

配對交易

一、前言

配對交易是指同時買入和賣出兩個相似的商品，這些商品在價格走勢或基本上表現類似，透過買賣兩者的價差，消除大部分市場風險，從而獲取穩定的報酬率。尋找配對交易標的最簡單方法是計算兩個商品之間的相關係數。假設過去兩個金融商品的走勢相似，則預期未來也會保持相同的關係，因此可以買入被低估的商品並賣出被高估的商品。然而，這種方法缺乏嚴謹性，高相關係數並不意味著兩者的價差在出現時會逐漸收斂，它僅能反映短期內的價格聯動。例如，如果兩隻股票的價格隨時間一起上漲，則它們是正相關的，但若這兩隻股票在上漲的同時，價差繼續擴大而沒有收斂，則投資者將無法從中獲利。

二、摘要

配對交易的步驟可分為兩個主要階段：首先，選擇配對交易的標的；其次，找尋合適的買賣時機以獲取報酬。

在選擇**配對交易標的**的方法方面，市面上的論文提出了多種方法，包括：

- 距離法：計算兩商品的價差標準差。
- 共整合法：使用 Engle 和 Granger (1987) 提出的二階段共整合檢定法。
- 隨機過程法：計算配對價差的長期均數復歸速度 (Kappa)。
- 機器學習：將股票分群以尋找配對標的。
- 主成份分析：透過將股票報酬率降維找尋特徵。

在確定**買賣時機**方面，可以使用以下方法：

- 技術分析：對兩商品線性回歸的殘差技術分析
- Copula 函數：計算條件機率，當條件機率異常大或小時進行交易。

此報告旨在探討在不同方法下，哪種配對交易的績效表現最好，因此對上述方法進行實驗，並且由於投資者採用配對交易的目的是賺取穩定報酬，但配對交易並非真正的套利，當兩商品的走勢不如預期時，虧損是可能的，因此，此報告的目的是找到期望值最高，且在包含交易成本的情況下勝率仍高的策略。

三、研究方法

歷史上能獲取數據的上市和上櫃股票總共為 2499 支，其中包含已下市或下櫃的股票，若要對每兩組股票進行配對，則需進行 $2499 * 2499 = 6,245,001$ 次計算。在配對交易中，若利用二階段共整合檢定法選取配對組，計算殘差時需耗費大量算力，儘管使用平行運算，仍需數十小時才能完成所有計算。為了以最短的時間驗證這些方法能否找到績效不錯的配對交易，實驗開始前，我們隨機選取了 100 支股票，並在後續實驗中使用這 100 支股票進行測試，為了不要出現過度擬合、前瞻風險的狀況，在此報告中一率使用 2007 至 2020 年的資料找尋配對標的，然後於 2021 至 2024 年做回測。

1. 方法一（尋找標的：共整合檢定法；買賣時機：布林通道）

重現「配對交易與機器學習在台灣股票市場之應用」論文的方法，其先使用 Engle and Granger（1987）提出的二階段共整合檢定法，先判斷兩商品的價格是否呈現非定態，然後再利用最小平方方法預估兩者的回歸係數，使兩者的關係能用簡單的線性解釋，再透過檢驗線性回歸的殘差是否為定態序列來確定共整合關係，這種方式能夠驗證配對的金融商品間是否具有長期均衡關係，確保未來有均值回歸的可能性。至於買賣時機的部分，此論文利用布林通道在兩商品的殘差之上操作。

透過二階段共整合檢定，在隨機選 100 個股票中，總共可以發現 375 組配對標的。

$$Pstock_1 = a * Pstock_2 + b$$
 a是回歸係數，b是常數。

當操作1股stock₁時，需要操作round((b + stock₂ * a)/stock₂)股stock₂。

兩者的殘差為Pstock₁ - a * Pstock₂ - b。

若當天收盤的殘差大於布林通道的上軌，則賣出 stock1，買入 stock2，直到殘差收斂回均線加一倍標準差的值；若當天收盤的殘差小於布林通道的下軌，則買入 stock1，賣出 stock2，直到殘差收斂回均線減一倍標準差的值。

