durations and inter-trade intervals

Consider the adjusted cointegration error and assume that ϵ_t in Eq.(2) follows an AR(1) process, i.e:

$$\epsilon_t = \phi\epsilon_{t-1} + a_t, \text{ where } a_t \sim \text{ i.i.d } N(0, \sigma_a^2). \tag{18}$$

As explained in Section 1, the trade duration is the time between opening and closing a trade. For a U-trade, a trade is opened when ϵ_t is higher than or equal to the pre-set upper-bound U and it is closed when ϵ_t is less than or equal to 0 which is the mean of the adjusted cointegration error. Suppose ϵ_t is at U , so a U-trade is opened. To calculate the expected trade duration, we would like to know the time needed on average for ϵ_t to pass 0 for the first time. Thus, calculating the expected trade duration is the same as calculating the mean first-passage time for ϵ_t to pass 0 for the first time, given the initial value is U. Let TD_U denote the expected trade duration corresponding to the pre-set upper-bound U. Using Eq.(12), TD_U is defined as follows:

$$TD_U := E(\mathcal{T}_{0,\infty}(U)) = \lim_{b\to\infty} \frac{1}{\sqrt{2\pi}\sigma_a} \int_0^b E(\mathcal{T}_{0,b}(s)) \exp\left(-\frac{(s-\phi U)^2}{2\sigma_a^2}\right) ds + 1. \quad (19)$$

As for trade duration, the inter-trade interval is the waiting time needed to open a trade after the previous trade is closed. For a U-trade, if there is an open U-trade while ϵ_t is at 0 during trading, the trade has to be closed. To calculate the expected inter-trade interval, we would like to know the time needed on average for ϵ_t to pass the pre-set upper-bound U for the first time, so we can open a U-trade again. Thus, calculating the expected inter-trade interval is the same as calculating the mean first-passage time for ϵ_t to pass U given the initial value is 0. Let I_U denote the expected inter-trade interval for the pre-set upper-bound U.

$$I_U := E(\mathcal{T}_{-\infty,U}(0)) = \lim_{-b\to-\infty} \frac{1}{\sqrt{2\pi}\sigma_a} \int_{-b}^U E(\mathcal{T}_{-b,U}(s)) \exp\left(-\frac{s^2}{2\sigma_a^2}\right) ds + 1. \quad (20)$$

3.4 Number of trades over a trading horizon

The expected number of U-trades $E(N_{UT})$ and the expected number of periods corresponding to U-trades $E(N_{UP})$ over a time horizon [0,T] are defined as follow:

$$E(N_{UT}) = \sum_{k=1}^{\infty} kP(N_{UT} = k)$$

and

$$E(N_{UP}) = \sum_{k=1}^{\infty} kP(N_{UP} = k)$$

.

In this subsection, we want to derive the expected number of U-trades $E(N_{UT})$ over a specified trading horizon. However, it is difficult to evaluate the exact value of $E(N_{UT})$. Thus, a possible range of values of $E(N_{UT})$ is provided.

As explained in Section 1, $period_U$ is defined as the sum of the trade duration and the inter-trade interval for U-trades. Thus, the expected $period_U$ is given by,

$$E(period_U) = TD_U + I_U.$$

First, we will evaluate the expected number of $period_U$'s $E(N_{UP})$ in the time horizon [0,T] as it has a direct connection to the trade duration and the inter-trade interval. Then, the relationship of N_{UT} and N_{UP} will be used to obtain a possible range of values of $E(N_{UT})$.

Let $period_{Ui}$ denote the length of the period corresponding to the ith U-trade. Thus,

$$T \geq E\left(\sum_{i=1}^{N_{UP}} (Period_{Ui})\right) = \sum_{k=1}^{\infty}\left[\sum_{i=1}^{k} E(Period_{Ui})\right] P(N_{UP} = k). \quad (21)$$

Since the period depends on the distribution of ϵ_t, which is a stationary time series, $E(period_{Ui})$ will be the same for all i. Thus, $E(period_{Ui}) = E(period_U)$ and

$$T \geq E(period_U) \sum_{k=1}^{\infty} kP(N_{UP} = k) = E(period_U)E(N_{UP}). \quad (22)$$

