國立政治大學金融學系研究所

碩士學位論文

# 統計套利下動態共整合關係之跨商品應用

# Application of Cross-Commodity Statistical Arbitrage Base on Dynamic Cointegration

指導教授：林士貴 博士

王釧茹 博士

研究生：徐語辰 撰

中華民國 一百零八 年 六 月



DOI:10.6814/NCCU201900124

# 謝辭

對於這篇論文，首先我要感謝我的指導老師林士貴教授與王釧茹教授，老師花了非常多的時間與我討論，對於我的研究提供許多建議與研究方向，並一次又一次幫我挑出論文不足之處，真的非常感謝老師的指導和付出。再來我要感謝Coinful Capital 執行長 Hank，在我研究期間不斷的根據我的研究給我建議與方向，讓我對統計套利有更深的一層領悟，我的研究也因此更加豐富。接下來我要感謝希奇資本(Xi Qi Capital)的陳琪龍教授，在碩一時多次與我討論並教導我統計套利相關概念。

接著我要感謝我的同門李昱穎和李杰穎，除了在論文方面我們也互相討論及督促彼此的進度，也一同討論學習程式交易這個新的領域，當然也要感謝班上其他同學，像許晏寧幫我解決了很多關於程式上的技術問題，還有很多同學，這裡就不一一列舉，你們讓我在碩士班兩年的生活更加的豐富及充實。

最後，我要感謝我的父母與爺爺奶奶，謝謝你們在生活上給予我充足的資源，讓我無後顧之憂地完成我的學業，在我面對困難時也適時地給我意見及鼓勵，並無條件支持我的任何決定，讓我從電資學院轉換跑道學習金融，沒有你們就沒有今日的我，未來踏入職場我會更加努力，不會讓你們失望的！

*我們用數據與模型當作翅膀，嘗試逃離市場的牢籠，但最終還是回到了經濟與政治的舞台。*

徐語辰　謹至于

國立政治大學　金融研究所

民國一〇八年六月



DOI:10.6814/NCCU201900124

# 統計套利下動態共整合關係之跨商品應用

學生:徐語辰　　　　　　　　　　　　　　　　　　　　指導教授：林士貴 博士

國立政治大學金融系

## 摘要

本研究根據 Vidyamurthy (2004)以及後續相關文獻所提出的統計套利配對交易方法對台灣股票市場進行實證研究。本文使用的模型為 Engle and Granger (1987)提出的二階段共整合檢定。我們利用上述模型檢定芝加哥交易所期貨一分鐘資料，找出具共整合性質之配對，利用技術指標─布林通道與 OU 過程找出價格異常的時間點進行交易，建構統計套利投資組合；本研究進一步將中位數反轉定律 Andrew(2009)加入，用於預測共整合殘差走勢，建構中位數反轉定律結合布林通道與 OU 過程之統計套利策略並建構投資組合。實證結果顯示和 Avellaneda and Lee (2010)結果相同，市場上確實存在市場中立性的報酬，且兩個策略的投資組合皆有優於大盤的績效和穩健性；此外中位數反轉定律確實有幫助我們減少進場次數提高勝率，並且使投資組合的最大虧損下降。

***關鍵詞：共整合、統計套利、布林通道、OU 過程、中位數反轉定律***



DOI:10.6814/NCCU201900124

# Application of Cross-Commodity Statistical Arbitrage Base on Dynamic Cointegration

Student: Yu-Chen, Hsu Advisor : Dr. Shih-Kuei, Lin

Department of Money and Banking, National Chengchi University

## Abstract

This paper used the statistic arbitrage method according to Vidyamurthy (2004) and other papers based on this book. This paper followed papers to conduct empirical research on Chicago Mercantile Exchange market. The models used in this paper is two-steps cointegration test that proposed by Engle and Granger (1987). We tested CME futures through the above models to test cointegration, and find the investable pairs. After finding out investable pairs, we used Bollinger Band and OU process to find out abnormal stock price to trade. Then we constructed the portfolio to study its performance. This study further adds the median reversal law by Andrew(2009) to predict cointegral residual and constructs a strategy with Bollinger Band and OU process model. The result shows that the strategy helping us find market neutral return, which is the same as the result of Avellaneda and Lee (2010). Furthermore, our portfolio is also better than investing in benchmark. Median reversal law truly helps us reduce trading frequency and decrease drawdown.

***Keywords：Cointegration 、Statistic Arbitrage 、Bollinger Band 、OU Process 、Median Reversal Law***



**DOI:10.6814/NCCU201900124**

# 目錄





DOI:10.6814/NCCU201900124

DOI:10.6814/NCCU201900124

# 圖目錄





DOI:10.6814/NCCU201900124

DOI:10.6814/NCCU201900124

# 表目錄







DOI:10.6814/NCCU201900124

# 第一章　緒論

## 第一節　　研究動機

近年來全球金融市場開放，加上 2008 年金融海嘯以來世界各國的貨幣寬鬆政策，又由於國際金融局勢日趨複雜，政治與經濟情勢變動對國際市場產生巨大影響的可能性漸增，諸如英國脫歐、川普當選美國總統、中美貿易戰等……，皆造成全球金融市場價格大幅波動，傳統利用總體經濟資料、歷史交易資訊來判斷長期持有投資策略，其績效穩健性開始受市場風險挑戰。

為規避此類系統性風險(System Risk)，最常使用的方法便是市場中立性投資策略(Market Neutral Strategy)，其概念是利用多個關聯性很高的資產，採用其價格間的連動關係，同時建立多空部位投資組合來消除大部份市場風險，降低投資組合績效與市場走勢相關性獲得穩定報酬，而在多種市場中立性投資策略中，配對交易(Pairs Trading)便是其中一種常用且建構困難度較低的方式。

配對交易(Pairs Trading)是一種利用二、三個資產配置，達到市場中性概念的投資策略，最早於 1980 年代中期華爾街著名投行 Morgan Stanley 的數量交易員 Nunzio Tartaglia 成立的數量分析團隊提出並使用，當時獲得優異績效，至今仍為許多投資機構投資人所使用，亦為許多研究的研究對象。而基於配對交易概念後發展的統計套利(Statistical Arbitrage)研究又如雨後春筍般的出現，故本研究期望能在這複雜的國際局勢中，藉由統計套利策略在芝加哥期貨交易所(CME)的五種期貨中的可行性分析及獲利能力分析，探討特定市場是否存在建構統計套利這種市場中立性交易策略的機會，期望能從中找到能有效分散市場風險並具有穩定獲利能力的投資策略，供從事相關交易及研究工作的工作參與者參考。



DOI:10.6814/NCCU201900124

## 第二節　　研究目的

Vidyamurthy (2004)定義統計套利是依據 Engle and Granger (1987)二階段共整合檢定法，檢定高度相關的兩資產間是否存在長期均衡關係，若確定其有長期的均衡關係，一旦兩者之間出現了異常的走勢，就可能產生套利的機會。當我們利用技術分析或量化方法判定其走勢異常，便會建構一個統計套利策略，買進價格相對被低估的資產，賣出價格相對被高估的資產。當未來兩者之間的價差得到修正，便可以進行相反的平倉操作來獲取利潤。接著我們進一步使用統計與計量知識去嘗試估計價差短期內穩定的均值與波動度，並預測它迴歸的機率與速度。統計套利被廣泛的應用，主要原因是：首先，統計套利的收益與市場互相獨立，即市場中立性，也就是說它與市場的上漲或者下跌無關；其次，其收益波動性相對於一般交易較小；第三，其收益相對於一般交易穩定。

本研究使用統計套利策略概念，應用芝加哥商品交易所(CME)的五個資產上，第一步對資產價格平滑化與消除雜訊處理，第二步找出具有共整合關係的配對組合時段，在使用共整合動態校準的方式，對兩兩資產進行統計套利模型建構。

當模型建構完成後，第三步我們嘗試加入適當的假設，分別使用布林通道或 OU 過程判斷價差異常時機，建構統計套利交易策略，倘若共整合關係崩解，則馬上解除模型並且重新尋找套利機會。

最後，對該序列做出預測，提高模型準確度。希望透過本研究得到的實驗結果可以讓避險基金做為參考，提供新的市場中立性交易策略，也希望透過這樣的方式建構出可以有效獲利的交易策略。





DOI:10.6814/NCCU201900124

## 第三節　研究架構

本研究論文架構共分為五部分，各章內容分述如下：

### 第一章、　緒論

說明本研究之研究動機與目的，並說明研究之流程用以清楚展示研究之主題以及流程。

### 第二章、　文獻探討

說明共整合檢定及 OU 過程是如何應用於交易策略、在來回顧如何處理具有雜訊的價格資料，最後用機率證實預測可行性的文獻回顧。

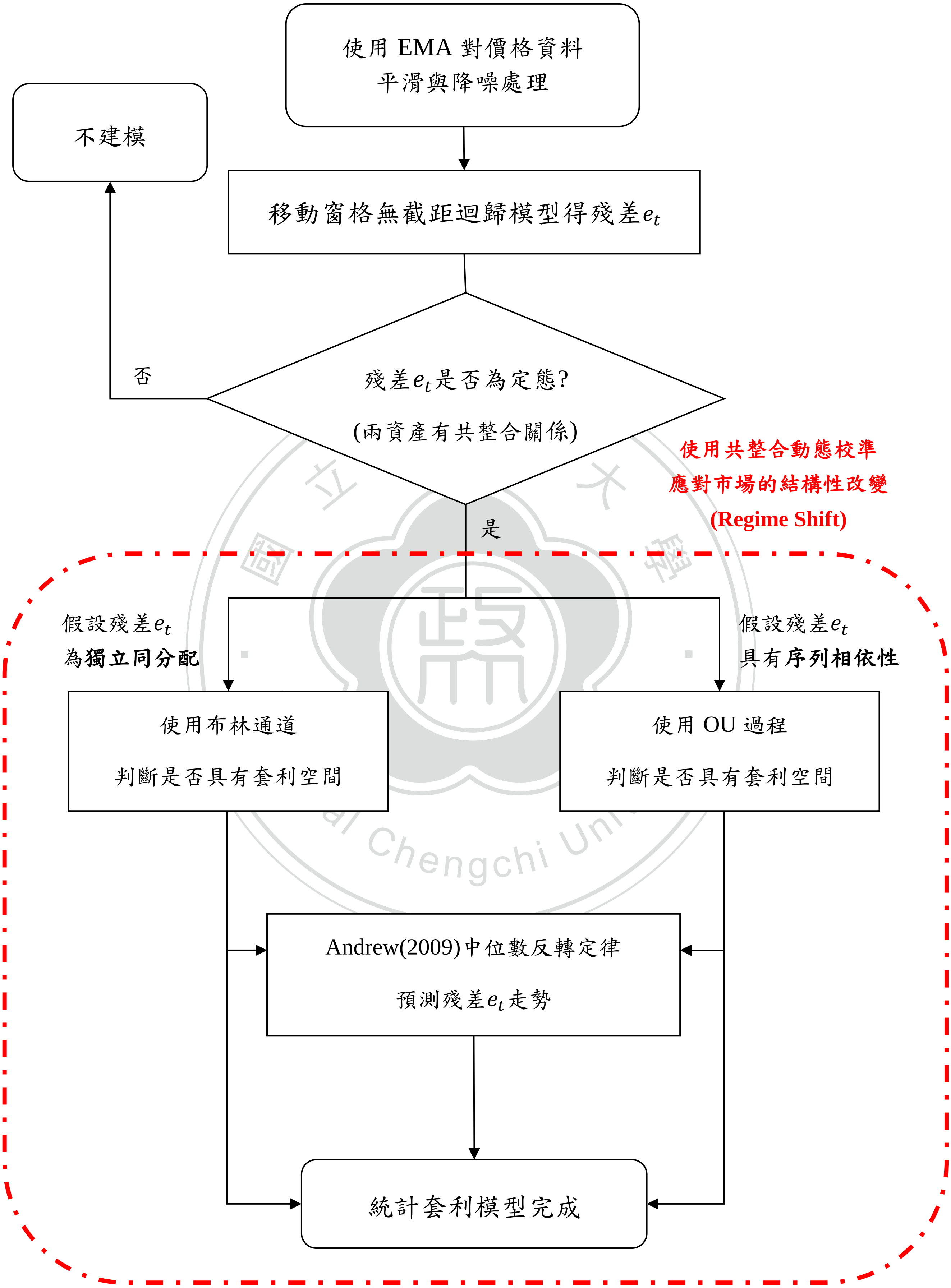
### 第三章、　研究方法

首先會從統計套利的概念開始講起，第二步介紹建模所需要的資料是如何被篩選以及降低雜訊，第三步說明模型如何被選擇以及原因，第四步講解支撐統計套利的核心理論-共整合關係，第五步說明模型如何根據市場的改變進行動態校準，第六步講解模型如何對市場異常進行判斷解讀，第七步則是根據歷史資料對市場做出預測，最後做出量化模型。

下頁以圖解的方式說明模型建構的整體流程:



DOI:10.6814/NCCU201900124






DOI:10.6814/NCCU201900124

第四章、　實證結果與分析

本研究的實證結果與分析，分為選用一分鐘頻率的原因、所選用的商品介紹，以及布林通道及 OU 過程統計套利策略實證結果。

第五章、　總結與展望

針對實證結果分析進行說明與整理，並對後續研究方向提出建議以及可改進之方向。

# 第二章　文獻探討

## 第一節　共整合統計套利

統計套利是廣泛被使用的交易策略，在學術界是相當重要的研究議題，相關的文獻內容也很龐大，本研究僅探討與本研究使用的統計套利方法相關之重要論文。Vidyamurthy (2004)中選擇使用的 Engle and Granger (1987)二階段共整合檢定法，檢定兩個高度相關的價格時間序列間是否存在長期均衡關係，二階段共整合檢定法首先以單根檢定確認配對的成份資產價格為 I(1)序列(一階整合序列)，接著將其中一個資產價格序列為自變數，另一資產作為應變數建立簡單線性迴歸關係，在對其殘差進行單根檢定，若結果為顯著，表示殘差為定態，則兩價格序列間具有共整合關係，可嘗試以其為標的建立統計套利策略。在確定兩資產有共整合性質後，便需判斷交易的進出場訊號，在交易訊號方面，Vidyamurthy (2004)設置布林通道，並去計算其標準差與迴歸均值，若殘差序列偏差超過此最佳門檻值為啟動交易條件。



DOI:10.6814/NCCU201900124

## 第二節　　指數移動平均在量化模型建構上的優勢

Andrew (2009)中提到價格中時常帶有過多的雜訊，會使得計量模型變的不準確且失去預測意義，所以必須針對原始資料進行平滑化與降噪處理，去除不必要的雜訊。他使用的方式是先將價格採用指數移動平均(以下稱 EMA,Exponential Moving Average)。一般常用的簡單移動平均(以下稱 SMA,Simple Moving Average)採用固定長度的資料窗格，並且全部使用相同的權重，對照來說，EMA 方案則是對整個歷史資料，應用了一個隨著指數遞減的權重值。這樣做的好處是儘管過了很久的大型價格衝擊事件，對於現在還是有一定的影響力，只是對於目前的計量模型估計與預測來說，比較靠近現在的資料，影響力還是比較大。以這種方式投射出來的預測模型，它的形式就跟 SMA 類似，是一種線性的關係。不過，計算出來的值是不同的。EMA 的計算，是以遞迴的方式計算。它的參數可以決定過去的觀察值，將會以多快的速度，變得與目前的估計值不相關(等效的說法是，它決定了要取多久之前的歷史資料，來計算目前的估計值)。Andrew(2009)提到只有在原始序列出現巨幅度地劇烈變化時，SMA 與 EMA 才會出現可觀而明顯的差異，如果原始序列遭遇了跳空變化，或者是有明確的趨勢時，這兩種情況下，SMA 都無法處理的很好，而 EMA 所表現出來的彈性，在這兩方面卻能展現出很明顯的功用。

## 第三節　　中位數反轉定律

Andrew(2009)中提到關於機率定理於價差移動上的運用，它呈現的結果是一個簡單的機率定理，表明了基本的反轉定律，保證在一個有效率的市場中，必然會存在著價差反轉的現象。Andrew (2009)證明了即使價差序列分配並非常態且具有厚尾現象，還是潛在有反轉的現象。總而言之，所有的分配，都有可能出現反轉的現象。這個定理的出現為統計套利補強了理論的不足，並且讓後續的研究



DOI:10.6814/NCCU201900124

有更穩固的理論支持。

中位數反轉定律告訴我們，如果價差為獨立同分配，那$t_0$的價差大於中位數，則$t_1$的價差小於$t_0$的價差的機率為百分之七十五，反之亦然。這個結果讓我們要解決的問題從價差序列是 Random，背後沒有相同的隨機分配，降低難度變成價差序列 Stochastic，每一個時刻的價差雖是隨機，但是背後具有相同的隨機分配。

# 第三章　研究方法

## 第一節　統計套利理論

統計套利的名稱源於它的核心功能，即利用歷史資料發現統計意義上的持續現象，並且通常這些現象是基於統計面的。此類統計意義上的持續關係可以存在於資產的價格與公司最新申報的營利數據上。這樣的關係也可以存在於多個資產的相對價格上。比如一個資產的價格水平和另一個資產的價格水平的差異。而我們所要研究的目標是屬於後者，關於兩資產間的價格差異。

我們可將兩檔不同的資產一個時段的價格以$\{P_i\}_{i=1}^{N}$表示，兩檔資產價格可能存在連動關係如圖 1:

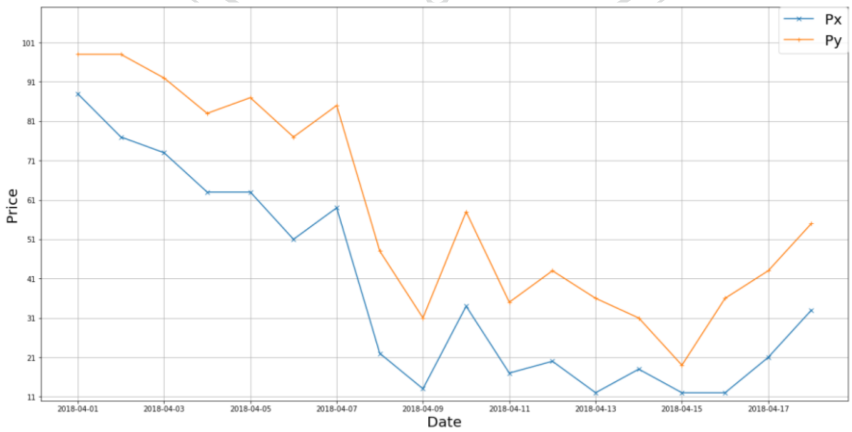



圖 1:兩資產走勢示意圖



DOI:10.6814/NCCU201900124

我們可以從圖 1 中發現兩檔資產的似乎存在某種連動關係，接著我們可以運用簡單迴歸分析，算出兩檔資產的連動關係:

$$P_{y,t} = \hat{\beta}_0 + \hat{\beta}_1 P_{x,t} + e_t \tag{1}$$

我們將式子移項後可以得到:

$$P_{y,t} - \hat{\beta}_1 P_{x,t} = \hat{\beta}_0 + e_t \tag{2}$$

此式代表的意義是我們從歷史價格資料得知，透過資產 Y 之價格$P_{y,t}$與資產 X 之價格$P_{x,t}$的線性組合，可以獲得長期均衡的$\hat{\beta}_0$加上殘差$e_t$。若 $P_{y,t} - \hat{\beta}_1 P_{x,t}$ 太大(即$P_{y,t} - \hat{\beta}_1 P_{x,t} > \hat{\beta}_0$)，則代表$P_{y,t}$過高，$P_{x,t}$過低，我們預測它會迴歸均值$\hat{\beta}_0$，而進行賣空一單位資產 Y，買入$\hat{\beta}_1$單位資產 X，反之當$P_{y,t} - \hat{\beta}_1 P_{x,t}$太小(即$P_{y,t} - \hat{\beta}_1 P_{x,t} < \hat{\beta}_0$)，則代表$P_{y,t}$過低，$P_{x,t}$過高，我們預測它會迴歸均值$\beta_0$，而進行買入一單位資產 Y，賣出$\hat{\beta}_1$單位資產 X,如圖 2:

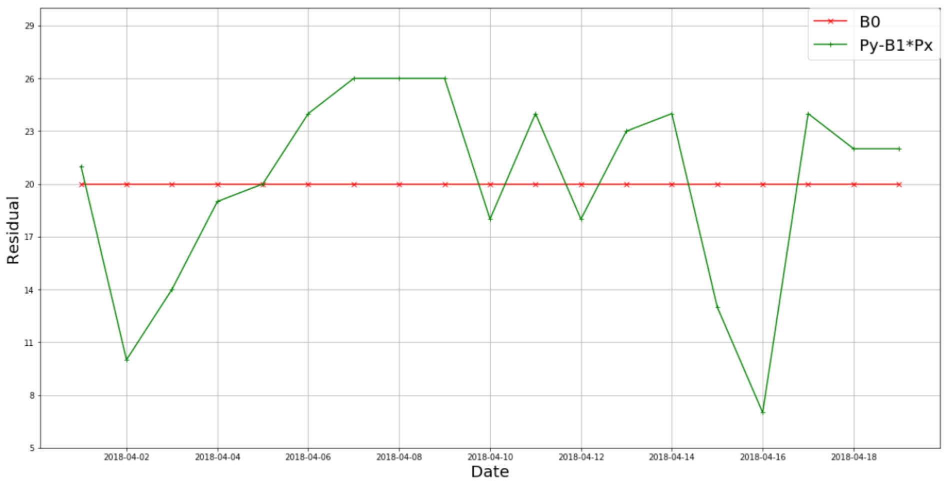



圖 2:均值迴歸示意圖



DOI:10.6814/NCCU201900124

這種市場中立性投資策略(Market Neutral Strategy)*[1]的好處是我們能避免系統性風險(System Risk)*[2]，在市場嚴重下跌時不會造成過大的損失。總結一下這小節，在建構統計套利策略時我們利用兩檔資產進行配對組合，找到一個適當持有比例便可以得到不受系統性風險影響的套利組合。

## 第二節　指數移動平均與報酬率形式

本章節我們將對資料降噪並定義輸入資料與介紹何為當期報酬率的概念。傳統的統計套利，將價格資料直接帶入並建置模型，但這樣會產生一個問題。依據市場效率假說(簡稱 EMH,Efficient Market Hypothesis)，我們假設市場為無效率市場，資產存在一個真實價值，但在受到消息時，價格並不會一次反應到位，有時會過度反應有時則會反應不足，這時的我們所看到的價格並非資產的真實價值，而是有偏差的，那若以這個價格帶入模型則模型會完全失準，以下圖 3:

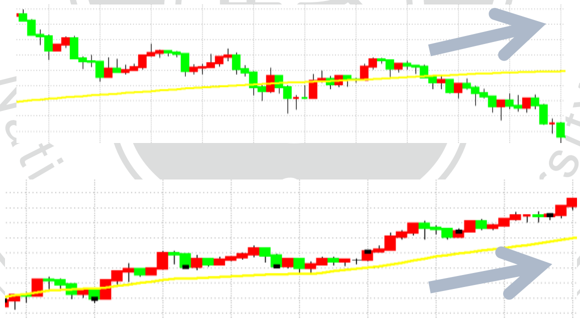


圖 3:真實價值與價格偏差圖

[1] 市場中立性投資策略(Market Neutral Strategy):市場中立策略是指屬於 Relative Value Strategy(相對價值策略)的一種投資策略，市場上某些資產的價格變化時，其變化的幅度和時間不同。利用這種差別去獲取利潤。並利用同時做多做空來降低系統性風險。

[2] 系統性風險(Systematic Risk）:證券投資風險包括系統性風險和非系統性風險，按照風險形成的原因，又可以分為外部因素造成的風險和投資者自身內因造成的風險。所謂系統性風險是指由於全局性的共同因素引起的投資收益的可能變動，這種因素以同樣的方式對所有證券的收益產生影響。



DOI:10.6814/NCCU201900124

我們假設兩條黃色線分別為資產 X 的真實價值與資產 Y 的真實價值，其實它們是存在連動關係的(齊漲)，但由於市場非完全有效而存在價格偏差，使得兩資產價格看起來呈現反向關係，如果我們只以這段期間的價格當作輸入資料，則會得出兩資產為負相關的結論。為了解決這個問題，我們使用指數移動平均(EMA)與當期報酬率的概念去平滑、降低雜訊，並試著找出資產的真實價值，以利後面量化模型的建構。

EMA 是 Exponential Moving Average 的簡稱,即指數移動平均。是時間序列分析中常用到的一種類型平均值。簡單來說，EMA 就是一個加權平均值。它的特別之處在於:

1. 隨著時間流逝，舊觀察值的權重將會呈現指數衰退(Exponential Decay)。
2. 等間距的時間序列$\{X_t：t = 1, \ldots\ldots, \infty\}$(等間距：即兩個相鄰樣本之間的時間間隔是不變的)。
3. 其中參數為$\lambda$，$\lambda > 0$掌控著指數衰減的程度，$\lambda$越大，權重隨時間衰減的越快。

這邊給出 EMA 的迭代公式定義:

$$EMA_t = \begin{cases} EMA_0, & t = 1 \\ (1-\lambda)EMA_{t-1} + \lambda X_t, & t \geq 2 \end{cases} \tag{3}$$

其中$EMA_0$是 EMA 的初始值。從$EMA_t$的迭代公式出發，持續迭代進去$EMA_{t-1}$，$EMA_{t-2}$，…………，$EMA_{t-k}$，我們得到:

$$EMA_t = \lambda[X_t + (1-\lambda)X_{t-1} + (1-\lambda)^2 X_{t-2} + \cdots (1-\lambda)^k X_{t-k}] + (1-\lambda)^{k+1} EMA_{t-k-1}$$

假設我們有無窮多個觀察值，那麼可以把上式進一步展開為:

$$EMA_t = \frac{X_t + (1-\lambda)X_{t-1} + (1-\lambda)^2 X_{t-2} + \cdots + (1-\lambda)^k X_{t-k} + \cdots}{1 + (1-\lambda) + (1-\lambda)^2 + (1-\lambda)^3 + \cdots + (1-\lambda)^k + \cdots}$$



DOI:10.6814/NCCU201900124

其中我們用到了泰勒序列$\sum_{n=0}^{\infty}(1-\lambda)^n = \frac{1}{\lambda}$。

我們使用 EMA 的兩個主要原因如下:

1. 從以上公式我們可以看出 EMA 的優點在於對整個歷史資料，應用了一個隨著時間衰退的權重值。這麼一來，儘管是過了很久的大型事件，對於現在還是有一定的影響力，只是對於目前的估計與預測來說，比較靠近現在的資料，影響力還是比較大的。

2. 由於我們使用的價格資料最小粒度是一分鐘，一分鐘的價格資料常常帶有許多的雜訊與價格偏差，若把帶有雜訊的資料直接丟入量化模型當中，會大幅降低量化模型的可用性，EMA 在這方面很好的幫助我們平滑資料。

接下來我們將比較價格與報酬率建模的優劣。

我們一般對於資產價格的報酬率定義為:

$$\frac{P_t - P_{t-1}}{P_{t-1}} \text{ 或 } ln\frac{P_t}{P_{t-1}}$$

其中$P_t$為資產價格，前者為離散型的報酬率，後者是為了在計算上方便的連續型報酬率，但這兩種報酬率形式對於統計套利模型的建構是不友善的，因為我們的目標在於找出可以交易的預測模型。我們理想中模型是能夠預測兩檔資產價格差異會均值迴歸，這樣我們才有辦法透過兩檔資產算出適當的對沖比例進行交易，但若直接以價格作為輸入資料又會發生上述所說的雜訊問題，且實務上兩資產常常歷史均價差異極大，若直接當作輸入資料，所產生的模型，會有不小的偏差。

以下整理報酬率與價格對統計套利建模的優劣比較:



DOI:10.6814/NCCU201900124

1. 報酬率(適合建置量化模型):

優點:

(1). 在統計套利模型建構上，更容易找到定態時間序列，若以定態時間序列作為迴歸分析的輸入資料，不會產生假性迴歸(Supurious Regression)。

(2). 絕大多數具有健康成交量的金融資產報酬率都有隱性邊界(例如：極少發生連續兩天漲 50%這種狀況)，在這樣的條件下，每檔資產的輸入資料大多落在一定的範圍(約-10%~10%)之間，這種有邊界的資料對模型建製的穩定性是比較好的。

缺點:

(1). 由上述報酬率的定義可知報酬率是流量的概念(只與前一期有關)，但對統計套利類型的交易策略而言我們需要的是存量(歷史價值的累積)，我們沒有辦法使用會均值迴歸的報酬率去建構交易策略。就一檔資產來說，我們在市場上交易的是商品的價格，我們可以說若以上述報酬率形式所做出來的交易模型，實務上而言並沒有辦法執行交易。

2. 價格(適合實際做交易):

優點:

(1). 價格是市場上最直接明確的資訊，以統計套利模型來說，我們可以直接透過價格差異是否均值迴歸在市場上進行交易買賣。

缺點:

(1). 如果價格未經過處直接拿來建構統計套利量化模型，很容易發生假性迴歸(Supurious Regression)的問題，原因在於價格絕大多數時間都不是定



DOI:10.6814/NCCU201900124

態序列。

(2). 統計套利的兩個標的常常歷史均值差異極大(例如:大立光(3008)與大同(2371),相差大於 100 倍) ，兩個價格資料量級不同，在這種情形下，以相同地位當作模型輸入資料，會導致量級大的輸入主導整個模型。

(3). 如本節開頭所說，我們所看到的價格並不一定代表資產的真實價值。

我們知道報酬率與價格的優缺點後，我們嘗試把價格資料做以下的處理:

$$X_t = \frac{P_{x,t} - EMA_{x,t}}{EMA_{x,t}}\ ,\ Y_t = \frac{P_{y,t} - EMA_{y,t}}{EMA_{y,t}} \tag{4}$$

其中$P_{x,t}$為資產 $X$ 在 t 時間點的價格, $P_{y,t}$為資產 $Y$ 在 t 時間點的價格。它代表的意義是當期報酬率，核心概念是我們認為資產的理論價值是$EMA_{x,t}$(或$EMA_{y,t}$)，而價格$P_{x,t}$(或$P_{y,t}$)與理論價值的差異即為該資產當下相對於歷史價值的報酬(或短期價格偏差)。

這種處理方式結合了報酬率與價格的優點，它能夠反應具有深遠影響但事發以久的重大事件訊息，也能有效平滑短期內過多雜訊的問題，且擁有報酬率具有邊界的好特性，定態的特性更比價格好上許多讓模型補足統計理論上的不足。

## 第三節　無截距迴歸模型

本章節我們將會探討傳統的統計套利模型與我們所使用的模型有何差異，並決定本論文想要深入研究的目標。

傳統的統計套利使用的模型為簡單迴歸模型:

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + e_t \tag{5}$$



DOI:10.6814/NCCU201900124

$$X_t = P_{x,t}, Y_t = P_{y,t}$$

其中$P_{x,t}$為資產 $X$ 在 t 時間點的價格, $P_{y,t}$為資產 $Y$ 在 t 時間點的價格。

但在上節我們提到了我們的輸入資料是當期報酬率,若把輸入資料帶入模型後經過移項會得到:

$$Y_t - \hat{\beta}_1 X_t = \hat{\beta}_0 + e_t$$

$$X_t = \frac{P_{x,t} - EMA_{x,t}}{EMA_{x,t}}\ ,\ Y_t = \frac{P_{y,t} - EMA_{y,t}}{EMA_{y,t}} \tag{6}$$

其中的$Y_t - \hat{\beta}_1 X_t$在統計套利模型中的意涵為買入一單位的 Y 資產的當期報酬率並且賣出$\hat{\beta}_1$(對沖比率)單位的 X 資產的當期報酬率，會留下無法解釋的$\hat{\beta}_0$，$\hat{\beta}_0$在這也代表報酬率，但它並不是當期超額報酬率，我們無法解釋$\hat{\beta}_0$這個報酬率的意義，對後續殘差$e_t$(當期超額報酬率)的研究也會產生問題*[3]，所以我們捨棄這種做法，採用無截距迴歸模型當作本論文的模型。

$$Y_t = \hat{\beta}_1 X_t + e_t$$

$$X_t = \frac{P_{x,t} - EMA_{x,t}}{EMA_{x,t}}\ ,\ Y_t = \frac{P_{y,t} - EMA_{y,t}}{EMA_{y,t}} \tag{7}$$

我們對兩資產的當期報酬率作無截距迴歸模型，在將它移項會得到:

[3] 其實換成無截距迴歸模型的形式，我們從統計套利模型的角度出發，一樣是沒辦法再實務上交易當期報酬率的。但是若我們改變思維，把整個式子看成是一種訊號的概念，看 X,Y 的當期報酬率是否過高或過低，進而去產生交易即可。



DOI:10.6814/NCCU201900124

$$Y_t - \hat{\beta}_1 X_t = e_t$$

$$X_t = \frac{P_{x,t} - EMA_{x,t}}{EMA_{x,t}}\text{，}\ Y_t = \frac{P_{y,t} - EMA_{y,t}}{EMA_{y,t}} \tag{8}$$

其中$Y_t$是 Y 資產的當期報酬率，$X_t$是 X 資產的當期報酬率，$\hat{\beta}_1$是對沖比率，$e_t$殘差則代表當期超額報酬率。至此，我們已經建立完成我們要研究的目標，也就是殘差$e_t$(當期超額報酬率)。我們希望透過殘差$e_t$均值迴歸的特性，交易資產 X 與資產 Y，獲取報酬。

## 第四節　共整合關係與 DF 單根檢定

本章節將探討 Engle and Granger 二階段共整合檢定與 Dickey and Fuller 單根檢定，這兩個統計檢定在研究中的意義皆為確認我們從兩資產 X,Y 進行配對後所產生出的殘差$e_t$(當期超額報酬率)，是否為定態序列。若$e_t$為定態序列，才有接下去進行統計研究的意義。若為非定態序列，則後續研究將不具統計可信度。

Engle and Granger (1987)提出共整合的理論，是指兩個或多個時間序列存在一個長期穩定均衡的關聯性，若兩個以上非定態的時間序列變數經過線性組合後形成的新序列為定態，則稱他們之間具有共整合關係。具體來說此多個序列分別為非定態，但具有共同的趨勢，使其經線性組合後可將共同隨機趨勢消除，成為一個新的定態時間序列。

在檢定兩個序列是否具有共整合關係前，須先確定兩序列的整合階次(Order of Integration)相同，一時間序列的整合階次為該序列成為定態序列前須經過的最少差分(Difference)次數，舉例來說若時間序列 $r_t$ 經過 d 次差分後始為定態,則 $r_t$ 之整合階次為 d，一般表示成 $r_t$~I(d)，整合階次相同的兩序列才有存在共整合關係的可能，大部分的總體經濟變數或價格序列整合階次通常為 1，經過一階差分後