(1) 績效

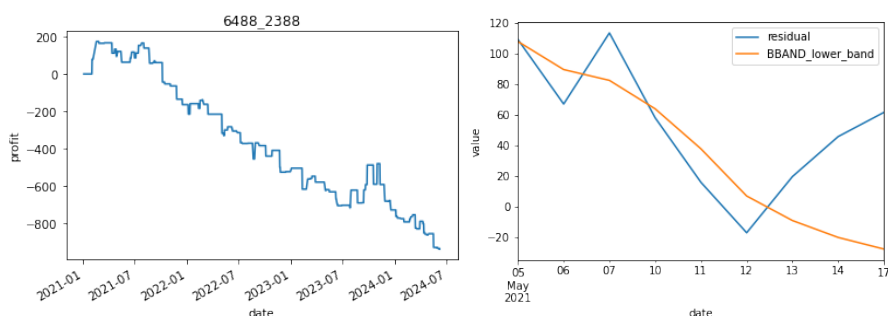
方法一	
交易時間	2021/01-2024/06
配對組數	375
最大虧損（操作 1 股 stock1）	-1943
最佳盈利（操作 1 股 stock1）	5192
勝率	0.56
期望值	90.48

(2) 含手續費的績效

方法一	
交易時間	2021/01-2024/06
配對組數	375
最大虧損（操作 1 股 stock1）	-2041
最佳盈利（操作 1 股 stock1）	5155
勝率	0.19
期望值	3.29

(3) 缺點

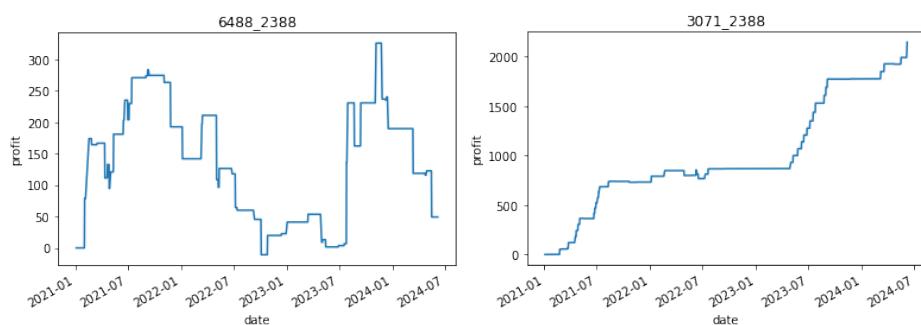
下左圖為透過此操作而虧損的配對交易（6488.tw, 2388.tw），原因是當殘值有趨勢時，布林通道也會跟著調整，因此當殘值不斷減少，在殘值剛碰到下軌道便買入 stock1、賣出 stock2，可能會導致 stock1 買在高點，然後 stock2 賣在低點，搭配下右圖可以很好地發現，在 2021/05/05 我們期望殘值會向上彈回均線減掉一倍標準差的值，然而殘值卻不斷往下掉，造成虧損。



2. 方法二（尋找標的：共整合檢定法；買賣時機：布林通道+殘值正負）

在觀察方法一的例子後，發現若殘值在布林通道下軌道之下且殘值為負，則殘值向上回歸的機率較高；反之，若殘值在布林通道上軌道之上且殘值為正，則殘值向下回歸的機率較高。因此，方法二除了增加殘值正負的限制外，其他操作步驟都和方法一相同。

下左圖為在方法一虧損的例子（6488.tw, 2388.tw），在套用方法二下由虧轉盈，而下右圖為在方法一盈利的例子（3071.tw, 2388.tw），儘管增加了條件，仍能獲得不錯的報酬。



(1) 績效

方法二	
交易時間	2021/01-2024/06
配對組數	375
最大虧損（操作 1 股 stock1）	-672
最佳盈利（操作 1 股 stock1）	8437
勝率	0.69
期望值	196.96

（2）含手續費的績效

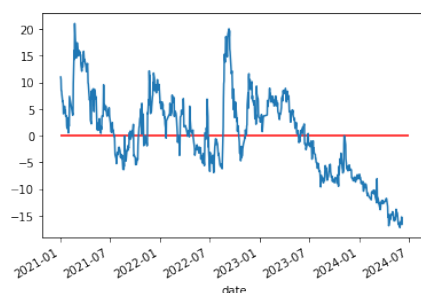
方法二	
交易時間	2021/01-2024/06
配對組數	375
最大虧損（操作 1 股 stock1）	-716
最佳盈利（操作 1 股 stock1）	8309
勝率	0.33
期望值	139.78

（3）缺點

雖然方法二的勝率有所提升，但方法一中提到的布林通道會隨著趨勢調整的缺點仍未解決。

3. 方法三（尋找標的：共整合檢定法；買賣時機：均線+殘值正負）

雖然在配對交易中假設殘值終將均值回歸，但有時殘值如下圖不會馬上回歸，若我們使用均值回歸的技術指標來操作，會因為殘值的不斷擴大而造成虧損。



因此，在方法三中實驗利用殘值的長短均線來產生交易訊號，是否能提升勝率。方法三的作法為，透過殘值的 2 天及 5 天均線進行操作，當殘值為負且快線（2 天均線）大於慢線（5 天均線）時，視為殘值將繼續向上擴大，因此買入 stock1、賣出 stock2，直到快線與慢線在殘值為正時產生死亡交叉，或者殘值變為負數時平倉，而當殘值為正且快線（2 天均線）小於慢線（5 天均線）時，視為殘值將繼續向下擴大，因此買入 stock2、賣出 stock1，直到快線與慢線在殘值為負時產生黃金交叉，或者殘值變為正數時平倉。

