Research Summary

Mohammadreza Manzaritavakoli[1] and Hamidreza Arian[2]

[1] *The Graduate School of Management and Economics at Sharif University of Technology*
[2] *RiskLab Middle East,Tehran,Iran*

Summarised:11/23/2021

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

Heni Puspaningrum, YanXia Lin, Chandra M. Gulati

Date:

Received: April 27, 2009

Revised: March 17, 2010

The scope of the article:

The goal of this work is to provide a numerical algorithm for estimating the average trade duration, average inter-trade interval, and average number of trades, and then using that information to establish best pre-set boundaries that maximizes the minimal overall profit.

Abstract:

A pairs trade is a trading strategy that involves matching a long position with a short position in two stocks with a high correlation. In determining which stocks maybe a pair, Banerjee, Dolado, Galbraith, and Hendry (1993) and Vidyamurthy (2004) confirmed that the cointegration method is more powerful than the correlation criterion for extracting profit capacity in transient pricing anomalies among stock prices driven through common underlying factors.This research investigates how the pre-set boundaries used to open a trade might influence the minimum total profit over a particular trading horizon using stationary features of cointegration error after an AR(1) process.The higher the pre-set trading boundaries, the higher the profit per trade, but the lower the number of trades. Using the average trade duration and average inter-trade interval, the number of trades over a specific trading horizon can be approximated. Both of these numbers are approximated using the mean first-passage time as an analogy for any pre-set boundaries.

Literature:

Pairs trading works by taking advantage of temporary price anomalies between related stocks that are in long-run equilibrium. Once this happens, one stock will be overvalued in comparison to the other. We can then invest in a two-stock portfolio (a pair), selling the overvalued stock (short position) and buying the cheap stock (long position). After the stocks have settled back into their long-run correlation, the trade is closed out by taking the opposite position of these stocks. Profit is made from the short-term price differences between the two stocks. Pairs trading is a market-neutral investment method since the profit is not dependent on market movement.

In recent years, profit from the pairs trading method has been lower than it was when it was firstly used. Gillespie and Ulph (2001), Habak (2002), Hong and Susmel (2003), and Do and Faff (2008), on the other hand, demonstrated that considerable positive returns may still be achieved.

The classic pairs trading approach select two stock as paired stocks whose price are heavily correlated.These techniques are model-free, as Do, Faff, and Hamza (2006) point out, and as a result, they are not prone to model misspecification and misestimation.They do not, however, provide for the prediction of convergence time or expected holding period.Some academics are currently working on pairs trading using a stochastic method, as first proposed by Elliott, van der Hoek, and Malcolm (2005)3, who explicitly modeled the mean-reverting behavior of the spread (the difference between the two stock values) in a continuous-time context. The spread was assumed to follow an Ornstein-Uhlenbeck process, which is actually an AR(1) process in a continuous term, in these pairs trading based on stochastic techniques.They did not, however, explain how to select trade pairings so that the spread can follow an Ornstein-Uhlenbeck process. Additionally, the spread's mean-reverting feature cannot be proved.

The cointegration technique, as revealed by Banerjee, Dolado, Galbraith, and Hendry (1993), can be used to extract profit potential since the cointegration relationship ensures that the two stocks have a long-run stationary relationship (equilibrium). "Lin, McCrae and Gulati (2006) developed a pairs trading strategy that enables traders to obtain a preset profit. 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 umber of trades and the minimum total profit, and then using these to find optimal pre-set boundaries (thresholds) to open the pair trades"