Thus,

$$E(N_{UP}) \leq \frac{T}{E(period_U)} = \frac{T}{TD_U + I_U}. \quad (23)$$

As for the derivation that leads to (23),

$$T < E\left(\sum_{i=0}^{N_{UP}+1} (period_{Ui})\right) = E(Period_U)E(N_{UP} + 1), \quad (24)$$

giving

$$E(N_{UP}) > \frac{T}{E(Period_U)} - 1 = \frac{T}{TD_U + I_U} - 1. \quad (25)$$

Thus,

$$\frac{T}{TD_U + I_U} \geq E(N_{UP}) > \frac{T}{TD_U + I_U} - 1. \quad (26)$$

However, the relationship between number of U-trades (N_{UT}), and number of $period_U$'s (N_{UP}) is $N_{UT} = N_{UP}$ or $N_{UT} = N_{UP} + 1$.

Thus,

$$\frac{T}{TD_U + I_U} + 1 \geq E(N_{UP}) + 1 \geq E(N_{UT}) \geq E(N_{UP}) > \frac{T}{TD_U + I_U} - 1. \quad (27)$$

Table 2 shows the estimation of the number of U-trades over T = 1000 observations for some AR(1) processes using the theory presented above. We use $\hat{N}_{UT} = \frac{1000}{TD_U + I_U} - 1$ to estimate the expected number of U-trades within [0,T]. The average trade duration for U-trades TD_U and the average inter-trade interval for U-trades I_U are calculated using the integral equation approach.

Table 3 shows the simulation results of the number of U-trades as a comparison to the theoretical results in Table 2. 1000 observations are generated from the model described in Eq.(18) for each simulation and each simulation is independently

Table 2: Estimation of the number of U-trades using an integral equation approach with $U = 1.5$

ϕ	σ_a^2	TD_U	I_U	$\hat{N}_{UT} = \frac{1000}{TD_U + I_U} - 1$
0.5	0.49	3.9181	40.6074	21.459
	1	3.5401	14.6006	54.125
	4	3.0469	5.5679	115.079
-0.5	0.49	1.2329	32.6253	28.535
	1	1.3666	10.523	83.1071
	4	1.5626	3.6220	191.879

Table 3: Simulated number of trades for an AR(1) process using T = 1000 observations and $U = 1.5$

ϕ	σ_a^2	TD_U	I_U	N_{UT} simulation	$\hat{N}_{UT} = \frac{1000}{TD_U + I_U} - 1$
0.5	0.49	4.054(0.585)	42.801(12.340)	21.725(5.277)	20.342
	1	3.780(0.308)	15.153(1.838)	52.650(5.226)	51.817
	4	3.407(0.255)	6.254(0.466)	103.000(6.421)	102.508
-0.5	0.49	1.170(0.084)	28.958(4.760)	32.025(5.091)	32.191
	1	1.242(0.072)	9.206(0.952)	95.175(8.311)	94.706
	4	1.385(0.051)	3.030(0.151)	225.500(8.741)	225.500

repeated 40 times. The values in parentheses are the standard deviations. In calculating the trade duration for each simulation, we start to open a trade when ϵ_t exceeds U. In calculating the inter-trade interval, the trade is closed when ϵ_t goes below zero. This is done because in the simulation, ϵ_t is a discrete time process. Thus, it is hard to obtain the exact time for ϵ_t at U and 0. We calculate the average trade duration TD_U, the average inter-trade interval I_U and the number of U-trades N_{UT} for each simulation. At the end of all 40 repeated simulations, we calculate the mean of TD_U, I_U and N_{UT} from all simulations as well as the standard deviations. Furthermore, the last column shows the number of trades using $\hat{N}_{UT} = \frac{1000}{TD_U + I_U} - 1$. From Table 3, we can conclude that if we can estimate the average of trading duration and the average of intra-trade interval correctly, the formula $\hat{N}_{UT} = \frac{1000}{TD_U + I_U} - 1$ can be used to estimate the number of U-trades.

Comparing the number of U-trades results in Tables 2 and 3, we see that for $\phi = 0.5$, the estimates of the number of U-trades using the integral equation are higher than those given by the simulation results. The opposite happens if $\phi = -0.5$. The difference is due to a slight difference in the framework underpinning the theory of integral equations and that for simulation from real data.