（1）績效

方法三	
交易時間	2021/01-2024/06
配對組數	375
最大虧損（操作 1 股 stock1）	-24218
最佳盈利（操作 1 股 stock1）	47590

勝率	0.68
期望值	1914.43

(2) 含手續費的績效

方法三	
交易時間	2021/01-2024/06
配對組數	375
最大虧損（操作 1 股 stock1）	-24303
最佳盈利（操作 1 股 stock1）	46734
勝率	0.43
期望值	1822.39

(3) 缺點

此方法的勝率跟期望值雖然表現不錯，但是最大虧損在提及的三個方法中是最大的，這給予了在做配對交易時，可以增加停損機制的啟發。

4. 方法四（尋找標的：PCA+OPTICS；買賣時機：布林通道+殘值正負）

重現 Medium 的文章「PAIRS TRADING: USING MACHINE LEARNING FOR THE SELECTION OF PAIRS」，由於此文章實驗在美股，因此好奇若實驗在台股績效如何。首先利用主成分分析（PCA）將實驗開始時隨機選取的 100 支股票報酬降至兩維度，然後使用 OPTICS 演算法將股票分群，接著在同一群中的股票進行兩兩配對。



由於布林通道加上殘值正負的方法（方法二）操作勝率最高，因此在使用 PCA 和 OPTICS 演算法進行分群和配對後，我們利用布林通道和殘值正負的方法來操作股票。

(1) 績效

方法四	
交易時間	2021/01-2024/06
配對組數	177
最大虧損（操作 1 股 stock1）	-447

最佳盈利（操作 1 股 stock1）	1782
勝率	0.75
期望值	89.17

由上表可見，選擇到的配對比共整合檢定方法少，但勝率有所提高。由於在現實中資金有限，因此方法四相較以上的方法，在實際執行中的可行性更高。

（2）含手續費的績效

方法四	
交易時間	2021/01-2024/06
配對組數	177
最大虧損（操作 1 股 stock1）	-494
最佳盈利（操作 1 股 stock1）	1770
勝率	0.33
期望值	56.19

5. 方法五（尋找標的：SSD+Kappa+STD；買賣時機：布林通道+殘值正負）

重現論文「建立配對交易之綜合指標」中的方法，該論文在選擇配對標的時，採用了配對股票的股價標準化後價差之平方總和（SSD）、均數復歸模型計算配對價差長期均值回歸的速度（Kappa）、配對標的之價差標準差（STD）三個指標，綜合指數即為這三個數值相加，並選擇綜合指數最大的前 50%股票進行配對交易。

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t$$

其中，

- dB_t ：標準布朗運動，定義機率空間
- θ ：配對價差長期平均水準
- κ ：配對價差之回復速度

均數復歸模型：，K 為 Kappa。

（1）績效

方法五	
交易時間	2021/01-2024/06
配對組數	86
最大虧損（操作 1 股 stock1）	-1241
最佳盈利（操作 1 股 stock1）	669
勝率	0.73
期望值	36.39

（2）含手續費的績效

方法五	
交易時間	2021/01-2024/06

配對組數	86
最大虧損（操作 1 股 stock1）	-1268
最佳盈利（操作 1 股 stock1）	664
勝率	0.32
期望值	-7.61

（3）缺點

由上表可知，此方法的勝率與方法四不相上下，而兩者的差別只差在選股的方式不同，但值得注意的是，此方法的最大虧損的絕對值比最佳盈利還要大，儘管勝率不錯，此方法的期望值相比其他方法低許多，顯示出此策略在大多數情況下可以賺錢，但利潤有限，而虧損的機率雖然不高，但一旦發生，虧損將會很大，而且目前尚未考慮交易成本的問題，若加上交易成本，此策略可能無法盈利，因此，若要在現實中執行策略，我更偏好方法四。此外，該論文希望找到 SSD 越小，Kappa 越大的配對，卻將三個指標加在一起，因此我認為這種選股方式並不恰當，且複雜化選股邏輯，導致無法找到理想的配對。