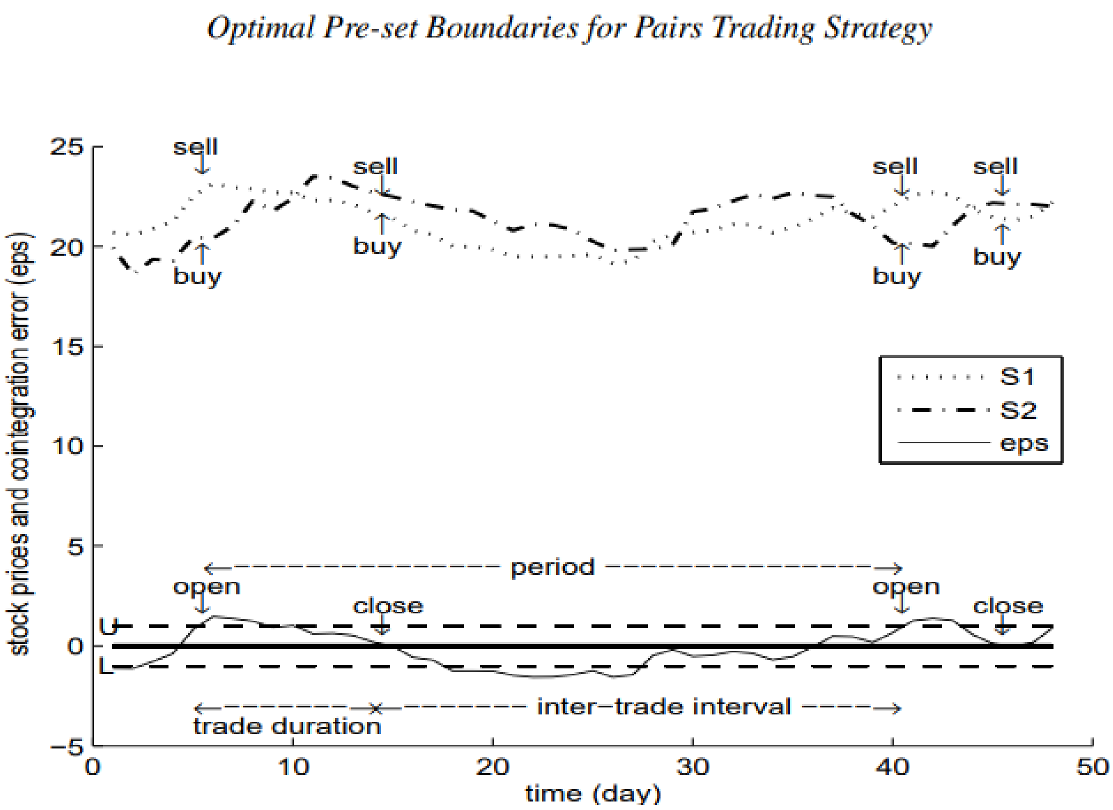

Figure 1: Example of two cointegrated shares

The research method and conclusion:

Consider two stocks, S1 and S2, whose prices are I(1), i.e., the stock price series are not stationary, but the first difference is. We say if these two stocks are cointegrated, then we have the below relationship between their prices:

$$P_{S1,t} - \beta P_{S2,t} = \mu + \varepsilon_t,$$

where ε_t is a stationary time series called cointegration errors and μ is the mean. A suitable linear combination of two I(1) non-stationary time series can sometimes lead in a stationary time series. If this occurs, the two I(1) series are said to be cointegrated. The boundaries, which we should set before starting a trade, act as thresholds to initialize a trade.There are two types of trades to consider: U-trades and L-trades. When the cointegration error is more than or equal to the pre-set upper-bound U, a U-trade is opened by selling S_2 of shares in S_1 and buying N_{S1} of shares in S_2, with the trade closing when the cointegration error is less than or equal to zero.The L-trade is the opposite. The lengths from the upper-bound U to the mean and from the lower-bound L to the mean are assumed to be the same since the cointegration error is considered to be a stationary process with a symmetric distribution. As a result, the number of U-trades and L-trades is predicted to remain the same. See Lin, McCrae, and Gulati for more information (2006).To keep things simple, we'll start with the U-trade scenario and work our way up to the optimum upper-bound.

"The following terms are required in further discussion. 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)." Maximizing the minimum total profit (MTP) over a defined trading period determines the optimality of the pre-set boundary values. The MTP is the pre-determined minimum profit per trade as well as the number of trades during the trading horizon. Lin, McCrae, and Gulati (2006) present the derivation of the pre-set minimum profit per trade, hence this study will focus on the expected number of trades. The distance between the pre-set boundaries and the long-run cointegration equilibrium also influences the number of trades. The higher the pre-set opening trade boundaries, the higher the minimum profit per trade, but the trade volume will be reduced. When it comes to reducing the boundary values, the opposite is true. The average trade duration and the average inter-trade interval are used to estimate the number of trades over a given trading horizon. Both of those values are approximated using an analogy to the mean first-passage timings for an AR(1) process for any pre-set limits. The mean first-passage times from Basak and Ho are evaluated using an integral equation approach in this research (2004).