4 Minimum Total Profit and the Optimal Pre-set Upper-bound

This section will combine the pre-set minimum profit per trade from Section 2 and the number of U-trades from Section 3 to define minimum total profit (MTP) over the time horizon [0,T]. The optimal pre-set upper-bound, denoted by U_o, is determined by maximizing the MTP.

Let TP_U denote the total profit from U-trades within the time horizon [0,T] for a pre-set upper-bound U. Thus,

$$TP_U = \sum_{i}^{N_{UT}} (\text{ Profit from the } i\text{th U-trade}).$$

Using Eqs.(4) and (27),

$$\text{Profit per trade} \geq U$$

and

$$E(N_{UT}) \geq \frac{T}{TD_U + I_U} - 1.$$

Table 4: Numerical results in determining optimal U

σ_a^2	$\phi = -0.8$		$\phi = -0.5$		$\phi = -0.2$	
	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o
0.25	91.7097	0.59	77.0414	0.5	66.4935	0.47
0.49	128.3609	0.83	107.8254	0.7	93.0673	0.65
1	183.3448	1.19	154.0117	1	132.9287	0.93
2.25	274.9967	1.78	230.9996	1.49	199.3710	1.4
4	366.6515	2.37	307.9922	1.99	265.8216	1.86
		$\approx 1.2\sigma_a$		$\approx \sigma_a$		$\approx 0.93\sigma_a$
		$\approx 0.72\sigma_\epsilon$		$\approx 0.87\sigma_\epsilon$		$\approx 0.91\sigma_\epsilon$

σ_a^2	$\phi = 0.2$		$\phi = 0.5$		$\phi = 0.8$	
	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o
0.25	55.1798	0.47	46.7138	0.53	34.7004	0.7
0.49	77.219	0.66	65.3655	0.74	48.5545	0.97
1	110.2877	0.95	93.3549	1.05	69.3438	1.39
2.25	165.4104	1.42	140.0095	1.58	103.9991	2.09
4	220.5361	1.89	186.6704	2.1	138.6582	2.78
		$\approx 0.95\sigma_a$		$\approx 1.05\sigma_a$		$\approx 1.4\sigma_a$
		$\approx 0.93\sigma_\epsilon$		$\approx 0.91\sigma_\epsilon$		$\approx 0.84\sigma_\epsilon$

Define the minimum total profit with the time horizon [0,T] by

$$MTP(U) := \left(\frac{T}{TD_U + I_U} - 1\right) U.^{7} \tag{28}$$

Then, considering all $U \in [0, b]$, the optimal pre-set upper-bound U_o is chosen such that $MTP(U_o)$ takes the maximum at that U_o. In practice, the value of b is set up as $5\sigma_\epsilon$ because ϵ_t is a stationary process, and the probability that $|\epsilon_t|$ is greater than $5\sigma_\epsilon$ is close to zero.[8]

The numerical algorithm to calculate the optimal pre-set upper-bound U is as follows:

1. Set up the value of b as $5\sigma_\epsilon$.
2. Decide a sequence of pre-set upper-bounds U_i, where $U_i = i \times 0.01$, and $i = 0, \ldots, b/0.01$.
3. For each U_i,
   (a) calculate $E(\mathcal{T}_{0,b}(U_i))$ as the trade duration (TD_{U_i}) using Eq.(19).
   (b) calculate $E(\mathcal{T}_{-b,U_i}(0))$ as the inter-trade interval (I_{U_i}) using Eq.(20).
   (c) calculate $MTP(U_i) = \left(\frac{T}{TD_{U_i}+I_{U_i}} - 1\right) U_i$.
4. Find $U_o \in \{U_i\}$ such that $MTP(U_o)$ is the maximum .

Examples of numerical results from some AR(1) processes are shown in Table 4. We use the model of an AR(1) process described in Eq.(18) and $T = 1000$. The table shows that for a given ϕ, U_o increases as σ_a increases. The last two rows of each ϕ show the approximation of U_o as a proportion of σ_a and σ_ϵ. Those approximations can be used as a general rule in choosing U_o. For example if we have the adjusted cointegration error ϵ_t with an AR(1) process and the ϕ is -0.5 or 0.5, quickly we can choose $U_o = \sigma_a$.