DOI:10.6814/NCCU201900124

即可成會定態序列。

Engle and Granger (1987)提出二階段共整合檢定法,用以檢定兩 I(1)序列是否具有共整合關係，假設 $X_t$、$Y_t$ 為一組非定態時間序列，兩者經一階差分後皆為定態，且假設此組序列配對僅存在一組共整合關係，首先估計兩者關係:

$$Y_t = \hat{\beta}_1 X_t + e_t \tag{9}$$

$$X_t = \frac{P_{x,t} - EMA_{x,t}}{EMA_{x,t}}，\ Y_t = \frac{P_{y,t} - EMA_{y,t}}{EMA_{y,t}}$$

接著對殘差項$e_t$進行 DF 單根檢定。

陳旭昇 (2013)時間序列資料對外在衝擊(Impulse)的反應大致上可分為兩類:定態(Stationary)時間序列及非定態(Non-stationary)時間序列。一個外生衝擊對非定態時間序列造成的影響可能不會隨時間消散，使其在時間的變動過程中留下永久性的影響並不斷累積，導致其具有隨機趨勢，即序列的平均數、變異數及各期間的共變異數並不固定且隨時間變動；定態時間序列僅會對外生衝擊產生暫時性的反應，一段時間過後此衝擊造成的影響將逐漸消失，使這樣的時間序列隨時間經過仍能保持穩定。若一時間序列$e_t$滿足下列性質，則我們稱$e_t$為定態:

1. $\mathrm{E}(e_t) = \mu$，即期望值$\mu$為固定常數

2. $Cov(e_t, e_{t-1}) = \gamma_l$ ，不同期間的共變異數只與落後期數$l$相關

上述定義為弱定態(Weak Stationary)，在大多數時間序列模型分析及迴歸分析中皆需要資料為弱定態，Granger and Newbold (1974)提出若以非定態變數進行迴歸分析將出現假性迴歸(Spurious Regression)的現象，即兩個不相關的序列變數出現極高的相關係數 $R^2$，因此在對時間序列資料進行分析前一般需要先檢定該



DOI:10.6814/NCCU201900124

序列是否為定態序列。

單根檢定(Unit Root Test)為檢定時間序列是否為定態的方法，若一時間序列具有單根則該序列有隨機趨勢，表示該序列為非定態序列，換而言之可藉由檢定一時間序列是否具有單根來確認其是否為定態序列。

本研究採用 Dickey and Fuller (1979)提出的 DF 檢定(Dickey-Fuller Test)，其模型為:

$$\Delta e_t = \emptyset e_{t-1} + \varepsilon_t \tag{10}$$

以上之模型檢定假設為:

$$H_0 : \emptyset = 0(\text{存在單根}); H_1 : \emptyset < 0(\text{不存在單根，即序列為定態})$$

其檢定統計量為:

$$\mathrm{DF} = \frac{\widehat{\emptyset} - 1}{\sqrt{\mathrm{Var}(\widehat{\emptyset})}} \tag{11}$$

如果拒絕了虛無假設，表示$e_t$沒有單根，為一定態時間序列；反之，該序列則為具單根現象之非定態序列。

許多研究會探討落後期數與時間趨勢對單根檢定的影響，而我們為簡化篩選流程，以利於在大量的統計套利時段中選取配對，僅選用 DF 檢定，且不考慮截距項$\beta_0$及時間趨勢$t$，以簡化單根檢定對後續統計套利之影響。

我們對殘差$e_t$進行 DF 檢定，若確認為定態序列則表示此序列在統計上可研究，也代表具有均值迴歸的特性，我們會以這段序列做為研究主題。至此，我們





DOI:10.6814/NCCU201900124

運用共整合關係與 DF 檢定確定了判斷模型可否建構的依據，即為殘差$e_t$是否為定態序列。

## 第五節　結構性改變與共整合動態校準

本章節會先介紹傳統使用的兩個模型建構方法的優點與缺點，接著會結合此二方法的優點嘗試新的模型建構方法。傳統的統計套利通常對資料做一次切割，將樣本區分為樣本內與樣本外資料，並在樣本內建立迴歸方程式，藉以算出$\hat{\beta}_1$(對沖比率)與共整合結果，而進行統計套利，如下圖 4:


圖 4:共整合狀態崩解圖

若我們取樣本內為 2016/10 以前(紅線左邊)，則很有可能這段時間兩個時間序列呈現共整合關係，且其對沖比率約會在 1(兩個資產價格糾纏在一起)上下，但如果我們把這個結果直接應用在樣本外會發現完全不符合實際情況。這種方法



DOI:10.6814/NCCU201900124

隱含的假設是市場不會發生結構性改變，兩資產的關係在未來會不斷重演。這個假設的好處在於兩資產的關係沒有被打破，那它們的價格關係將永遠不會改變，也就是我們所設計的量化模型會永遠有效果，但是很明顯如上圖來看，絕大多數資產的共整合關係都會發生結構性改變，使得這種建模方式在統計套利上來說太過於僵硬，沒辦法隨著市場結構性的改變去做應對。

另一種常用的方法為純粹移動窗格的方式，也就是每過一段時間，自動重新建構量化模型，如下圖 5:

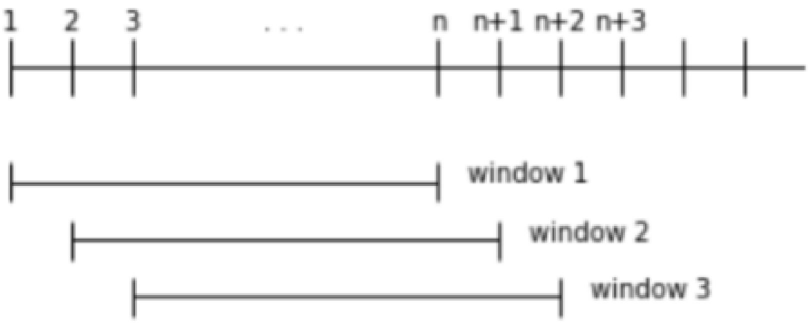



圖 5:移動窗格建模示意圖

這種方法改善了將樣本分為內與外的僵硬方式，進而做到了靈活的效果，但同時，這種方式也存在一個問題，那就是我們要如何決定我們的時間間隔參數?時間本身對於在盤中的交易行為本身並不具任何意義，舉個比方，假設我們的時間參數是一小時，也就是說我們每過一小時會重製我們的量化模型，但是市場的結構性改變並不會如你所願剛好每次都在一小時那刻發生，它有可能在三十分發生，又或是甚至二十分，更甚至五分...在這種固定時間重製模型的方式，基本上就是假設市場會在固定時間有結構性改變，但從歷史的角度來看，這項假設並不合理。

我們改善了上述兩種方式的缺點並保留其優點，運用的方式為以共整合持續的關係作為整個模型的存亡依據，也就是說我們模型會先以目前的時間點往回蒐



DOI:10.6814/NCCU201900124

集某段時間長度的資料，並且判斷這段時間序列是否存在於長期穩定的關係($e_t$符合統計性質的定態時間序列)，如果確認$e_t$為定態序列(兩資產具有共整合關係)，那就對這段時間的資料進行建模。模型的持續時間則是依據持續不斷的共整合檢定去作決定，倘若共整合檢定隨著時間的推移持續不斷通過，那我們的模型將保持固定，不做任何改動。這種建模方法隱含的假設是我們認為若一段時間序列在加入新資料後不斷的被確認它是定態，代表這段時間序列皆為同一母體所產出的隨機變數，也就是他們具有相同的期望值與變異數。倘若隨著時間的推進，新資料的加入使得共整合關係在某個時刻崩解，則我們的模型將會破壞，強制平倉正在進行的交易。此行為的意涵是我們認為若一段時間序列被確認它已經不是定態，那代表它發生了結構性改變。在這個前提之下，我們依據之前的資料所做出的量化模型將不在適用。模型破壞後，我們在重新找尋新的序列時段，並從最新時間點往回蒐集某段時間長度的資料，倘若不通過共整合檢定，則不建製模型，倘若通過，則重新建模。以下用圖解方式說明我們的建模過程,如圖 6:

每一個方塊代表隨著時間進行而產生的殘差$e_t$，橘色方塊係為具有共整合關係的定態序列，黑色方塊為共整合狀態消失。

Status1：取 N 個單位的資料確認是否具有共整合關係，共整合檢定通過，建製量化模型。

Status2：新資料$e_t$加入且通過共整合檢定，固定參數，沿用 Status1 建製的模型。

(隱含假設新進入的序列資料為同個分配產出的隨機變數)

Status3：同上

Status4：黑方塊出現，共整合關係崩解。

(隱含假設序列發生結構性改變，代表之前的模型不在具有效果，模型破壞)

Status5：共整合檢定不通過，序列並非定態，不建模。



DOI:10.6814/NCCU201900124

Status6：同上。

Status7：取 N 個單位的資料確認是否具有共整合關係，共整合檢定通過，建製量化模型。

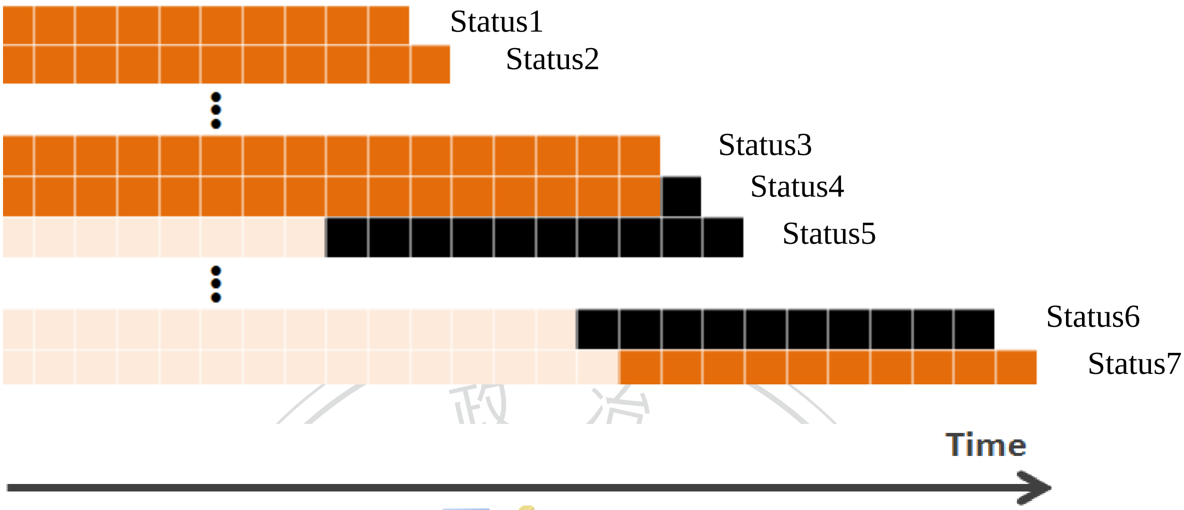



圖 6:共整合動態校準示意圖

至此，我們依據動態校準與共整合關係確定了模型能否使用的準則。

## 第六節　布林通道技術指標

六到八章節我們將提到如何判斷$e_t$是否異常，$e_t$異常代表市場具有統計上的可套利空間存在，我們希望$e_t$是具有均值迴歸特性的時間序列，那我們就可以透過資產 X 與資產 Y 的線性組合產生$e_t$，並且去依據$e_t$的歷史資料，判斷它目前所在的位置，預測它之後可能的走向，進行套利。本章節的假設為每筆殘差資料$e_t$皆為獨立同分配，我們利用信賴區間的概念去預測$e_t$是否過於偏差的，運用的方法類似技術指標步林通道。

Bollinger (1992)中提出了布林通道(Bollinger Bands)的概念，此通道為一個技術分析工具，主要以一條 n 日移動平均線為中心線，向上和向下各加減 c 倍標準差後，就構成了上界和下界，而由此上下界所圍成的帶狀區間即為布林通道。布



DOI:10.6814/NCCU201900124

林通道的原始設定為:

1. 上界(Upper Bond) = 中心線 + 2 個 20 日標準差

2. 中心線 = 20 日移動平均線

3. 下界(Upper Bond) = 中心線 - 2 個 20 日標準差

布林通道結合了移動平均線和統計學的標準差的概念。其基本的型態，是由 3 條軌道線組成的帶狀通道。中心線移動平均線為平均成本的概念，而上界、下界則為股價壓力及支撐訊號的判斷指標，利用統計的概念在價格偏離平均值過多時，認為價格會迴歸均值，利用此訊號進場獲利，一般為逆勢策略，在碰到上軌時做空標的，在碰到下軌時做多標的，布林通道的策略變化性多樣。

本研究的布林通道建立基於上一章所講述的共整合動態校準建模方法建製，不同於原本布林通道的點在若序列為定態，我們就不會隨時間推進更改通道的大小，直到共整合關係崩解後再次建模產生新的布林通道才會改變。

## 第七節 OU 過程與 AR(1)模型

本章節的前提假設為殘差$e_t$是具有序列相依性(也就是說，序列並不是獨立的)，我們建構時間序列模型去預測$e_t$是否過於偏差的，運用的方法為隨機過程中的 OU 過程。

OU 過程的正式表達如下:

$$de_t = \theta(\mu_e - e_t)dt + \sigma dW_t \tag{12}$$

其中$\theta$為復歸速度，$\sigma$為標準差，$\mu_e$為均值，$W_t$為維納過程(Wiener Process)。

OU 過程具有均值迴歸的特性的，當$e_t$>$\mu_e$時，在不考慮隨機變數$W_t$的情況下，



DOI:10.6814/NCCU201900124

$de_t$<0，使得$e_t$在那一瞬間有一個向下的斜率。

我們假設$e_t$是一個 OU 過程，為了簡化模型，我們考慮 OU 過程的離散形式：

$$e_t - e_{t-1} = \theta(\mu_e - e_{t-1}) + \sigma(W_t - W_{t-1}) \tag{13}$$

(在離散的情況下$\theta(\mu_e - e_t)$中的$e_t$把它當成$e_{t-1}$，因為離散狀況下沒有辦法達到"瞬間"，只能通過差分體現)

離散的維納過程描述的場景為：$W_1 \ldots\ldots W_t$ 時間序列的每一個增量(increment) $W_t - W_{t-s}$，$\forall\, t$ 都是獨立同分配的，$W_t - W_{t-s} \sim N(0, s)$，所以可簡化成：

$$e_t - e_{t-1} = \theta(\mu_e - e_{t-1}) + \varepsilon_t\,, \varepsilon_t \sim N(0, \sigma^2) \quad \forall\, t \tag{14}$$

上式中的$e_t$由前一期的$e_{t-1}$與均值的偏離程度$\mu_e - e_{t-1}$，和復歸速度$\theta$決定。

進一步可以簡化成帶漂移項(Drift Term)的 AR(1)過程：

$$\begin{gathered} e_t = a + be_{t-1} + \varepsilon_t \\ a = \theta\mu_e\text{，}b = 1 - \theta, \varepsilon_t \sim N(0, \sigma^2) \quad \forall\, t \end{gathered} \tag{15}$$

我們可以證明 $e_t \sim N(\mu_e, \frac{\sigma^2}{2\theta - \theta^2})$ 是弱定態序列。(證明請見附錄二)

為了簡化模型，我們使$\sigma^2 = 1$，並用 OLS 最小平方法對$a, b$進行估計進而算出$\widehat{\mu_e}$與$\hat{\theta}$，在利用$e_t$的分配建立信賴區間，當$\hat{\theta} > 0$且$e_t$偏離信賴區間時即判斷它為異常。OU 過程的建立與布林通道一樣是基於共整合動態校準建模方法建製。

## 第八節　中位數反轉定律

這章節講述一個簡單但具有嚴謹數學支撐的機率定律，這個定律在我們的研



DOI:10.6814/NCCU201900124

究中扮演的是的效果，希望透過這個定律能夠提高模型預測的準確度。Andrew(2009)提到關於價差移動理論上的基礎，保證在一個有效率的市場中，必然會存在價差反轉的現象。我們可以證明下式: (證明請見附錄一)

$$\mathrm{P}[(e_t > e_{t-1} \cap e_{t-1} < m) \cup (e_t < e_{t-1} \cap e_{t-1} > m)] = 0.75$$

隨機變數$e_t$是 t 時點的殘差，而每一時點的$e_t$，都是獨立同分配的隨機變數，在統計上等價於 Andrew(2009)的價差，$m$ 代表殘差中位數。如果用交易行為的方式描述的話，我們只要確認了 t-1 時點的殘差，並觀察到殘差大於中位數(或小於中位數)，就可以對 t 時點的殘差進行方向性預測，並且準確率高達 75%。

這個基礎定律並不只存在於理論上，Andrew(2009)之後對其放寬了許多假設，包括離散的狀況、非常數變異數、非常態分配、序列具有相關性。這個定律幫助我們有更好的統計準則可以預測$e_t$的走向，我們會將這個定律分別加入布林通道策略與 OU 過程策略，希望可以提高策略勝率。

## 第九節　統計套利策略的建構方法

### 一、　布林通道統計套利策略

通過篩選之配對時段，表示在該期間內兩檔資產經 Engle-Granger 二階段檢定具有共整合關係，利用 OLS 估計之參數$\hat{\beta}_1$，並計算殘差$e_t$:

$$Y_t = \hat{\beta}_1 X_t + e_t \tag{17}$$

$$X_t = \frac{P_{x,t} - EMA_{x,t}}{EMA_{x,t}} \text{，} \quad Y_t = \frac{P_{y,t} - EMA_{y,t}}{EMA_{y,t}}$$