"When the cointegration error is higher than or equal to the pre-set upper-bound U at time to, a trade is opened by selling N_{S1} of S_1 shares at time to for $N_{S1}S_1$,t0 dollars and buying N_{S2} of S_2 at time to for $N_{S2}S_2$, to dollars. When the cointegration error has settled back to its equilibrium at time t_c, the positions are closed out by simultaneously selling the long position shares for $N_{S2}S_2$,t_c dollars and buying back the N_{S1} of S_1 shares for $N_{S1}S_1$,t_c dollars. Profit per trade will be

$$P = N_{S2}\left(P_{S2,t_c} - P_{S2,t_o}\right) + N_{S1}\left(P_{S1,t_o} - P_{S1,t_c}\right)$$

According to the CCW rule as in Lin, McCrae and Gulati (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 $N_{S2} =$, the minimum profit per trade can be determined as follows:

$$P = \beta\left[P_{S2,t_o} - P_{S2,t_c}\right] + \left[P_{S1,t_c} - P_{S1,t_o}\right] = \left(\varepsilon_{t_c} - \varepsilon_{t_o}\right) \geq U$$

Thus, by trading the shares with the weight as a proportion of the cointegration coefficients, the profit per trade is at least U dollars." Accordingly, trading 1 unit share S1 and 1 unit share S2 in U-trades or L-trades would result in a minimum profit of at least U per trade.

In this research, we provide an alternative to Basak and Ho's cointegration-based pairs trading technique by applying an integral equation approach of the mean first passage time (2004). This section will lead you through how to estimate the number of trades over a certain trading horizon. To begin, we will review the mean first-passage time of the AR(1) process using an integral equation approach developed by Basak and Ho (2004). Second, using an integral equation method, a numerical scheme is offered to calculate the mean first-passage time of an AR(1) process. Finally, an analogy of the mean first-passage time is used to estimate the average trade duration and average inter-trades interval. Fourth, the average trade duration and average inter-trade interval are used to approximate the number of trades over a certain trading horizon.

Consider an AR(1) process:

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

where $-1 < \phi < 1$ and $\xi_t \sim$ i.i.d. $N\left(0, \sigma_\xi^2\right)$.
The first-passage time $T_{a,b}\left(y_0\right)$ is defined as

$$T_{a,b}\left(y_0\right) = \min\left\{t : Y_t > b \text{ or } Y_t < a \mid a \leq Y_0 = y_0 \leq b\right\}$$

Particularly,

$$T_a\left(y_0\right) = T_{a,\infty}\left(y_0\right) = \min\left\{t : Y_t < a \mid Y_0 = y_0 \geq a\right\}$$

and

$$T_b\left(y_0\right) = T_{-\infty,b}\left(y_0\right) = \min\left\{t : Y_t > b \mid b \geq Y_0 = y_0\right\}$$

We define a discrete-time real-valued Markov process Y_t on a probability space (W, F, 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 $Y_0 = y_0[a, b]$. The mean first-passage time over interval [a,b] of an AR(1) process, starting at initial state $y_0[a, b]$, is given by"

$$E\left(T_{a,b}\left(y_0\right)\right) = \int_a^b E\left(T_{a,b}(u)\right) f\left(u \mid y_0\right) du + 1$$

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

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

This is a Fredholm type of the second kind integral equation that may be numerically solved using the Nystrom method (Atkinson, 1997) as shown in the next subsection.

"Using the trapezoid integration rule (Atkinson, 1997):

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

where $u_0=a, u_j=a+jh, u_n=b, j=1,\ldots,n$ and the weights w_j for the corresponding nodes are

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

Thus, the integral term in (3.6) can be approximated by

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

Let $E_n\left(T_{a,b}\left(y_0\right)\right)$ denote the approximation of $E\left(F_{a,b}\left(y_0\right)\right)$ using n partitions. Thus, the expectation in using n partitions can be estimated by

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

Set y_0 as u_i for $i=0,1,\ldots,n$ and reformulate as follows

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

and then solve the following linear equations in to obtain an approximation of $E_n\left(P_{a,b}\left(u_j\right)\right)$."