The MTP can be used as a criteria to determine whether the stock pairs are worth to be traded. If we have limited funding to trade stocks in the market, and we have identified several stock pairs, we can choose the stock pair that give the maximum MTP.

5 Empirical examples

This section will investigate the application of the above pairs trading strategy. Since we do not apply real pairs trading in the stock market, we use empirical data

[7]We adopt the notation $MTP(U)$ since the Minimum Total Profit is a function of U.

[8]σ_ϵ is the standard deviation of ϵ_t

available in the internet (www.finance.yahoo.com.au). The empirical data is divided into two parts, namely in-sample data and out-sample data. The in-sample data is assigned as training period where we analysis the cointegration relationship and then determine the optimal pre-set optimal boundaries U_o and L_o. The out-sample data is assigned as trading period. The out-sample data is assumed still hold the same cointegration relationship with the in-sample data, so the pairs trading strategy can be applied to the out-sample data using the optimal pre-set optimal boundaries U_o and L_o obtained from the in-sample data.

There is no standard rule to choose how long the training period (in-sample data) and trading period (out-sample data) needed. However, the training period needs to be long enough so that we can determine that a cointegration relationship actually exists, but not so long that it is obsolete for the trading period. For trading period, it needs to be long enough to have opportunities to open and close trades and test the strategy, but it can not too long because it is possible that the cointegration relationship between the two stocks may change. We use 12-month training period and 6-month trading period with daily data as these periods correspond with the other study by Gatev *et al.*(1999, 2006), Gillespie and Ulph (2001) and Habak (2002).

This paper give two specific illustrations, BHP-RIO and TNS-TVL on the Australian Stock Exchange (ASX). The stocks of BHP-RIO and TNS-TVL are cointegrated and the cointegration error can be fitted with the AR(1) model. We use PcFiml (Doornik and Hendry, 1997) and PcGive (Hendry and Doornik, 1996) softwares to analyze the cointegration relationship of the data.

From the in-sample data, knowing that BHP-RIO and TNS-TVL are cointegrated and the cointegration error are AR(1) processes, the values of ϕ and σ_a can be estimated. The algorithm in Section 4 is applied to obtain the estimates of the optimal pre-set upper-bound U_o, the number of U-trades, the expected of trade duration and the estimates of the minimum total profits from U-trades for each pair of shares. As we have explained before, the number of trades and the minimum total profits, produced by the algorithm in Section 4 ,are for U-trades during the time horizon [0,T] only. As the ϵ_t from those share pairs are stationary processes and have symmetric distributions, in considering the L-trades, we can simply take the total number of trades and the total profit to be double the results from the algorithm above and the estimate of the optimal pre-set lower-bound is $L_o = -U_o$.

After we obtained the estimates of the optimal pre-set boundaries U_o and L_o, the pairs trading strategy is applied to the in-sample data to obtain the actual number of trades, total profits and the averages of the trade durations. If the adjusted cointegration error, ϵ_t, is above or at U_o a U-trade is opened by selling one unit share $S1$ and buying β unit shares $S2$ and then close the trade by doing the opposite position when ϵ_t is below or at zero. We can also open an L-trade when ϵ_t is below $L_o = -U_o$ by buying 1 unit share $S1$ and selling β unit shares $S2$ and then it is closed by doing the opposite position when ϵ_t is above or at zero. In the case of BHP-RIO, BHP is assigned as $S1$ and RIO is $S2$ while in the case of TNS-TVL, TNS is $S1$ and TVL is $S2$. There is no opening trade when the previous trade has not been closed. We can compare the theoretical results and the actual results from the in sample data whether the share pair is worth enough to be traded.

Using the optimal pre-set boundaries U_o and L_o as well as the cointegration relationship from the in-sample data, we apply the pairs trading strategy to the out-sample data. We calculate the profit and trade duration from each trade (U-trades as well as L-trades) and at the end, the total number of trades, the total profits and the averages of the trade durations are also calculated. The results from the out-sample data show whether the pairs trading strategy still works or not.

5.1 BHP-RIO

BHP Billiton and Rio Tinto are major operators in the mining sector. Both have diversified mining resources in Australia, as well as other countries, that define them as blue-chip stocks in the ASX.