其中$Y_t$是 Y 資產的當期報酬率，$X_t$是 X 資產的當期報酬率，迴歸資料長度取 60 分鐘，$\hat{\beta}_1$是對沖比率，殘差$e_t$則代表當期超額報酬率。



DOI:10.6814/NCCU201900124

接著在共整合動態校準的框架下，我們提出假設:

**殘差$e_t$為獨立同分配(Independent And Identically Distributed)。**

接著利用技術指標布林通道判斷進場做多及放空配對時機的訊號，可分為兩部分策略，其規則為:

1. 當殘差項$e_t$由下方向上穿越布林通道上緣 $\mu + 1.5\sigma$ 但小於$\mu + 2\sigma$時，進場放空資金比例為 1 單位的 Y 並做多資金比例$\hat{\beta}_1$單位的 X，並於$e_t$由上方向下穿越一倍布林通道上緣$\mu + 1\sigma$時，平倉出場。

2. 當殘差項$e_t$由上方向下穿越布林通道下緣 $\mu - 1.5\sigma$ 但大於$\mu - 2\sigma$時，進場做多資金比例為 1 單位的 Y 並放空資金比例$\hat{\beta}_1$單位的 X，並於$e_t$由下方向上穿越一倍布林通道下緣$\mu - 1\sigma$時，平倉出場。

其中μ為前 60 分鐘的$e_t$之平均，σ為前 60 分鐘的$e_t$之標準差，我們不另外進行參數的選擇，避免過度最佳化的問題。策略訊號的產生概念如圖 7:

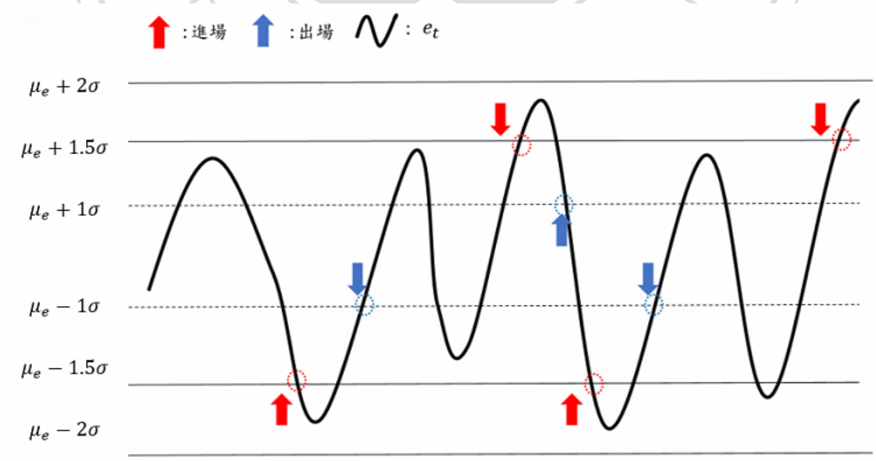



圖 7:布林通道策略示意圖

此布林通道策略的基本概念為當殘差項$e_t$偏離過多超出布林通道時，我們便認為兩資產相對報酬率在短期內有所偏差，進場建倉，建立多空對沖的投資組合，等



DOI:10.6814/NCCU201900124

未來$e_t$收斂回歸均值時平倉獲利。$\hat{\beta}_1$我們採用 0<$\hat{\beta}_1$<5 這個區間，防止短時間內價格過度偏差所產生的$\hat{\beta}_1$異常造成模型失增。

我們為求策略的簡易選用較基本的布林通道門檻值，僅以一點五倍標準差及兩倍標準差做為判訊號之通道，不再另用最佳化找尋最適合統計套利的通道標準差倍數，減少我們後面探討每個配對皆用不同參數組成投資組合時產生過度最佳化問題。

此外本研究為探討統計套利之市場中立報酬是否真實存在，因此在回測及交易中不考慮手續費及交易稅等交易成本，期望能真實看出配對交易的市場中立報酬效果。

## 二、 OU 過程統計套利策略

通過篩選之配對時段，表示在該期間內兩檔資產經 Engle-Granger 二階段檢定具有共整合關係，利用 OLS 估計之參數$\hat{\beta}_1$，並計算殘差$e_t$:

$$Y_t = \hat{\beta}_1 X_t + e_t \tag{18}$$

$$X_t = \frac{P_{x,t} - EMA_{x,t}}{EMA_{x,t}},\ Y_t = \frac{P_{y,t} - EMA_{y,t}}{EMA_{y,t}}$$

其中$Y_t$是 Y 資產的當期報酬率，$X_t$是 X 資產的當期報酬率，迴歸資料取 60 分鐘，$\hat{\beta}_1$是對沖比率，殘差$e_t$則代表當期超額報酬率。

接著在共整合動態校準的框架下，我們提出假設:

**殘差$e_t$具有序列相依性(Dependent Time Series)。**

接著利用 OU 過程判斷進場做多及放空配對時機的訊號，可分為兩部分策略，其規則為:



DOI:10.6814/NCCU201900124

1. 當殘差項$e_t$由下方向上穿越一點五倍標準差$\mu_e + 1.5\frac{\sigma}{\sqrt{2\theta-\theta^2}}$但小於兩倍標準差$\mu_e + 2\frac{\sigma}{\sqrt{2\theta-\theta^2}}$時，進場放空資金比例為 1 單位的 Y 並做多資金比例$\hat{\beta}_1$單位的 X，並於$e_t$由上方向下穿越一倍標準差$\mu_e + 1\frac{\sigma}{\sqrt{2\theta-\theta^2}}$時，平倉出場。

2. 當殘差項$e_t$由上方向下穿越一點五倍標準差$\mu_e - 1.5\frac{\sigma}{\sqrt{2\theta-\theta^2}}$但大於兩倍標準差$\mu_e - 2\frac{\sigma}{\sqrt{2\theta-\theta^2}}$時，進場做多資金比例為 1 單位的 Y 並放空資金比例$\hat{\beta}_1$單位的 X，並於$e_t$由下方向上穿越一倍標準差$\mu_e - 1\frac{\sigma}{\sqrt{2\theta-\theta^2}}$時，平倉出場。

其中$e_t \sim N(\mu_e, \frac{\sigma^2}{2\theta-\theta^2})$，$\mu_e$為 OU 過程前 60 分鐘的迴歸均值，$\frac{\sigma}{\sqrt{2\theta-\theta^2}}$為 OU 過程前 60 分鐘的波動程度，我們不另外進行參數的選擇，避免過度最佳化的問題。策略訊號的產生概念如下圖 8:

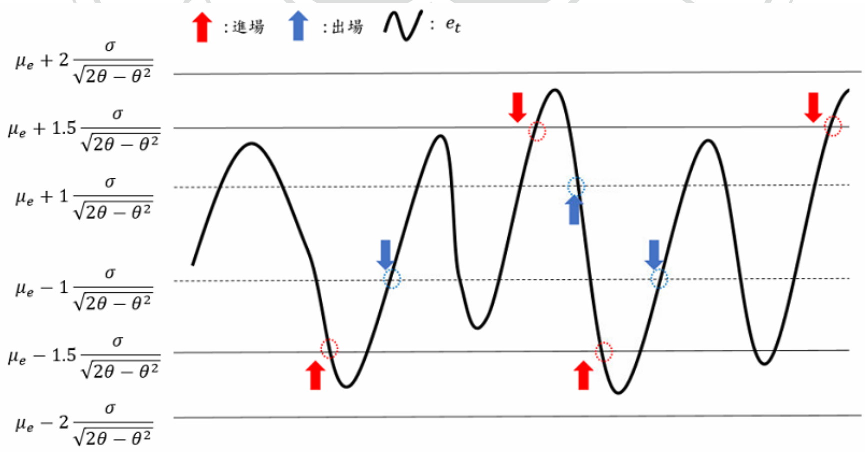



圖 8:OU 過程策略示意圖

此 OU 過程策略的基本概念為當殘差項$e_t$偏離過多超出範圍時，我們便認為兩資產相對報酬率在短期內有所偏差，進場建倉，建立多空對沖的投資組合，等未來



DOI:10.6814/NCCU201900124

$e_t$收斂回歸成正常水準時平倉獲利。

我們為求策略的簡易選用較基本的波動門檻值，僅以一點五倍及兩倍做為判訊號，不再另用最佳化找尋最適合統計套利的標準差倍數，減少我們後面探討每個配對皆用不同參數組成投資組合時產生過度最佳化問題。在迴歸速度$\theta$上，我們只採用$\theta$>0 的時間段，因$\theta$<0 代表發散而非收斂，$\hat{\beta}_1$我們採用 0<$\hat{\beta}_1$<5 這個區間，防止短時間內價格過度偏差所產生的$\hat{\beta}_1$異常造成模型失增。

此外本研究為探討統計套利之市場中立報酬是否真實存在，因此在回測及交易中不考慮手續費及交易稅等交易成本，期望能真實看出配對交易的市場中立報酬效果。

## 第十節　策略績效評估指標

1. 總報酬率(Total Return)

總報酬率為最直觀也最重要的評估指標之一，在統計套利中總報酬率高代表兩資產連動性高且套利空間大，相對的較低則代表資產連動性低或套利空間小。

2. 夏普率(Sharpe Ratio)

夏普率(Sharpe Ratio)為一個考慮報酬相對於風險的指標，其概念為每單位的風險能賺的報酬，數值越高表示績效越佳，最早於 Sharpe (1966)中被使於評估共同基金(Mutual Funds)之風險，在被廣泛應用後於 Sharpe (1994)正式發表取名。夏普值計算方法如下:

$$\text{Sharpe Ratio} = \frac{\mathrm{E}(R_p) - R_f}{\sigma_p} \tag{19}$$

其中$\mathrm{E}(R_p)$為投資組合預期報酬率，一般使用過去的平均報酬，$R_f$為無風險利率，



DOI:10.6814/NCCU201900124

我們設定為 1%；$\sigma_p$為投資組合的風險，一般使用投資組合報酬的標準差。

夏普率代表交易模型的穩定性。在統計套利中總報酬率很高但是波動太大，代表模型常常判斷錯誤導致對沖不完全造成方向性損失或方向性獲利，但這個並不是統計套利模型所追求的，一般高頻的統計套利模型，都會在盡量讓部位在持倉時間不受市場波動干擾，因此都會有較高的夏普值。

3. 每筆交易平均報酬率(Per Average Return)

每筆交易平均報酬率可以讓我們知道平均每次交易各資產之間所捕捉的套利空間大小，對套利空間大的組合往後可以進行閥值的優化，而不單單只以主觀判定兩倍標準差或一點五倍標準差。對套利空間小的組合，可以在前面篩選組合的過程中進行優化，淘汰或改用其他模型配飾。

4. 最大連續虧損(Maximum Drawdown,MDD)

最大連續虧損(Maximum,Drawdown,MDD) 就是最大的一筆連續虧損。為什麼對一個交易模型而言 MDD 如此重要呢? 因為一個交易模型不可能永遠賺錢，MDD 則是讓我們了解到策略歷史最大的連續虧損，透過 MDD 來評量一個策略在目前市場表現的狀況，若出現連續虧損已經超過歷史的 MDD 則要考慮，這個策略在市場是否有效，或是因為事件所導致的正常狀況。同時的 MDD 也可以做為需要多少超額保證金的依據。

5. 風險報酬比率(NPMDD Ratio)

此比率概念為每承受損失 1%的風險，所產生的預期獲利報酬率(%)，交易策略的此項指標要能夠為正，並且數值越高表示績效越佳。分母採用最大可能虧損率(Maximum Drawdown, MDD)，是一個將交易策略對最大可能虧損狀況承受能力的衡量指標。最大可能虧損報酬率的計算方法如下:



DOI:10.6814/NCCU201900124

$$NPMDD\ Ratio = \frac{NetProfit(\%)}{MDD(\%)} \tag{20}$$

其中 NetProfit(%)為策略總報酬率，MDD(%)為最大可能虧損。

風險報酬比率的概念類似於夏普率，但它比較常出現在槓桿交易當中，它關心的是策略能不能承受歷史最壞的狀況。因為槓桿交易常常會產生爆倉的情況，所以若 MDD 太大導致爆倉，則回測的報酬率在高都沒有用。當策略的 MDD 越小，且總報酬率越高，通常表示策略越穩定。高頻的統計套利策略中 NPMDD 都會有偏高的傾向，原因與夏普值類似，因其持倉極短(平均為 3~10 分鐘)，且有進行對沖，不容易受到持倉過久導致的方向性損失。

6. 勝率(Win Rate)

勝率的計算方法如下：

$$Win\ Rate = \frac{win\ trades}{Total\ trades} \tag{21}$$

我們以勝率去衡量模型的準確度，勝率過低(40%以下)潛在風險是預測模型失準，過高(90%)潛在風險是資料有誤亦或過擬合(overfitting)*[4]，一般經過縝密方式建模的高頻統計套利模型勝率約會在 60%~80%之間。

[4] 過擬合(Overfitting)：是指在調適一個統計模型時，使用過多參數。對比於可取得的資料總量來說，一個荒謬的模型只要足夠複雜，是可以完美地適應資料。當可選擇的參數的自由度超過資料所包含資訊內容時，這會導致最後（調適後）模型使用任意的參數，這會減少或破壞模型一般化的能力更甚於適應資料。過適的可能性不只取決於參數個數和資料，也跟模型架構與資料的一致性有關。此外對比於資料中預期的雜訊或錯誤數量，跟模型錯誤的數量也有關。



DOI:10.6814/NCCU201900124

# 第四章　實證分析

## 第一節　　實證資料與研究期間

本研究所使用之資料來源為芝加哥期貨交易所 (CME)資料庫，資料包含兩檔外匯期貨、兩檔指數期貨、一檔原物料期貨 (截至 2018 年 12 月 31 日，公開交易之期貨)，每一分鐘之開盤價格及收盤價格。資料期間為 2016 年 1 月 3 日 18:00 到 2018 年 12 月 31 日 17:00(樣本總共約 115 萬分鐘)，由於我們採用的是共整合動態校準的方式，所以不使用樣本內外的分法，也不在對參數進行額外最佳化，並且所有交易皆屬於當沖類型，不做留倉交易，每次交易持倉時間約為 3~10 分鐘之間。

由於每檔資產的跳動點數不同，在實務上要在做一層 Dollar Neutral*[5]，但我們為了簡化研究，暫先不做此操作，僅以期貨點數計算報酬率。

1.　選擇一分鐘的原因

首先說出結論，我們認為單純以價格來建立模型，市場週期越短(頻率越高)越容易預測，原因有兩個。

第一，市場週期如果是非常短的情況，那我們基本上可以假定這些價格都出自同一個時間序列模型(或分配)，這個推測背後的邏輯是在這些價格產生的過程當中，並未受到外力干擾，也就是說價格期間在這段走勢不受除了歷史資料以外的影響，若用數學式我們可以表達成:

$$P_t = F(P_{t-1}, P_{t-2} \dots P_0) + Noice \qquad (22)$$

[5] Dollar Neutral: 每檔資產跳動點數不同，我們應先對商品價格與點數進行計算，讓配對的資產跳動金額一致。



DOI:10.6814/NCCU201900124

但是週期越短也會出現一個致命性的問題，那就是雜訊(Noice)更多，所以對市場週期短的預測模型而言，如何降低價格的雜訊是一件很重要的事。

反過來說，為甚麼市場週期長，相對來說難預測?這邊的預測指的是單純使用成交量與價格做預測模型，而不考慮基本面&政治情勢&宏觀經濟這些因子。因為市場週期如果拉長，很容易受到消息與政策面的影響，也就是單純用價格建立預測模型，無法預知突發狀況，當這種突發狀況發生，價格就不在只是受到歷史資料影響，包括了四面八方新資訊引力對價格的拉扯，當這種情況發生，引導價格的數學式就不單單只有歷史價格那麼簡單了，可我們的模型並沒有這種歷史資料告訴它該如何應對，最後會以失敗告終。但市場週期拉長的好處在於，價格的資訊含金量是遠遠高於短期價格的，換句話說，雜訊的成份比較少，對於消除雜訊的工作也就沒那麼複雜且困難，以上是第一個選擇一分鐘做研究的原因。

第二個原因是關於跳空/除權息/增減資相關的問題。我們從跳空進行舉例，大多數的期貨市場/股票市場，因為不是 24/7/365*[6]的關係，都會有跳空，我們沒辦法直接從歷史價格去定義跳空的原因與價格行為的關係，倘若我們採用每日價格作為研究資料，這段跳空的時間(通常是休市時間)發生的事情我們根本沒辦法掌握，這會讓我們的模型變的非常不穩定，況且跳空還有一個問題，價格資料會變成是間斷性的離散資料。除權息與增減資也是一樣的道理，我們很難去找到每一檔商品的每一次跳空/除權息/增減資，去對它進行處理與數學定義。以上兩個原因就是我們使用分鐘數據的理由。

那我們為甚麼不使用五分鐘或十分鐘甚至一小時?這也是許多人可能會提到的問題。我們先回想模型的要求，其中有一個要求是模型是不留倉的，在不留倉這個條件下，我們每個新模型都是從每一天的開市進行建構，也就是說我們有一