$$\begin{pmatrix}1-K\left(u_0,u_0\right) & -K\left(u_0,u_1\right) & \ldots & -K\left(u_0,u_n\right)\\ -K\left(u_1,u_0\right) & 1-K\left(u_1,u_1\right) & \ldots & -K\left(u_1,u_n\right)\\ \vdots & \vdots & \vdots & \vdots\\ -K\left(u_n,u_0\right) & -K\left(u_n,u_1\right) & \ldots & 1-K\left(u_n,u_n\right)\end{pmatrix}\begin{pmatrix}E_n\left(T_{a,b}\left(u_0\right)\right)\\ E_n\left(T_{a,b}\left(u_1\right)\right)\\ \vdots\\ E_n\left(T_{a,b}\left(u_n\right)\right)\end{pmatrix}=\begin{pmatrix}1\\1\\ \vdots\\1\end{pmatrix}$$

where

$$K\left(u_i,u_j\right)=\frac{h}{2\sqrt{2\pi}\sigma_{\xi}} w_j \exp\left(-\frac{\left(u_j-\phi u_i\right)^2}{2\sigma_{\xi}^2}\right) \quad \text{where } i,j=1,\ldots,n$$

To make the length of partition h the same in each case, the paper employs distinct b and n. The results suggest that $h=0.1$ is acceptable to provide simulation-like results.

We need to know how long it takes on average for ε_t to pass 0 for the first time in order to compute the expected trade duration. Computing the expected trade duration is thus equivalent to calculating the mean first-passage time for ε_t to pass 0 for the first time, assuming the initial value is U. Let TD_U stand for the estimated trade duration that corresponds to the upper-bound U:

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

Same for calculating expected inter-trade interva, we can calculate the mean-first passage time for ε_t to pass U, assuming the initial value is 0. Let I_U stand for the estimated inter-trade interval that corresponds to the upper-bound U:

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

For U-trades, $Period_U$ is defined as the sum of the trade duration and the inter-trade interval, as mentioned before. As a result, the expected $Period_U$ is:

$$E(Period_U) = I_U + TD_U$$

First, we'll examine the expected number of $Period_U$'s ($E(N_{UP})$) in the time horizon $[0,T]$, which is linked to trade duration and inter-trade interval. Then, to generate a potential range of $E(N_{UT})$ values, the relationship between the expected number of U-trades and the expected number of periods corresponding to U-trades $E(N_{UP})$ will be employed, the result is:

$$\frac{T}{TD_U+I_U} + 1 \geq E\left(N_{UP}\right) + 1 \geq E\left(N_{UT}\right) \geq E\left(N_{UP}\right) > \frac{T}{TD_U+I_U} - 1$$

The optimal pre-set upper-bound, represented by U_0, is determined by maximizing the minimum total profit (MTP).For a pre-determined upper-bound U, let TP_U represent the total profit from U-trades across the time range $[0,T]$. Thus,

$$TP_U = \sum_{i=1}^{N_{UT}} (\text{ Profit from the } i \text{ th U-trade })$$

$$\text{Profit per trade} \geq U \quad \text{and} \quad E\left(N_{UT}\right) \geq \frac{T}{TD_U+I_U} - 1$$

$$\mathrm{MTP}(U) := \left(\frac{T}{TD_U + I_U} - 1\right) U$$

"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 $5\sigma_\varepsilon$ because ε_t is a stationary process, and the probability that $|\varepsilon_t|$ is greater than $5\sigma_\varepsilon$ is close to zero. [15] The numerical algorithm to calculate the optimal pre-set upper-bound U is as follows: 1. Set up the value of b as $5\sigma_\varepsilon$. 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 $E\left(T_{0,b}\left(U_i\right)\right)$ as the trade duration (TD_{U_i}) using (3.13). (b) calculate $E\left(T_{-b,U_i}(0)\right)$ as the inter-trade interval (I_{U_i}) using (3.14). (c) calculate $MTP\left(U_i\right) = \left(\frac{T}{TD_{U_i}+I_{U_i}} - 1\right) U_i$ 4. Find $U_o \in \{U_i\}$ such that MTP (U_o) is the maximum."

Heni Puspaningrum; **Yan-Xia Lin & Chandra M. Gulati**