This paper uses the daily closing price of the two stocks from 2 January 2004 to 30 December 2004 as in-sample data and 3 January 2005 to 30 June 2005 as out-sample data. From the in-sample data, cointegration relationship of the two stocks is obtained as follows:

$$BHP_t - 0.61248RIO_t = \epsilon_t^*, \tag{29}$$

and the adjusted cointegration error

$$\epsilon_t = \epsilon_t^* - 7.3884, \tag{30}$$

and then fit ϵ_t as an AR(1) process as follows:

$$\epsilon_t = 0.8994\epsilon_{t-1} + a_t, \tag{31}$$

where $\sigma_\epsilon = 0.6479$.

Using the in-sample data i.e. $T = 251$ observations from 2 January 2004 to 30 December 2004 , and letting $\phi = 0.8994$ and $\sigma_a = \sqrt{1-\phi^2}\sigma_\epsilon = 0.2055$, we obtain the following estimates from the numerical algorithm in Section 4:

1. optimal pre-set upper-bound $U_o = 0.81$ and lower-bound L_o = -0.81,

2. total Number of trades (U-trades + L-trades) = 16.00,

3. minimum Total Profit (U-trades + L-trades)= $0.81 \times 16 =$ \$12.96,

4. expected trade duration = 13.74 days,

5. 1 trade (either a U-trade or an L-trade) per 15.70 days.

The above results are the estimation from the theory using the in-sample data. We also want to know the results when the pairs strategy explained in Section 2 is applied to the in-sample data and using U_o = 0.81 and L_o = -0.81. We obtain the following actual results:

1. number of trades (U-trades + L-trades)= 9,

2. total profit (U-trades + L-trades)= \$ 10.23 ,

3. average profit per trade = \$1.14,

4. average trade duration = 11.91 days,

5. on average, 1 trade (either a U-trade or an L-trade) per 27.9 days.

Comparing the estimation results from the theory and the actual results from the in-sample data show that the actual number of trades and the actual total profit are less than the estimation. However, we still get some profit and we always observe the profit per trade is higher than 0.81 dollars which is the optimal pre-set upper-bound U_o (the average profit per trade = \$1.14).

We also want investigate whether the pairs trading strategy using the out-sample data will also produce profit. Assume that the out sample data (T = 124 observations) still follows the models in Eqs.(29) and (31), so we can apply the same pair strategy and apply $U_o = 0.81$ and $L_o = -0.81$ as for the in-sample data. From the out-sample data, we obtain

1. number of trades (U-trades + L-trades): 4,

2. total profit (U-trades + L-trades): \$ 4.77 ,

3. average profit per trade = \$1.19,

4. average trade duration = 22.75 days,

5. on average, 1 trade (either a U-trade or an L-trade) per 31 days.

Comparing the trading results from the in-sample data and the out-sample data, the results are not too different (notice that the number of observations of the out-sample data is half of the in- sample data). The significant different from the both results is the average trade duration. However, from the out sample data we still always obtain profit per trade which is higher than 0.81 dollars (the average profit per trade = \$1.19).

5.2 TNS-TVL

Transonic Travel Ltd (TNS) and Travel.com.au (TVL) are travel companies listed on the ASX. In this study we consider the daily closing price of the two stocks from 2 January 2004 to 30 December 2004 as in-sample data and from 3 January 2005 to 30 June 2005 as out-sample data. From the in-sample data, we obtain the cointegration relationship of the two stocks to be:

$$TNS_t - 0.26659TVL_t = \epsilon_t^*, \tag{32}$$

with the adjusted cointegration error

$$\epsilon_t = \epsilon_t^* - 15.43, \tag{33}$$

and we fit ϵ_t as an AR(1) process as follows:

$$\epsilon_t = 0.9465\epsilon_{t-1} + a_t, \tag{34}$$

where $\sigma_\epsilon = 1.256258$.

With the indicated 251 days in-sample data, and $\phi = 0.9465$ and $\sigma_a = \sqrt{1-\phi^2}\sigma_\epsilon$, we obtain the following estimates from the numerical algorithm in Section 4:

1. optimal upper-bound U_o = 1.00 and lower-bound L_o = -1.00,
2. number of trades (U-trades + L-trades)= 10.91,
3. minimum Total Profit (U-trades + L-trades)= 1×10.91 = \$10.91,
4. expected trade duration = 18.00 days,
5. 1 trade (either a U-trade or an L-trade) per 18.42 days.