[6] 24/7/365:市場全年無休一年 365 天,一週 7 天,一天 24 小時運作,代表市場不會產生跳空。



DOI:10.6814/NCCU201900124

段時間是沒辦法做交易的(模型需要累積資料)，我們使用的共整合時間參數為 60 分鐘，這個數字不能太小也不能太大，太小的結果是若我們用 20 個一分鐘收盤價做出的模型，資料量太少，在統計上大數法則(Law of large numbers)*[7]不成立(至少要求 30 筆樣本以上,越多越好)，但太多的話我們開市建模時間太長，例如我們參數為 600，那代表開市前 600 分鐘都沒有交易訊號，一整天的交易日只有 23 小時，我們當然希望越快能交易越好。同理，如果我們用一小時價格會發生甚麼事?我們一天的資料量只有 23 筆，完全不足以建構模型，五分鐘的話也只有 23x12=274 筆，所以這就是我們不使用五分鐘以上價格的原因，至於其它時間;如:二分鐘、三分鐘...則可以做嘗試。最完美的情況是我們直接從委託簿(OrderBook)*[8]自行模擬成交狀況產出逐筆資料(Tick)進行建模，但那資料量過於龐大且複雜，我們為了簡化研究(但不過度簡化)，選擇了一分鐘價格。

2. 歐元/美元外匯期貨

表 1:歐元/美元外匯期貨規格表

| | |
|---|---|
| 合約規模 | 12,500 歐元<br>如果美元/歐元= 1.3000，則合約= 16,250 美元(=12,500 歐元 x 1.3000 美元/歐元)。 |
| 交易時間 | CME Globex：<br>週日：下午 5:00-次日下午 4:00(中部時間)。<br>週一至週五：下午 5:00-次日下午 4:00(中部時間)，但週五於下午 4:00 休市，並於週日下午 5:00 重開。 |

[7] 大數法則(Law of large numbers):在數學與統計學中，大數法則又稱大數定律、大數律，是描述相當多次數重複實驗的結果的定律。根據這個定律知道，樣本數量越多，則其算術平均值就有越高的機率接近期望值。

[8] 委託簿(Orderbook): 訂單簿是交易所（期貨/證券交易所）用於記錄特定金融資產中買方和賣方的訂單（手動或電子）列表。 匹配引擎使用該簿確定可以履行哪些訂單，即可以進行哪些交易。



DOI:10.6814/NCCU201900124

| | |
|---|---|
| 最小變動價位 | 0.0001 美元/歐元（=1.25 美元）。 |
| 產品簡稱 | EUR。 |
| 上市合約 | 以 3 月季度週期的 2 個月（3 月、6 月、9 月、12 月）。 |
| 結算方法 | 可交割。 |
| 交易終止 | 合約月份第三個週三之前的第二個營業日（通常是週一）中部時間上午 9:16。 |

3. 英鎊/美元外匯期貨

表 2:英鎊/美元外匯期貨規格表

| | |
|---|---|
| 合約規模 | 62,500 英鎊。 |
| 交易時間 | CME Globex：<br>週日：下午 5:00-次日下午 4:00(中部時間)。<br>週一至週五：下午 5:00-次日下午 4:00(中部時間)，但週五於下午 4:00 休市，並於週日下午 5:00 重開。 |
| 最小變動價位 | 完整交易：0.0001 美元/英鎊增幅（6.25 美元）。<br>連續月份價差（僅 Globex）： 0.00001 美元/英鎊增幅（0.625 美元）。<br>所有其他價差合併： 0.0001 美元/英鎊增幅(6.25 美元)。 |
| 產品簡稱 | GBP。 |
| 上市合約 | 前 3 個連續月份上市合約和以 3 月季度周期的 20 个月（3 月、6 月、9 月、12 月）。 |



DOI:10.6814/NCCU201900124

| | |
|---|---|
| 結算方法 | 可交割。 |
| 交易終止 | 合約月份第三個週三之前的第二個營業日（通常是週一）中部時間上午 9:16。 |

4. E-mini S&P 500 指數期貨

表 3：E-miniS&P500 指數期貨規格表

| | |
|---|---|
| 合約規模 | 50 美元 x 標準普爾 500 指數。 |
| 交易時間 | CME Globex：<br>週日：下午 5:00-次日下午 4:00(中部時間)。<br>週一至週五：下午 5:00-次日下午 4:00(中部時間)，但週五於下午 4:00 休市，並於週日下午 5:00 重開。<br>下午 3:15-3:30 之間交易暫停。 |
| 最小變動價位 | 完整交易：0.25 指數點=12.50 美元 跨期價差：0.05 指數點=2.50 |
| 產品簡稱 | SPX。 |
| 上市合約 | 以 3 月季度週期的 5 個月（3 月、6 月、9 月、12 月）。 |
| 結算方法 | 財務結算。 |
| 交易終止 | 交易最遲可以在該合約月第 3 個周五上午中部時間 8：30 執行 對於指數收盤基差交易（BTIC），交易在該合約月第三個周五之前的周四中部時間下午 3 時終止。 |



DOI:10.6814/NCCU201900124

5. E-mini 那斯達克 100 指數期貨

表 4:E-mini 那斯達克 100 指數期貨規格表

| | |
|---|---|
| 合約規模 | 20 美元 x 那斯達克 100 指數。 |
| 交易時間 | CME Globex：<br>週日：下午 5:00-次日下午 4:00(中部時間)。<br>週一至週五：下午 5:00-次日下午 4:00(中部時間)，但週五於下午 4:00 休市，並於週日下午 5:00 重開。<br>下午 3:15-3:30 之間交易暫停。 |
| 最小變動價位 | 完整交易：0.25 指數點=5.00 美元。<br>跨期價差：0.05 指數點=1.00 美元。<br>BTIC 最小變動價位：0.05 指數點=1.00 美元。 |
| 產品簡稱 | NSX。 |
| 上市合約 | 以 3 月季度週期的 5 個月（3 月、6 月、9 月、12 月）。 |
| 結算方法 | 財務結算。 |
| 交易終止 | 交易最遲可能在合約月第 3 個週五上午 8:30 中部時間終止。 |

6. WTI 輕質原油期貨

表 5:WTI 輕質原油期貨規格表

| | |
|---|---|
| 合約規模 | 1,000 桶。 |
| 報價單位 | 每桶美元和美分。 |



DOI:10.6814/NCCU201900124

| | |
|---|---|
| 交易時間 | CME Globex：<br>週日：下午 5:00-次日下午 4:00(中部時間)。<br>週一至週五：下午 5:00-次日下午 4:00(中部時間)，但週五於下午 4:00 休市，並於週日下午 5:00 重開。 |
| 最小價格波幅 | 每桶 0.01 美元。 |
| 產品簡稱 | WTI。 |
| 上市合約 | 當前年度和未來 8 個日曆年以及 2 個額外連續合約月份上市的月度合約。在當前年度 12 月份合約交易終止後，將加入新日曆年及 2 個額外連續合約月份上市的月度合約。 |
| 結算方法 | 可交割。 |
| 交易終止 | 當前交割月份的交易須在交割月前一個月的 25 日之前的第 3 個營業日終止。若該月 25 日為非營業日，則交易須在 25 日前一個營業日之前的第 3 個營業日終止。若交易所法定假日時程表在原油期貨上市後發生變更，原上市到期日應仍然有效。 若原上市到期日被宣佈為公共假日，則到期日將為前一個營業日。 |

## 第二節　布林通道策略之統計套利實證結果

### 一、　布林通道策略(假設殘差$e_t$為獨立同分配)

我們假設殘差$e_t$為獨立同分配利用布林通道建構統計套利策略進行回測，而策略績效的計算及回測我們利用全樣本期 2016 年 1 月 3 日 18:00 到 2018 年 12 月 31 日 17:00(樣本總共約 115 萬分鐘)測試，計算過程皆使用報酬率當計算依據，即每一塊美金的資金在策略每次的進出場後可以取得的獲利，本研究計算的是樣本期間之總報酬率，以利於我們去計算績效的最大策略虧損報酬率，如表 6:



DOI:10.6814/NCCU201900124

表 6:布林通道策略績效表

| | 總報酬 | 夏普率 | 每筆平均報酬率 | MDD | 交易次數(多/空) | 風報比 | 勝率 | 單日最高獲利紀錄 |
|---|---|---|---|---|---|---|---|---|
| EURxGBP | 32.51% | 8.47 | 0.0072155% | -0.42% | 4554(2313/2241) | 77.98 | 64.85% | 1.96%(2016/06/23 ) |
| EURxNSX | 11.81% | 3.65 | 0.0057880% | -0.47% | 2060(1021/1039) | 25.21 | 57.96% | 1.12%(2016/06/23) |
| EURxSPX | 13.35% | 4.09 | 0.0065134% | -0.6% | 2082(1007/1075) | 22.15 | 58.71% | 1.62%(2016/06/23 ) |
| EURxWTI | 17.76% | 5.23 | 0.0075833% | -0.73% | 2365(1163/1202) | 24.28 | 59.56% | 0.91%(2016/06/23 ) |
| GBPxNSX | 15.74% | 2.99 | 0.0067852% | -1.87% | 2347(1168/1179) | 8.45 | 56.64% | 1.85%(2016/06/23 ) |
| GBPxSPX | 16.89% | 3.96 | 0.0069296% | -1.03% | 2468(1197/1271) | 16.38 | 57.49% | 0.61%(2016/06/23 ) |
| GBPxWTI | 14.80% | 2.94 | 0.0058948% | -1.54% | 2533(1277/1256) | 9.63 | 58.46% | 1.21%(2016/06/23 ) |
| NSXxSPX | 57.81% | 8.23 | 0.0070287% | -1.03% | 8431(4158/4273) | 55.94 | 70.21% | 1.99%(2018/02/05) |
| NSXxWTI | 18.12% | 2.61 | 0.0068500% | -1.54% | 2721(1342/1379) | 11.72 | 56.54% | 1.56%(2018/10/23) |
| SPXxWTI | 19.59% | 3.46 | 0.0072158% | -2.15% | 2951(1471/1480) | 9.09 | 58.45% | 1.41%(2018/02/06) |

首先我們可以從總報酬率(Total Return)開始看起，可以發現策略的報酬率都不算高，3 年總報酬率大約落在 15%~20%，年化後的報酬率落在 5%~7%之間，除了兩個配對組合特別高，分別是歐元配英鎊(EURxGBP)與小那指配小 S&P(NSXxSPX)，我們推測是這兩個配對本身的商品本質較類似，從交易次數就能窺知一二，確實這兩個商品的交易次數分別佔據榜中前兩名，而在夏普率(Sharpe Ratio)與風險報酬比率(NPMDD Ratio)方面也是這兩個商品最高，小那指配小 S&P(NSXxSPX)勝率更是來到 70.21%，意味這兩個商品更容易形成共整合關係，且關係不容易崩解。但在每筆平均報酬方面則是歐元配原油最高(GBPxWTI)，平均每筆獲利約 0.0076%，整體來說我們使用一分鐘的資料進行統計套利，交易模型偏向高頻交易，高頻交易的特徵也不意外的呈現在績效報表上，我們策略的持倉時間落在 3~10 分鐘之間，很少超過 10 分鐘的持倉時間，這點也直接反應在 MDD 上，最大 MDD 為小 S&P 配原油(SPXxWTI)的-2.15%，以交易策略來說並不算高。至於最後一欄單日最高獲利，我們發現最高的為小那指配小 S&P(NSXxSPX)，接近 2%左右，發生在 2018/02/05，這天美國道瓊股市暴跌一千多點。接著把所有的配對以一樣的初始資金進行投資組合建構，如圖 9 與表 7



DOI:10.6814/NCCU201900124

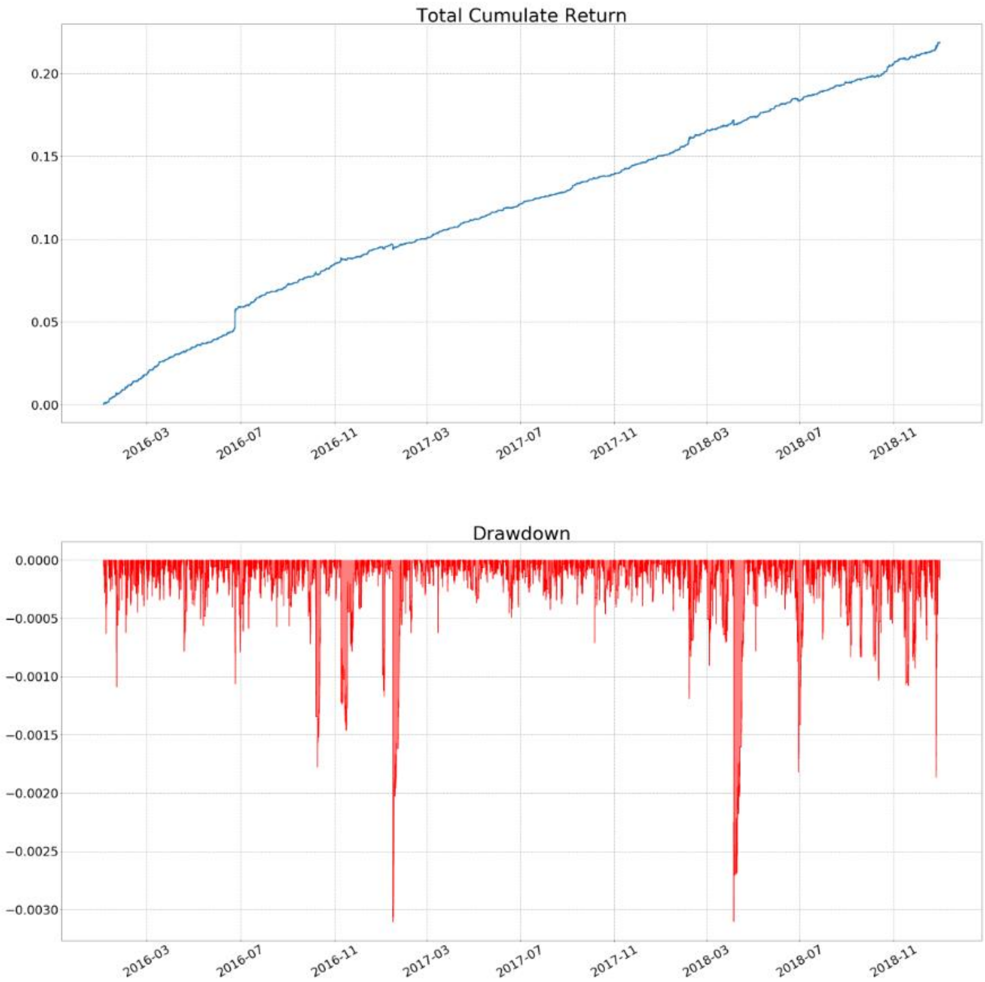



圖 9:布林通道策略投資組合績效圖

表 7:布林通道策略投資組合策略績效表

| | 單日獲利記錄<br>前 10 名 |
|---|---|
| | 1.082039%(2016/06/23) |
| Total Return：21.87% | 0.230716%(2016/11/08) |
| Sharpe Ratio：6.93 | 0.188500%(2018/10/23) |
| | 0.187582%(2018/02/05) |
| MDD：-0.31% | 0.153510%(2016/01/20) |
| | 0.150396%(2018/12/28) |
| NPMDD Ratio：70.35 | 0.139612%(2016/03/16) |
| | 0.136753%(2016/02/01) |
| | 0.126537%(2016/03/02) |
| | 0.122301%(2017/09/07) |



DOI:10.6814/NCCU201900124

報酬曲線相當的漂亮，看來假設殘差$e_t$為獨立同分配似乎是不錯的選擇，整體報雖然酬不高(21.87%)，年化大約 7%左右，但這本來就是市場中性策略的特性，它最大的優點是非常的穩定，MDD 只有-0.31%，很適合當成機構投資人進行需要很穩定報酬的基金策略使用(如：養老基金)，年化夏普 6.93 也是非常漂亮的數字，單日獲利紀錄則發生在 2016/06/23 的 1.08%，並且很多配對的最高獲利日都在這天，這天是個特殊的日子，我們後面會對這進行解釋。

## 二、 布林通道策略加入中位數反轉定律

接著我們把布林通道策略直接加入中位數反轉定律的條件試試，樣本期間與計算方法皆不變，如表 8:

表 8:布林通道+中位數反轉定律策略績效表

| | 總報酬 | 夏普率 | 每筆平均報酬率 | MDD | 交易次數(多/空) | 風報比 | 勝率 | 單日最高獲利紀錄 |
|---|---|---|---|---|---|---|---|---|
| EURxGBP | 35.08% | 9.26 | 0.0044155% | -0.44% | 8019(4072/3947) | 79.16 | 65.68% | 1.29%(2016/06/23) |
| EURxNSX | 15.75% | 4.97 | 0.0042882% | -0.64% | 3705(1838/1867) | 24.5 | 60.48% | 1.27%(2016/06/23) |
| EURxSPX | 16.23% | 5.24 | 0.0042025% | -0.39% | 3908(1994/1914) | 42.06 | 60.36% | 0.51%(2016/06/23) |
| EURxWTI | 20.11% | 6.0 | 0.00477% | -0.58% | 4257(2060/2197) | 34.93 | 62.02% | 0.58%(2016/06/23) |
| GBPxNSX | 22.16% | 4.82 | 0.0052058% | -0.80% | 4292(2161/2131) | 27.66 | 60.70% | 1.03%(2016/06/23) |
| GBPxSPX | 26.33% | 5.42 | 0.0060225% | -1.27% | 4410(2211/2199) | 20.68 | 61.92% | 0.65%(2016/10/07) |
| GBPxWTI | 22.39% | 5.51 | 0.0050820% | -0.87% | 4439(2251/2188) | 25.74 | 61.30% | 0.56%(2018/02/08) |
| NSXxSPX | 68.75% | 10.55 | 0.0051947% | -0.92% | 13525(6758/6767) | 74.55 | 74.48% | 0.82%(2016/06/23) |
| NSXxWTI | 31.50% | 4.54 | 0.0072137% | -1.34% | 4486(2232/2254) | 23.55 | 60.73% | 1.31%(2018/12/27) |
| SPXxWTI | 28.25% | 5.33 | 0.0063011% | -1.31% | 4924(2456/2468) | 21.52 | 62.53% | 1.26%(2016/11/08) |