6. 方法六（尋找標的：Tau+Rho；買賣時機：Copula）

重現論文「採用關聯式結構模型執行配對交易在台灣股票市場」，利用關聯式結構模型（Copula）來了解隨機變數之間的關係。在前面提到的方法中，二階段共整合檢定假設股價報酬率呈現常態分配，然而現實中股價報酬率通常不呈現常態分配，且配對標的通常不屬於線性關係。因此，關聯式結構模型是一種用於描述多個隨機變量之間相依結構的函數，能夠幫助分析不呈線性關係的兩個隨機變數之間的關係，並計算其聯合機率。

在此論文中，先透過 Kendall's Tau 和 Spearman's Rho 係數來選擇配對標的，與 Pearson 相關係數不同的是，這兩個係數可以用來分析非線性關係，當兩個係數皆大於 0.6 時，即選取這些配對股票作為交易標的。由於此方法在篩選配對標的時條件較嚴苛，因此從實驗開始時隨機選取的 100 支股票中進行配對篩選，最終獲得了 50 組配對標的。接著，論文中利用 Copula 函數來計算條件機率：

$$P(Y \leq y | X = x) = \frac{\partial C(u, v)}{\partial u} \bigg|_{u=F(x)}$$

左圖的 X 為 stock2 的股價，Y 為 stock1 的股價，當條件機率超過 0.95 時，賣出 stock1，買進 stock2，直到條件機率收斂至 0.55（論文中的參數）時平倉；若條件機率低於 0.05 時，買進 stock1，賣出 stock2，直到條件機率收斂至 0.45（論文中的參數）時平倉。

（1）績效

方法六	
交易時間	2021/01-2024/06

配對組數	50
最大虧損（操作 1 股 stock1）	-84.5
最佳盈利（操作 1 股 stock1）	84.5
勝率	0.5
期望值	-3.55e-15

（2）含手續費的績效

方法六	
交易時間	2021/01-2024/06
配對組數	50
最大虧損（操作 1 股 stock1）	-109
最佳盈利（操作 1 股 stock1）	74
勝率	0.46
期望值	-6

（3）缺點

倘若將此方法選到的 50 支股票運用布林通道加上殘值正負值的方法操作：

方法六（改變買賣操作，不含手續費）	
交易時間	2021/01-2024/06
配對組數	50
最大虧損（操作 1 股 stock1）	-417
最佳盈利（操作 1 股 stock1）	1510
勝率	0.74
期望值	69.02

方法六（改變買賣操作，含手續費）	
交易時間	2021/01-2024/06
配對組數	50
最大虧損（操作 1 股 stock1）	-455
最佳盈利（操作 1 股 stock1）	1486
勝率	0.2
期望值	34.51

因此判斷利用 Copula 條件機率來決定買賣時機點並不是好方法，或許更換條件機率的參數可以得到更好的結果。

未來可優化之處：

- 許多論文提及運用機器學習來找尋適當的買賣點，例如「透過機器學習及標記技術建構配對交易策略」。然而，這些論文提及的主要是監督式機器

學習。在本報告中，尚未發現勝率接近 9 成的操作方式，因此難以標記買賣點給機器學習。在未來的優化中，除了繼續研究買賣時機並交給機器學習外，或許也可以嘗試非監督式學習。

- 上述實驗參雜許多參數，在未來沒有時間壓力之下，可以試著優化參數，看是否能提升績效。
- 由於配對的商品價格可能不如預期收斂，因此可以加入停損的機制。
- 嘗試其他金融商品，目前只以隨機選取的 100 支股票做實驗。

結論：

在不包含手續費的狀況下，方法三（尋找標的：共整合檢定法；買賣時機：均線+殘值正負）雖然勝率最低，但其期望值最高，而勝率最高的則是方法四（尋找標的：PCA+OPTICS；買賣時機：布林通道+殘值正負）。觀察所有策略在不包含手續費的狀況下，皆可得正的期望值，顯示配對交易的存在，但由於許多時候兩商品出現的價差不足以支付手續費，因此在包含手續費的狀況下，所有策略的績效都變差。

若要在現實中執行配對交易，則考慮交易成本是必要的，在納入交易成本的狀況下，方法三（尋找標的：共整合檢定法；買賣時機：均線+殘值正負）的期望值最高，但最大虧損的值也較大，而方法二（尋找標的：共整合檢定法；買賣時機：布林通道+殘值正負）的勝率比方法三高一些，為較穩健一些的交易策略。

然而以上的勝率跟期望值是在交易所有配對標的下的結果，在現實中如果資金有限，則方法六（尋找標的：**Tau+Rho**）在改變買賣操作至利用布林通道加上殘值正負值的狀況下，只選擇了 50 組配對標的，在含手續費之下期望值達 34.51，也是不錯的策略。