Analogous to the BHP-RIO case, we apply the pairs strategy explained in Section 2 for the in-sample data with U_o = 1.00 and L_o = -1.00, and we obtain the actual results:

1. number of trades (U-trades + L-trades)= 7,

2. total profit (U-trades + L-trades)= \$ 13.49 ,

3. average profit per trade = \$1.93,

4. average trade duration = 19.71 days.

5. on average, 1 trade (either a U-trade or an L-trade) per 35.85 days.

and from the out sample data (T = 124 observations), we obtain:

1. number of trades (U-trades + L-trades)= 2,

2. total profit (U-trades + L-trades)= \$ 3.68 ,

3. average profit per trade = \$1.84,

4. average trade duration = 32.5 days,

5. on average, 1 trade (either a U-trade or an L-trade) per 62 days.

From the TNS-TVL case, we see that the actual results for total profit and trade duration from the in-sample data are not too different with the estimation results from the theory, and even the actual total profit is significantly higher than the estimate. Furthermore, we always get a profit which is higher than 1 dollar from each trade (the average profit per trade = \$1.93 and \$1.84 from in-sample data and out-sample data respectively). However, the results from out-sample data are not quite good as we have only 2 trades and the the average trading duration is quite high (about 1 month). Perhaps, this result reflects that the out sample data does not quite follow the models in Eq.(32).

6 Conclusion

In this paper we have given a methodology to choose the optimal pre-set boundaries for pairs trading strategy based on cointegration technique and give a quantitative method to evaluate the average trade duration, the average inter-trade interval, and the average number of trades. The optimality in term of maximizing the minimum total profit over the specified trading horizon is developed by combining cointegration technique, the cointegration coefficient weighted rule, and the mean first-passage time using an integral equation approach.

The pairs trading strategy is applied to empirical data from two pair samples: BHP-RIO and TNS-TVL. Even though from the BHP-RIO case we can not obtain results as high as projected, we always obtain a profit per trade higher than the optimal pre-set upper-bound U_o. The actual total profit from both pair cases are quite similar to the estimates. For the TNS-TVL case, the results from the out sample data are not quite good. Perhaps, these results are due to the out sample data not quite following the models developed from the in sample data. Adjustment to the model may need to be made when using out sample data.

The above strategy can be extended if we set the minimum profit per trade as the minimum profit required (P_r), for example to meet the trading cost. We can trade $N_{S1} = \lfloor \frac{P_r}{U} \rfloor$ of S1 shares and $N_{S2} = \lfloor \beta \frac{P_r}{U} \rfloor$ of S2 shares to obtain the minimum profit per trading to be at least P_r. If we want to restrict the money invested in the trade to amount of I, we trade $N_{S1} = \lfloor \frac{I}{(P_{S1,t_o}+\beta P_{S2,t_o})} \rfloor$ of S1 and $N_{S2} = \lfloor \beta \frac{I}{(P_{S1,t_o}+\beta P_{S2,t_o})} \rfloor$ of S2 when we open a trade and then will get minimum profit per trade of UN_{S1}. [9]

We are aware that in large groups of stocks, the cointegrated stocks may not follow the assumptions given in this paper. For example, the cointegration relationship may disappear in the future, or the cointegration error may not be symmetric or may not an AR(1) process. Whether the technique displayed in this paper works or not only relies on two conditions:(1) within the in-sample data, there is a linear combination of stocks to form an AR(1) series, (2) such relationship does not significantly change in the trading period (out-sample data). Further investigations are warranted to explore different assumptions. In this paper we have established a framework that may be applied for a cointegrated stock pair with AR(1) cointegration error.

Acknowledgements

The authors would like to thank John Rayner, Martin Bunder, and Bernie Wilkes for their insightful and useful comments.

References:

Alexander, C. (1999). Optimal Hedging Using Cointegration. *Philosophical Transactions of the Royal Society*, A357,2039-2058.

Atkinson, K. (1997). *The Numerical Solution of Integral Equations of the Second Kind.* Cambridge University Press.

Avery-Wright, R. *Market Neutral (Pairs Trading) Explained.* http:// compareshares.com.au/expert_CFPs11.php.

Banerjee, A., Dolado, J., Galbraith, J. and Hendry, D. (1993). *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data.* New York, Oxford University Press.

Basak, G.K. and Ho, K.−W.R. (2004). Level-crossing Probabilities and First-passage Times for Linear Processes. *Advance Applied Probability*, **36**, 643-666.

[9] $\lfloor a \rfloor$ denotes the maximum integer number less than or equal a.

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day Inc.

Doornik, J.A. and Hendry, D.F. (1997). *Modeling Dynamic Systems Using Pc-Fiml 9.0 for Windows.* International Thomson Business Press.

Elliott, R.J., van der Hoek, J. and Malcolm W.P. (2005). Pairs Trading. *Quantitative Finance*, **5**, 271-276.

Engle, R.F. and Granger, C.W.J. (1987). Co-integration and Error Correction Representation, Estimation and Testing. *Econometrica*, **58**, 251-276.

Ehrman, D.S. (2006). *The Handbook of Pairs Trading: Strategies Using Equities, Options, and Futures.* John Wiley & Sons Inc.

Fama, E.F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Political Economy*, **96**, 246-273.

Gatev, E.G., Goetzmann, W.N. and Rouwenhorst, K.G. (1999). *Pairs Trading: Performance of a Relative Value Arbitrage Rule.* Working Paper 7-32, National Bureau of Economic Research.

Gatev, E.G., Goetzmann, W.N. and Rouwenhorst, K.G. (2006). Pairs Trading: Performance of a Relative Value Arbitrage Rule. *The Review of Financial Studies,* **19**, 797-827.

Gillespie, T. and Ulph, C. (2001). *Pair Trades Methodology: A Question of Mean Reversion.* Derivative Analysis Department, Salomon Smith Barney.

Goodboy, D. *A Basic Introduction to Pairs Trading.* http://biz.yahoo.com

Granger, C.W.J. (1986). Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economics and Statistics*, **48**, 213-228.

Grant, A. (2003). *Simulation and Application of a New Pairs Trading Strategy Involving Cointegration.* Bachelor of Mathematics and Finance Honours Thesis, University of Wollongong.

Habak, C. (2002). *Pairs Trading: Applying Cointegration to a Pairs Trading Strategy.* Bachelor of Mathematics and Finance Honours Thesis, University of Wollongong.

Harris, R.I.D. (1995). *Using Cointegration Analysis in Econometric Modeling.* London, Prentice Hall.

Hendry, D.F. and Doornik, J.A. (1996). *Empirical Econometric Modeling Using PcGive 9.0 for Windows.* International Thomson Business Press.

Hendry, D.F. and Juselius, K. (2000). Explaining Cointegration Analysis: Part II. *Energy Journal*, **22**, 75-120.

Herlemont, D. (2003). Pairs Trading, Convergence Trading, Cointegration. *YATS*

Finances and Technology. www.yats.com.

Hong, G. and Susmel, R. (2003). *Pairs-Trading in the Asian ADR Market*. http://www.bauer.uh.edu/rsusmel/Academic/ptadr.pdf.

Johansen, S. (1988). Statistical Analysis of Cointegrating Vectors. *Journal of Economic Dynamics and Control*, **12**, 231-254.

Lin, Y.−X., McCrae, M. and Gulati, C. (2003). *Necessary Conditions for a Minimum Profit Level in Pairs Trading: A Cointegration-based Procedure*. Workshop on Mathematical Finance, University of Wollongong.

Lin, Y.−X., McCrae, M. and Gulati, C. (2006). Loss Protection in Pairs Trading Through Minimum Profit Bounds: A Cointegration Approach. *Journal of Applied Mathematics and Decision Sciences*, (doi: 10.1155/JAMDS/2006/73803).

Liu, X., Song, H. and Romilly, P. (1997). Are Chinese stock markets efficient? A cointegration and casuality analysis. *Applied Economics Letters*, **4**, 511-515.

Narayan, P.K. (2005). Are the Australian and New Zealand stocks prices nonlinear with a unit root? *Applied Economics*, **37**, 2161-2166.

Stone, C. *Finding Profit in Pairs*. http://www.investopedia.com.

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

Whistler, M. (2004). *Trading Pairs - Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies*. John Wiley.