首先從勝率開始看，每一個組合都有 60%以上的勝率，雖然不及 Andrew(2009)中位數定律的 75%，但對於原來的策略來說平均提高 5%~10%，尤其是小那指配小 S&P(NSXxSPX)，勝率達到 74.48%，相當接近 75%了，再來我們看到總報酬率，平均來說也是有所上升，大多圍繞在三年 20%上下，策略穩定



DOI:10.6814/NCCU201900124

度方面，年化夏普值與風報比也都有抬升，小那指配小 S&P(NSXxSPX)的年化夏普值更是來到 10.55，以交易策略來說算是非常高了，單日最高獲利方面出現在小那指配原油(NSXxWTI)，發生在 2018/12/27 的 1.31%，這個日子剛好是美國股市歷史上最大漲幅，道瓊工業指數大漲一千多點。接著把所有的配對以一樣的初始資金進行投資組合建構，如圖 10 與表 9:

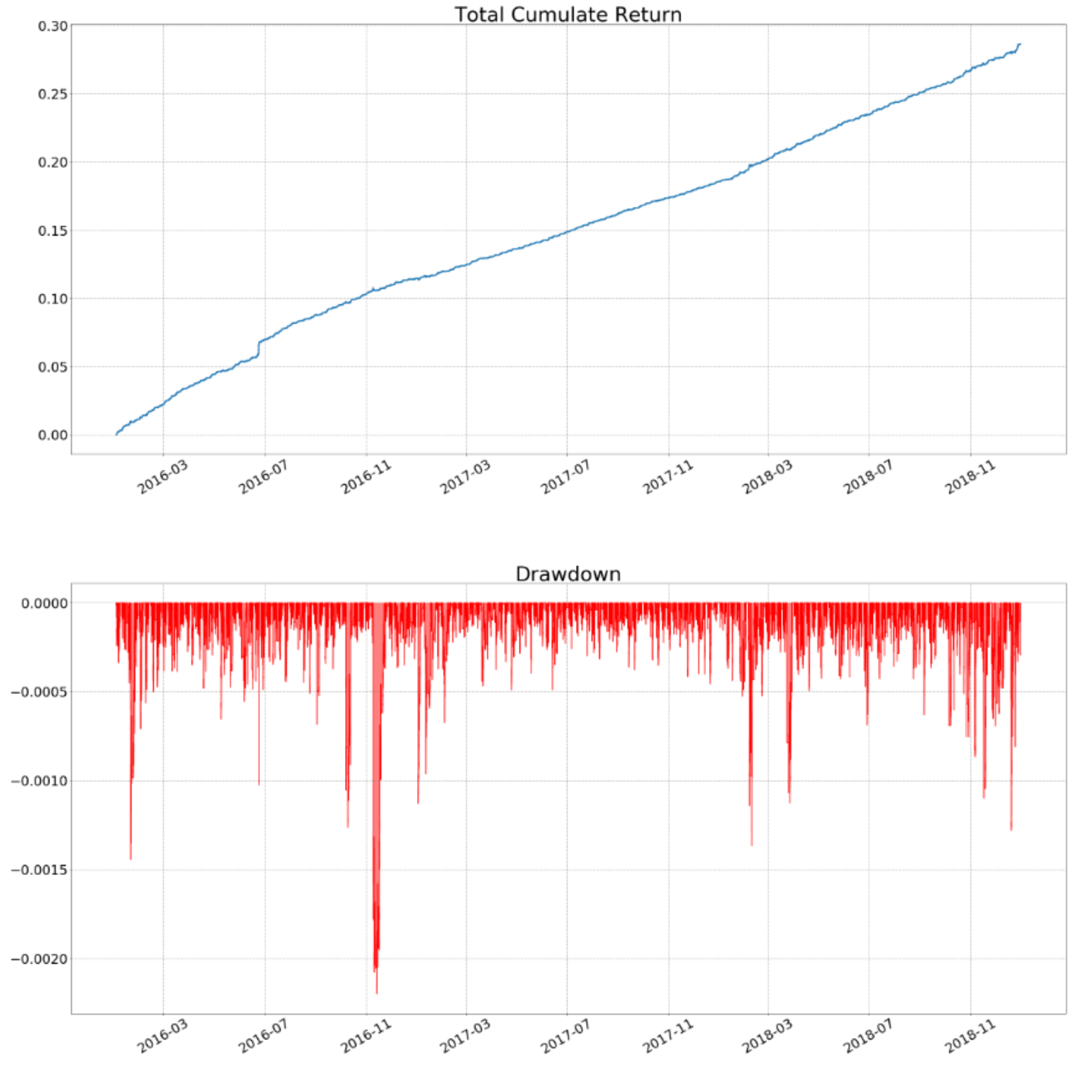



圖 10:布林通道策略+中位數反轉定律投資組合績效圖



DOI:10.6814/NCCU201900124

表 9:布林通道+中位數反轉定律投資組合策略績效表

| | 單日獲利記錄<br>前 10 名 |
|---|---|
| Total Return：28.64%<br>Sharpe Ratio：10.51<br>MDD：-0.22%<br>NPMDD Ratio：130.27 | 0.717027%(2016/06/23)<br>0.244528%(2016/06/24)<br>0.237241%(2016/11/08)<br>0.235510%(2018/12/27)<br>0.146842%(2018/10/25)<br>0.139062%(2016/02/18)<br>0.137423%(2016/01/20)<br>0.137015%(2018/02/05)<br>0.133338%(2018/02/06)<br>0.133293%(2016/02/01) |

加入中位數定律後，整體績效都是向上提升，風報比到 130.27，在一次用結果反推，看來假設殘差$e_t$為獨立同分配似乎非常合理，就連加入機率論的方法都很符合結果。而單日最高獲利紀錄同樣是 2016/06/23，但不同於原策略的 1.08%，是 0.717%，我們還沒辦法解釋關於這個下降的原因。

總結這一小節，看起來我們利用 EMA 平滑資料，並利用共整合動態校準架構調整模型，最後假設殘差$e_t$為獨立同分配是很符合市場情況的，雖然我們尚未克服實際交易中的摩擦成本，但若單純以研究資產價格行為之面向來看，我想以結果反推我們的假設，算是合理的。

## 第三節　OU 過程之統計套利實證結果

### 一、　OU 過程策略(假設殘差$e_t$具有序列相依性)

我們利用 OU 過程建構統計套利策略進行回測，而策略績效的計算及回測我們利用全樣本期 2016 年 1 月 3 日 18:00 到 2018 年 12 月 31 日 17:00(樣本總共約 115 萬分鐘)測試，計算過程皆使用報酬率當計算依據，即每一塊美金的資



DOI:10.6814/NCCU201900124

金在策略每次的進出場後可以取得的獲利，本研究計算的是樣本期間之總報酬率，以利於我們去計算評估績效的最大策略虧損報酬率，如表 10:

表 10:OU 過程策略績效表

| | 總報酬 | 夏普率 | 每筆平均報酬率 | MDD | 交易次數(多/空) | 風報比 | 勝率 | 單日最高獲利紀錄 |
|---|---|---|---|---|---|---|---|---|
| EURxGBP | 23.41% | 3.71 | 0.0019188% | -1.13% | 12334(6260/6074) | 20.61 | 50.62% | 1.67%(2016/06/23 ) |
| EURxNSX | 7.28% | 1.37 | 0.0010047% | -2.11% | 7338(3610/3728) | 3.45 | 48.63% | 0.68%(2016/06/26) |
| EURxSPX | 6.89% | 1.33 | 0.000945% | -2.31% | 7396(3615/3781) | 2.98 | 48.79% | 1.12%(2016/06/23 ) |
| EURxWTI | 11.99% | 1.91 | 0.0014606% | -1.41% | 8319(4199/4120) | 8.51 | 48.35% | 1.35%(2016/06/23 ) |
| GBPxNSX | 8.27% | 0.88 | 0.0010114% | -6.40% | 8286(4072/4214) | 1.29 | 48.12% | 4.16%(2016/06/23 ) |
| GBPxSPX | 13.76% | 1.74 | 0.0016917% | -2.32% | 8241(4044/4197) | 5.92 | 49.37% | 2.6%(2016/06/23 ) |
| GBPxWTI | 12.88% | 1.59 | 0.0015085% | -1.97% | 8642(4421/4221) | 6.53 | 49.56% | 2.26%(2016/06/23 ) |
| NSXxSPX | 40.07% | 3.74 | 0.0027530% | -2.98% | 16139(8560/7579) | 13.45 | 55.91% | 1.37%(2018/03/29) |
| NSXxWTI | 14.18% | 1.17 | 0.0015616% | -5.14% | 10347(5615/4732) | 2.76 | 49.81% | 1.58%(2018/10/24) |
| SPXxWTI | 14.45% | 1.81 | 0.0020348% | -4.49% | 10536(5596/4940) | 3.91 | 49.64% | 1.87%(2018/02/06) |

首先一樣從總報酬率開始看起，看起來都不是很高，最低的歐元配小 S&P(EURxSPX)三年報酬率 6.89%，換算年化約 2.2%左右，最高的組合一樣是小納指配小 S&P(NSXxSPX)的 40.07%，換算年化約 12.2%左右，但依舊比不上前面布林通道策略的報酬率，從夏普值與風險報酬比率也看得出不如前面來的優秀，最大可能虧損(MDD)方面，英鎊配小那指(GBPxNSX)來到了-6.40%，而這個組合的報酬也不高，8.27%，MDD 偏大與報酬較低導致夏普值上不去，交易次數的話普遍偏多，我們在加入勝率一起看，勝率大部份都沒有破 50%，對於這種偏高頻的統計套利模型而言是不及格的，最後看到單日最高獲利紀錄，最高發生在英鎊配小那指(GBPxNSX)，發生在 2016/06/23 的 4.16%，是三年總報酬率約一半，我們認為這種過高的單日報酬率已經不單單只有市場中立性的報酬了，或多或少參雜了不少的方向性獲利在裡面。接著我們把所有的配對以一樣的初始資金進行投資組合建構，變成一個完整的統計套利策略，如圖 11 與表 11:



DOI:10.6814/NCCU201900124

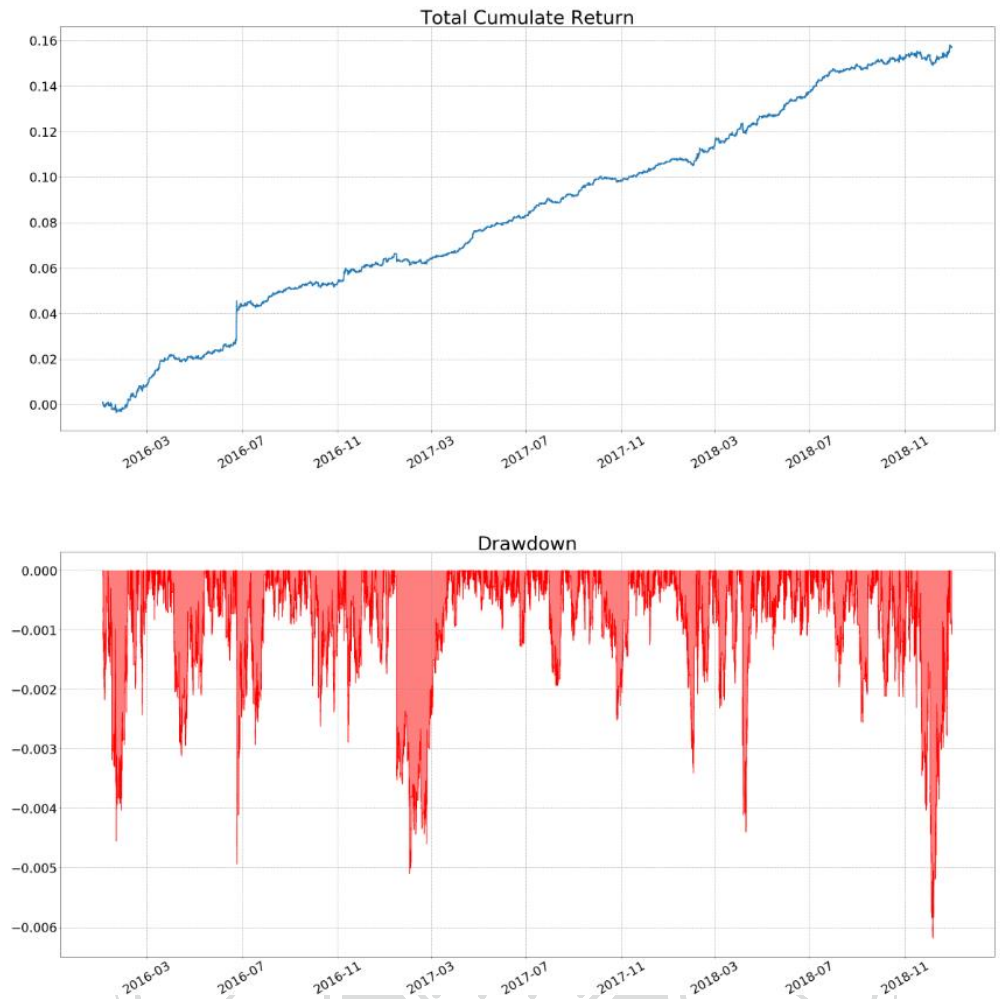



圖 11:OU 過程策略投資組合績效圖

表 11:OU 過程投資組合策略績效表

| | 單日獲利記錄<br>前 10 名 |
|---|---|
| Total Return：15.71%<br>Sharpe Ratio：3.29<br>MDD：-0.62%<br>NPMDD Ratio：25.4 | 1.510234%(2016/06/23)<br>0.287145%(2018/10/24)<br>0.280058%(2016/11/08)<br>0.255346%(2018/03/01)<br>0.216835%(2018/04/24)<br>0.193376%(2016/02/24)<br>0.192891%(2017/04/23)<br>0.184459%(2018/12/18)<br>0.180553%(2016/06/07)<br>0.179607%(2018/09/13) |





DOI:10.6814/NCCU201900124

報酬曲線看起來還是很不錯，呈現 45 度向上，三年總報酬 15.71%，相比之下稍微低了一點，年化夏普率 3.29 還算可以接受，總體而言較布林通道策略差，但優勢在策略暴發力比較高，單日最高獲利來到 1.51%，同樣發生在 2016/06/23。相比於布林通道假設殘差$e_t$為獨立同分配，OU 過程假設殘差$e_t$具有序列相依性似乎從結果上沒那麼貼近實際。

## 二、 OU 過程策略加入中位數反轉定律

接著我們把 OU 過程策略直接加入中位數反轉定律的條件試試，樣本期間與計算方法皆不變，如表 12:

表 12:OU 過程+中位數定律策略績效表

| | 總報酬 | 夏普率 | 每筆平均報酬率 | MDD | 交易次數(多/空) | 風報比 | 勝率 | 單日最高獲利紀錄 |
|---|---|---|---|---|---|---|---|---|
| EURxGBP | 23.39% | 6.63 | 0.0041732% | -0.50% | 5655(2811/2844) | 46.6 | 63.14% | 0.99%(2016/06/23) |
| EURxNSX | 6.53% | 2.19 | 0.0026519% | -0.97% | 2491(1239/1252) | 6.75 | 58.1% | 0.32%(2016/06/23) |
| EURxSPX | 12.08% | 3.67 | 0.0043894% | -0.69% | 2784(1387/1397) | 17.54 | 59.86% | 0.83%(2016/06/26) |
| EURxWTI | 9.30% | 3.04 | 0.0032533% | -0.76% | 2890(1502/1388) | 12.24 | 59.25% | 0.34%(2016/07/20) |
| GBPxNSX | 10.59% | 2.97 | 0.0037108% | -0.79% | 2881(1419/1462) | 13.41 | 58.60% | 0.46%(2018/04/27) |
| GBPxSPX | 14.48% | 3.21 | 0.0046942% | -0.84% | 3116(1534/1582) | 17.3 | 59.59% | 1.19%(2016/06/23) |
| GBPxWTI | 16.58% | 4.38 | 0.0052438% | -0.53% | 3192(1607/1585) | 31.34 | 60.63% | 1.20%(2016/06/24) |
| NSXxSPX | 50.92% | 8.61 | 0.0053677% | -0.92% | 9862(5141/4721) | 55.0 | 68.92% | 1.34%(2016/06/23) |
| NSXxWTI | 16.14% | 2.63 | 0.0054418% | -1.40% | 3102(1577/1525) | 11.52 | 58.73% | 1.47%(2016/11/08) |
| SPXxWTI | 17.39% | 3.53 | 0.0057935% | -1.63% | 3335(1669/1666) | 10.68 | 58.79% | 1.63%(2016/11/08) |

從績效表中可以看出中位數定律對於 OU 過程策略也有正向幫助，尤其是在策略穩定度方面，年化夏普值與風報比都比原本沒加入之前高出接近 2 倍，算是非常大的成長，這點也應證中位數定律在序列相依性的假設下同樣試用，值得注意的是最大的每筆平均報酬率(Per Avg Return)這個指標，在每一種模型下結果都不同，我們也還沒辦法推測出它的原因，這也是之後的研究目標之一，



DOI:10.6814/NCCU201900124

而在單日最高獲利紀錄上產生在小 S&P 配原油(SPXxWTI)，發生在 2016/11/08 的 1.63%，這一天也是非常特殊的日子，我們下一小節會一一進行分析與解釋。

接著我們把所有的配對以一樣的初始資金進行投資組合建構，變成一個完整的統計套利策略，如圖 12 與表 13:

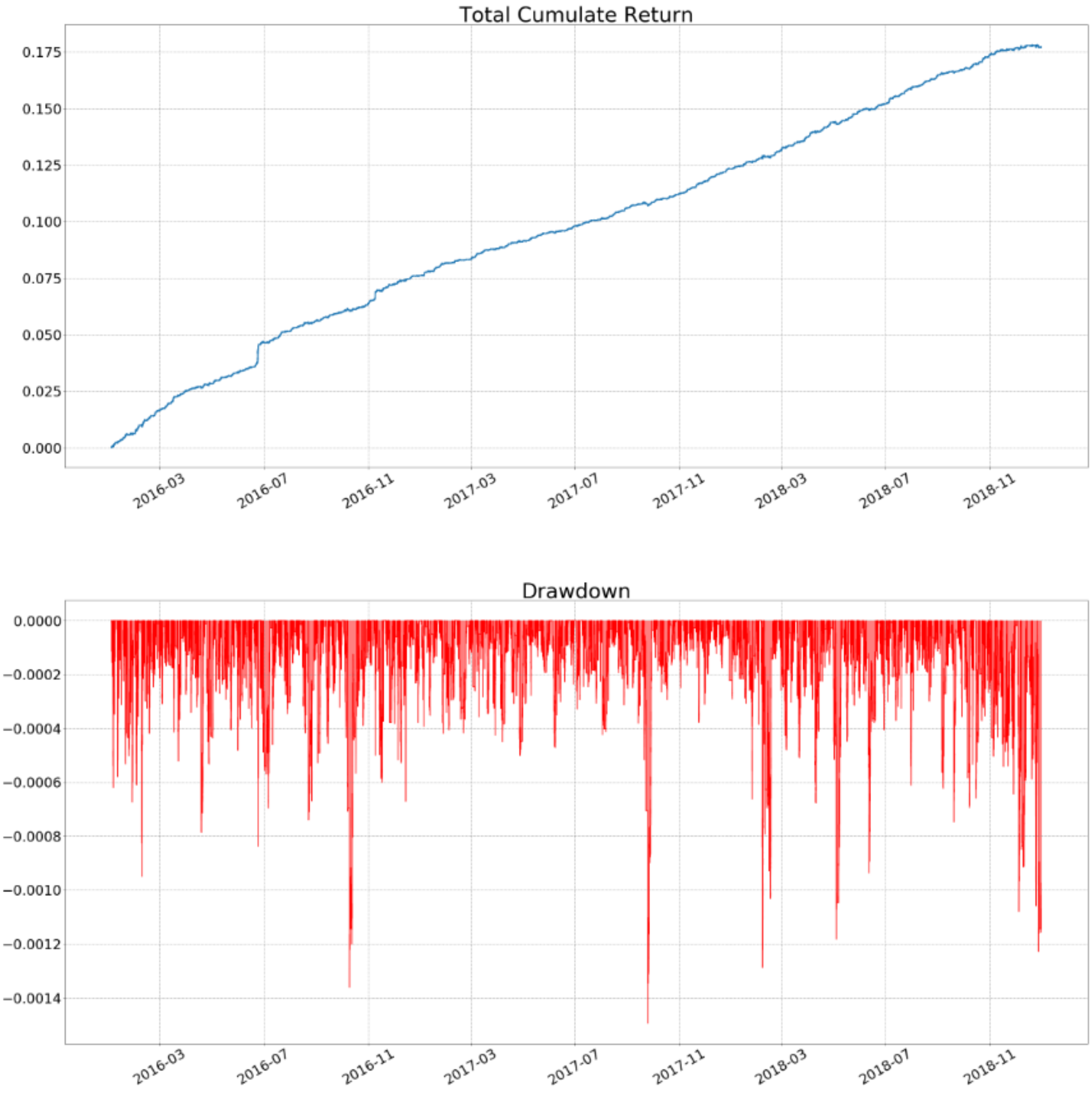



圖 12:OU 過程策略+中位數反轉定律投資組合績效圖



DOI:10.6814/NCCU201900124

表 13:OU 過程+中位數定律投資組合策略績效表

| | 單日獲利記錄<br>前 10 名 |
|---|---|
| Total Return：17.72%<br>Sharpe Ratio：7.79<br>MDD：-0.15%<br>NPMDD Ratio：118.57 | 0.567816%(2016/06/23)<br>0.349596%(2016/11/08)<br>0.267304%(2016/06/24)<br>0.145976%(2016/02/01)<br>0.134224%(2016/03/16)<br>0.129186%(2017/01/05)<br>0.126260%(2016/02/24)<br>0.116604%(2018/02/07)<br>0.112332%(2018/07/05)<br>0.108189%(2018/04/03) |

報酬曲線明顯比沒用中位數定律好很多，整體來說策略變得很穩定，從夏普值 3.29 變成 7.79 就能看出，報酬率小幅度提高，單日獲利最高紀錄又出現了熟悉的日子 2016/06/23 的 0.57%，第二高 2016/11/08 的 0.35%與第三高 2016/06/24 的 0.27%，似乎也在前面也都有看到。

總結這一小節，一開始我們假設殘差$e_t$具有序列相依性使用 OU 過程建立信賴區間，相比之下效果似乎不如布林通道來的好，但加入中位數定律後卻穩定了不少，我們推測是殘差$e_t$偏向獨立同分配的可能性更高一些，還有一點，單日最高獲利紀錄頻繁出現某些日子，這使得我們又在對這進行相關研究。

## 第四節　黑天鵝事件驅動研究結果

我們從策略績效中發現許多配對組合的最大獲利日是 2016/06/23，正好是一個很重要的黑天鵝事件，英國脫歐公投。我們從個別配對組合歐元配英鎊(EURxGBP)來嘗試研究這天到底發生甚麼事，如圖 13:



DOI:10.6814/NCCU201900124

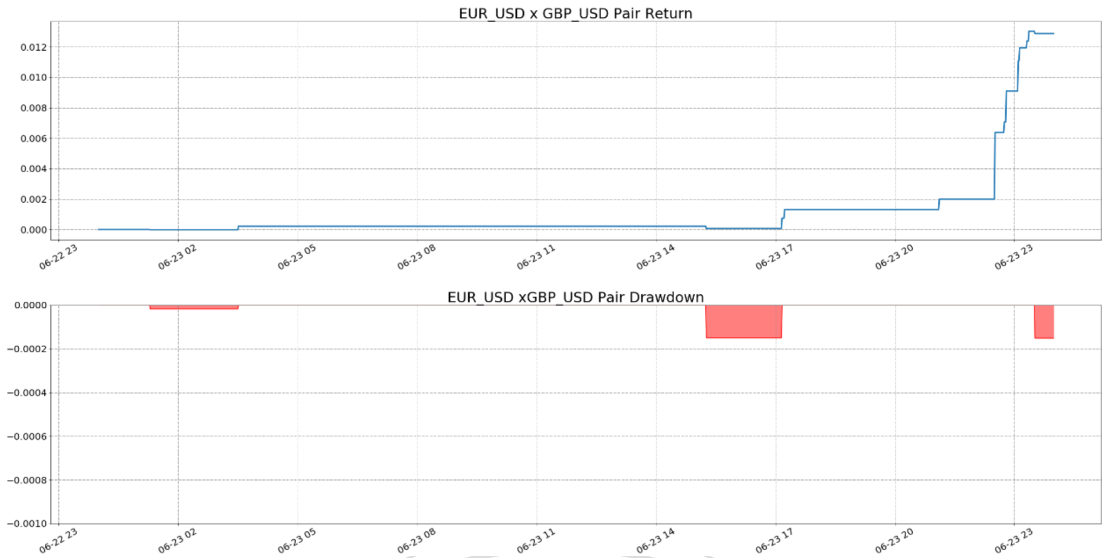



圖 13:英國脫歐公投-單日績效圖(EURxGBP)

我們可以看到權益曲線幾乎沒有回落的跡象，且單日報酬率高達 1.2%，接著我們看歐元與英鎊這天的走勢,如圖 14:

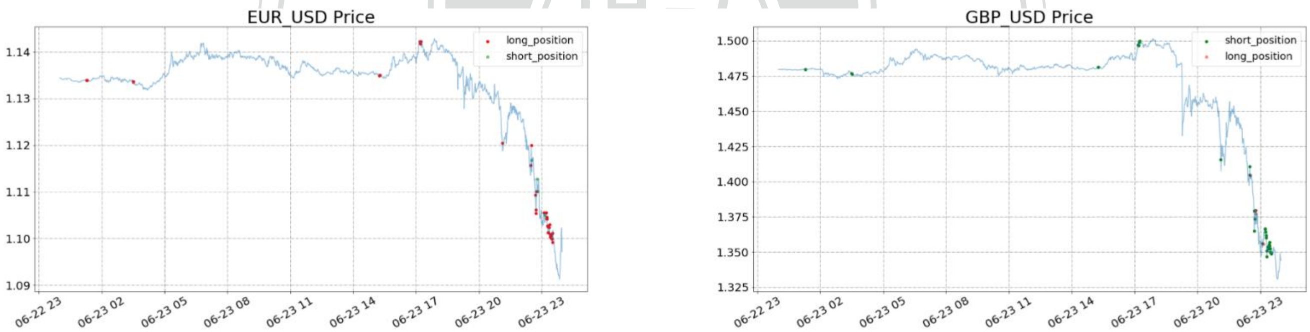
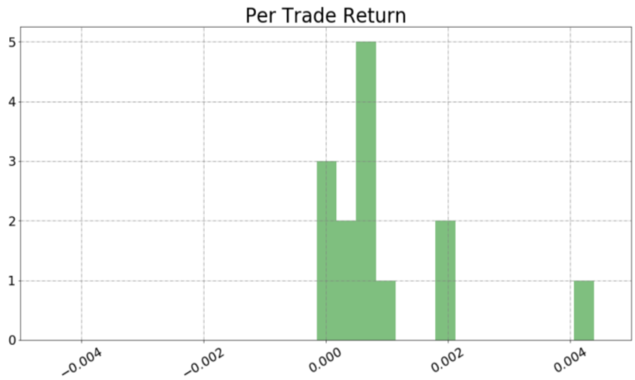



圖 14:英國脫歐公投-單日走勢與交易訊號圖(EURxGBP)

紅色的點代表做多，綠色的點代表做空，很明顯可以看到歐元與英鎊在這天走勢非常類似(拉長樣本期間走勢是完全不同的) 。兩資產同時受到了英國脫歐的影響，在走勢大幅下跌時模型判定有共整合關係，反而是兩資產波動較小時(2016/06/23 00:00~2016/06/23 15:00)沒什麼交易，似乎跟一般認知的均值回歸策略有所差異，接著我們看這一天的每筆交易報酬率分佈,如圖 15:



DOI:10.6814/NCCU201900124




| | |
|---|---|
| 單日總報酬率 | 1.28% |
| 勝率 | 78.5% |
| 交易次數 | 14 |
| 平均每筆交易報酬 | 0.09% |

圖 15:英國脫歐公投-單日報酬率分佈圖(EURxGBP)

橫軸代表交易報酬率，縱軸為次數，我們可以發現整個分佈為右偏而且厚尾，我們搭配勝率一起看，發現勝率高達 78.5%，比全樣本期間的 64.85%還高將近 15%左右，而平均每筆交易報酬 0.09%，更是比原本的 0.004%高了二十二倍。

接下來我們看第二個例子，2016/11/08，很巧的是今天正好也是黑天鵝事件，美國總統川普當選。我們選取小 S&P 配原油(SPXxWTI)來嘗試研究這一天所造成的影響，首先看到圖 16:

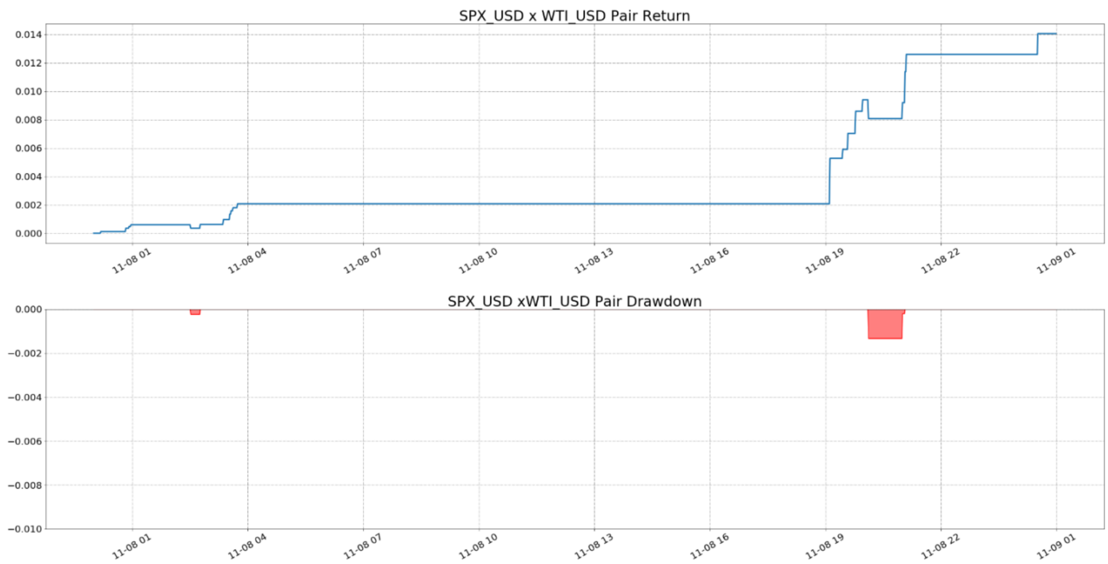



圖 16:美國總統大選-單日績效圖(SPXxWTI)



DOI:10.6814/NCCU201900124

獲利曲線基本上指有兩段非常微小的虧損，虧損持續時間都不超過一小時，獲利不斷創新高，單日總報酬率高達 1.4%，接著我們看圖 17:

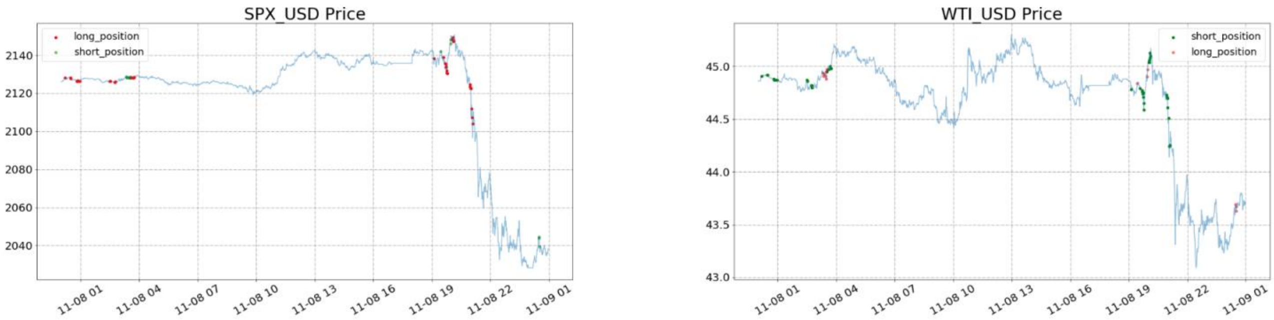
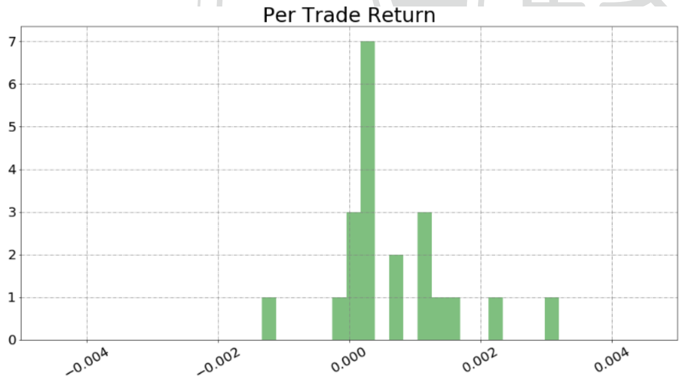



圖 17:美國總統大選-單日走勢與交易訊號圖(SPXxWTI)

紅色的點代表做多，綠色的點代表做空，我們發現在這天，小 S&P 走勢與 WTI 走勢在前段不相同，但在後段下跌時非常類似，並且這段期間也是模型通過共整合關係頻繁觸發交易的時間點，接著我們看這天每筆交易報酬率分佈,如圖 18:




| | |
|---|---|
| 單日總報酬率 | 1.41% |
| 勝率 | **90.4%** |
| 交易次數 | 21 |
| 平均每筆交易報酬 | 0.067% |

圖 18:美國總統大選-單日報酬率分佈圖(SPXxWTI)

我們可以看到勝率高達 90.4%，比全樣本的 62.53%高出將近 30%，代表這兩個資產在這一天的共整合關係非常明確，也就是殘差$e_t$完全按照我們想要的方向走，總共交易 21 次，只預測失敗 2 次，並且幅度都很小，整個報酬率分佈跟歐元配英鎊(EURxGBP)在英國脫歐那天相同，右偏厚尾，而平均每筆交易報酬 0.067%，也比原本的 0.0063%高出約十倍。



DOI:10.6814/NCCU201900124

最後取 OU 過程加入中位數定律後的投資組合最大獲利前十天，看看是否真的都是具有黑天鵝亦或特殊事件發生:

表 14:OU 過程+中位數定律策略事件驅動表

| 單日獲利記錄<br>前 10 名 |
|---|
| 0.567816%(2016/06/23) 英國脫歐公投日<br>0.349596%(2016/11/08) 美國總統大選-川普當選<br>0.267304%(2016/06/24) 英國脫歐公投開票日<br>0.145976%(2016/02/01)<br>0.134224%(2016/03/16)<br>0.129186%(2017/01/05)<br>0.126260%(2016/02/24)<br>0.116604%(2018/02/07)<br>0.112332%(2018/07/05) 中美貿易戰-美國對華徵稅正式生效<br>0.108189%(2018/04/03) 中美貿易戰-美國宣佈加徵新 1133 項商品關稅 |

我們發現前十大獲利日其中有一半皆是國際情勢發生重大轉變的事件當天，我們對這種現象的解讀是來自 Avellaneda and Lee (2010)，他們發現在股票波動正常的水準時用超過二十個因子去分析資產報酬率，才能找到真正的超額報酬。因此我們推測資產報酬率在絕大多數情況下確實是多因子模型，但是當重大事件發生時，重大事件對於資產報酬率的影響會變得非常巨大，遠遠高出其他因子，也就是逐漸趨近單因子模型，那我們利用另一資產規避掉這個因子的風險時，超額報酬即會體現。

我們以圖解的方式說明，在沒有重大事件驅動市場時，資產報酬率比較難以預測，我們的統計套利模型只規避了最主要的全球經濟因子，但是每個資產所受到的引力來源並不相同(甚至可以說有很大的差異)，所以我們沒有辦法完全的規避多因子風險後萃取出超額報酬，以原油配小 S&P 為例(WTI&SPX)，如圖 19:





DOI:10.6814/NCCU201900124

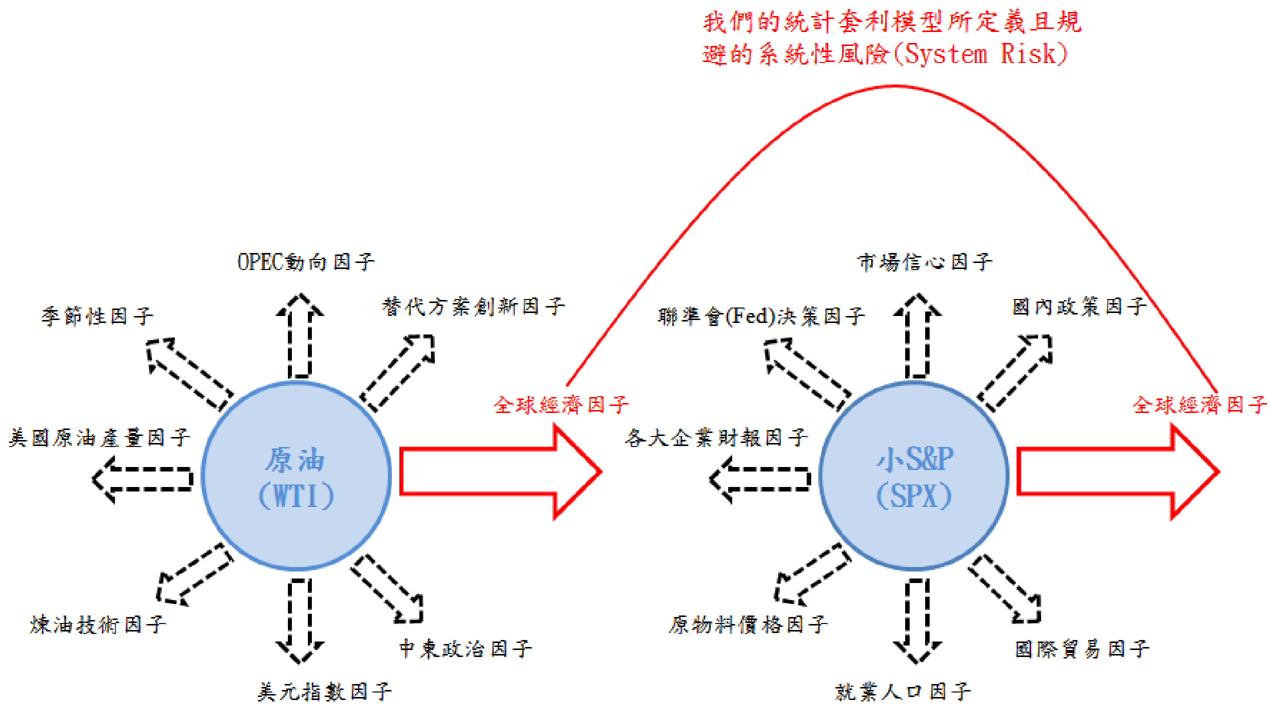



圖 19:正常波動度下多因子模型示意圖

紅色實線箭頭是我們有考慮到的全球經濟因子，黑色虛線箭頭則是模型沒有考慮的其他因子，我們的統計套利模型對這兩資產的鏈結即是規避全球經濟因子後賺取超額報酬。但我們規避它後，殘留下來的其他因子，我們模型沒有對這些因子做處理，而它們會影響到我們模型的預測能力，這些因子所帶來的影響被模型當成超額報酬，殘差$e_t$。

但是英國脫歐當天，全球的經濟情勢皆受到其大幅影響，該事件因子在短時間內支配並驅動了全球絕大部份資產報酬率的走向，這代表只要我們規避該事件因子的風險，我們就更有可能找出真正的超額報酬，殘差$e_t$。如圖 20:



DOI:10.6814/NCCU201900124

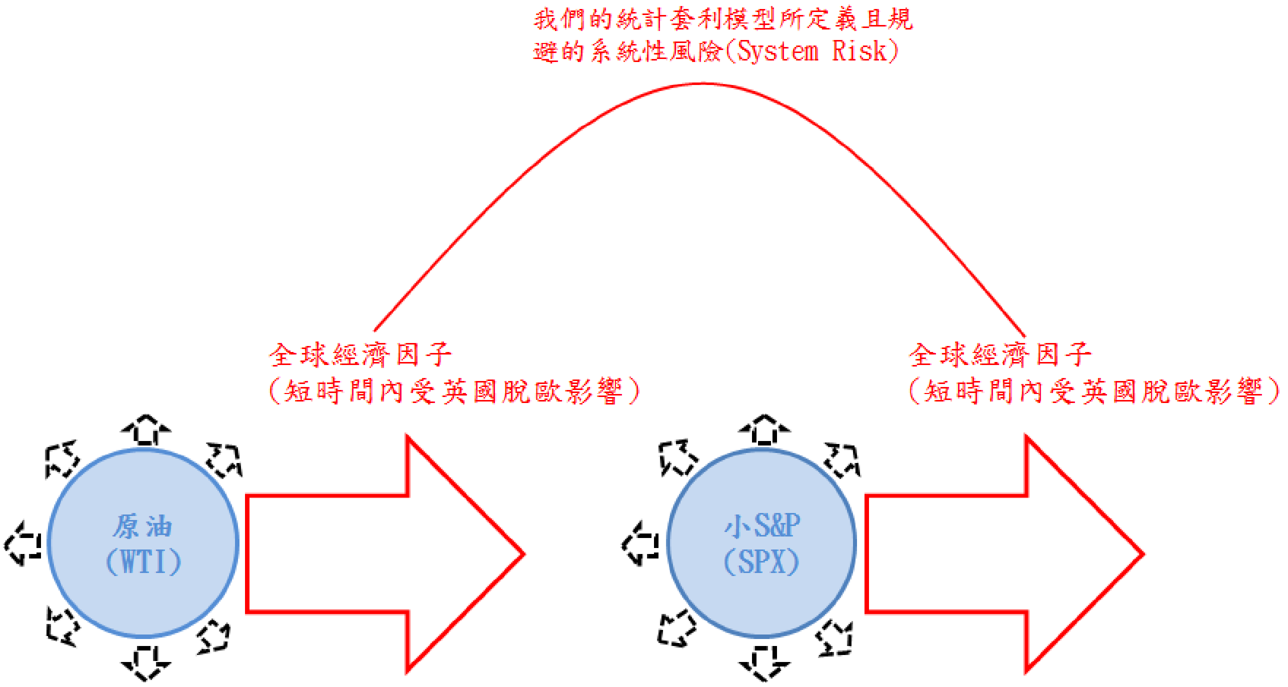



圖 20:事件驅動下趨近單因子模型示意圖

該事件因子影響力非常大，同時也讓其他因子的影響力變小，那殘差$e_t$就更能往我們所預期的方向前進，進而獲取超額報酬。

總結這一小節，我們從結果反推出最大獲利日的原因，即為事件驅動，在事件發佈當下，我們雖無法得知資產價格會往哪個方向移動，但我們知道資產報酬率會趨近單因子模型，並且一定會有移動，短期市場也一定會發生結構性改變，規避這個因子後所獲得的，即為超額報酬，殘差$e_t$。

# 第五章　建議與結論

## 第一節　結論

我們根據 Vidyamurthy (2004)中的統計套利方法做出改善，依據 Engle and Granger (1987)二階段共整合檢定法，建立統計套利模型，探討市場的超額報酬是否真的存在。首先我們改善了傳統配對交易使用價格建構模型的問題，說明報酬



DOI:10.6814/NCCU201900124

率與價格建構模型的優劣，接著解釋我們的對價格資料的降噪流程，在完成資料初步篩選後，我們說明為甚麼使用無截距迴歸模型而並非簡單迴歸模型，再來我們為了防止市場結構發生改變，使用共整合動態校準框架監控模型，強調資料定態的重要性，定態的性質才能使量化模型有意義。在框架底下，我們採用兩種假設去分析殘差，其一為假設殘差是獨立同分配，我們採用技術指標布林通道，其二為假設殘差具有序列相依性，我們把殘差視為會均值迴歸的 OU 過程，最後我們使用 Andrew(2009)的中位數定律輔助殘差異常判斷，期望讓預測更準確。

實證研究我們首先解釋使用一分鐘價格而非其他頻率的原因，接著介紹研究的五檔期貨，分別是歐元期貨(EUR)、英鎊期貨(GBP)、小那斯達克指數期貨(NSX)、小 S&P 指數期貨(SPX)、原油期貨(WTI)。實證研究結果方面，假設殘差為獨立同分配的布林通道策略得到的結果相當優異，加入中位數定律後某些組合的勝率更是接近 75%，但我們也發現某些日期的獲利特別突出。而在假設殘差為序列相依性的 OU 過程策略中結果雖然也很不錯，但卻沒有布林通道策略來的優秀，加入中位數定律後卻也有畫龍點睛的效果，同時我們也從結果中發現與布林通道策略類似的現象，某些日期的獲利很高。接著我們嘗試對這些日期做研究，我們發現許多異常的日期皆有黑天鵝事件發生(如:英國脫歐公投、美國總統川普當選)，我們依循 Avellaneda and Lee (2010)的研究，推測資產報酬率會從多因子逼近單因子。

我們確實成功建構能夠預測短期市場走向的統計套利模型，但這模型本身還不完善，接著我們分別以實務面向與研究面向分別給想對這方面深入探討的人建議。

## 第二節　實務面向建議

以實務面來說，要克服的點如下列:



DOI:10.6814/NCCU201900124

1. 摩擦成本(滑價與手續費):

實務面中最重要的就是交易成本，尤其我們的模型偏向高頻統計套利，交易成本更是不可忽略的點，我們建議可以嘗試把頻率降低到 2~5 分鐘去嘗試建模，並且在下單時不是只以市價單(Market Order)為考量，而加入限價單(Limit Order)的預排。若能對訊號產生當下的市場微結構進行分析更好，優化下單流程，就能改善摩擦成本帶來的損失。

2. 標的物選擇:

我們選擇的 CME 五檔商品主因是成交量大並且為逐筆搓合，最重要的是為期貨，在放空上並沒有太大困難，但若嘗試把模型運用在台灣，目前台灣符合上述性質的資產只有台灣股價指數加權期貨(TXF)，台灣個股期貨的交易量與買賣價差(Bid-Ask Spread)不穩定，現貨來說雖然交易量穩定，但尚未實施逐筆搓合，且融券費用極高，所以在台灣股票類別我們沒辦法找到適合這模型的多標的池。但我們認為原物料商品期貨會是不錯的選擇，符合我們上述所說性質，且有足夠的歷史資料供回測。

3. 倉位與資金控管:

我們的模型倉位為二分法，只有開倉與平倉，單位皆為一，但我們可以根據其偏離的絕對值大小決定放入的資金量或倉位。

## 第三節　研究面向建議

1. OU 過程優化:

從實證結果我們發現 OU 過程並不如布林通道策略，但我們認為 OU 過程如果經過優化，或許能有不一樣的結果，例如我們只假設假設$\sigma^2 = 1$並不符合實際



DOI:10.6814/NCCU201900124

狀況，而是該對其進行嚴謹估計。$\theta$代表復歸速度，我們只對其做閥值的篩選，但它最大的用途是資金配置，復歸速度快配給更多的資金，而資金復歸小則配給小資金。

2. 價格與 EMA 半衰期研究:

我們整個模型並沒有對參數進行任何最佳化，且只用到一個參數 60(迴歸與共整合關係的資料時間長度)，但這數字沒有任何理論依據，只是根據我本人自己的想法取一個整數，EMA 的半衰期與參數息息相關，可以嘗試用時間序列模型分析價格，找出落後期數，搭配 EMA 找出適合的參數。

3. 殘差波動率與閥值研究:

策略的閥值也是一樣，我們自行定義一倍&一點五倍&二倍標準差，但這數字也缺少數據佐證，可以利用波動度模型(如 Garch 模型)嘗試研究殘差的波動大小，再從其中訂下閥值，這也能呼應實務面的交易成本，只要單筆交易獲利越高，受到交易成本的影響將會越小。

4. Beta 的研究:

模型中使用的 Beta 閥值為 0~5 之間，這點與前兩個建議相同，皆是無數據佐證自己定義的閥值，但 Beta 閥值在策略中至關重要，若能加入迴歸分析中的模型檢測，檢測該模型是否配適得當(如 R-square)，或許能提高模型不小的預測能力。





DOI:10.6814/NCCU201900124

# 參考文獻


[1] 陳旭昇，2013。時間序列分析: 總體經濟與財務金融之應用。臺灣東華。

[2] Andrew Pole.(2009).Statistical Arbitrage. *John Wiley & Sons*.

[3] Avellaneda, M. and J-H. Lee (2010). Statistical arbitrage in the US equities market.*Quantitative Finance*, Vol. 10(7), 761–782.

[4] Dickey, D. A., and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427-431.

[5] Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 251-276.

[6] Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2), 111-120.

[7] Said, S. E., and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599-607.

[8] Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management*, 21(1),49-58.

[9] Vidyamurthy, G. (2004). Pairs trading: quantitative methods and analysis (Vol. 217). *John Wiley & Sons*.




DOI:10.6814/NCCU201900124

# 附錄一　中位數定律證明(Andrew, 2009)

首先我們利用 $X=\ e_t$和 $Y=\ e_{t-1}$來簡化符號。因此，這兩個事件就可以分別表示成:

$$\{\mathrm{X} < Y \cap Y > m\}\text{和}\{\mathrm{X} > Y \cap Y < m\}$$

整個狀況拆解成兩個部份(從 Y>m 和 Y<m 不能同時發生)，因此個別部份的機率，就可以單純的採用個別機率的總和來表示。首先考慮被我們拆開來的第一部份:

$$Pr[X < Y \cap Y > m] = \int_m^\infty \int_{-\infty}^y f_{XY}(x,y)\,dx\,dy$$

其中$f_{XY}(x,y)$符號表示 X 和 Y 的聯合密度函數。根據獨立性的假設，所謂的聯合密度，也就是個別密度函數的乘積，而且在這裏的例子中，同時還具有同分配的條件(這也是假設而來)，其累積密度函數則用 *F(.)*表示，因此可以推導如下:

$$\begin{aligned}
\int_m^\infty \int_{-\infty}^y f_{XY}(x,y)\,dx\,dy &= \int_m^\infty \int_{-\infty}^y f(x)f(y)\,dx\,dy \\
&= \int_m^\infty F(y)f(y)\,dy \\
&= \int_m^\infty F(y)dF(y) \\
\int_m^\infty F(y)dF(y) &= \frac{1}{2}F(y)^2 |\,{}_m^\infty \\
&= \frac{1}{2}[(\lim_{t\to\infty} F(t)^2) - F(m)^2] \\
&= \frac{1}{2}[1-(\frac{1}{2})^2] \\
&= \frac{3}{8}
\end{aligned}$$



DOI:10.6814/NCCU201900124

對於被我們拆開的第二部份來說，在經過一個初始的代數簡化之後，結果還是會遵循著類似的推導過程。首先要注意一下，事件 Y<m 可以被表示為，兩個拆開的事件取聯集之後的結果:

$$Y < m \equiv \{(X < Y) \cap (Y > m)\} \cup \{(X > Y) \cap (Y < m)\}$$

根據中位數的定義(m 是分配的中位數)，事件 Y<m 的機率，正好是一半。兩個拆開的事件取聯集，在機率上是加法的關係，因此我們可以利用這個事實，將式子改寫成:

$$Pr[X > Y \cap Y < m] = \frac{1}{2} - \Pr[X < Y \cap Y < m]$$

現在，按照上面處理第一部份的方式推導:

$$\begin{aligned} Pr[X > Y \cap Y > m] &= \frac{1}{2} - \int_m^\infty \int_{-\infty}^y f_{XY}(x, y)\, dx\, dy \\ &= \frac{1}{2} - \int_m^\infty F(y) dF(y) \\ &= \frac{1}{2} - \frac{1}{2}[(\lim_{t\to\infty} F(t)^2) - F(m)^2] \\ &= \frac{1}{2} - \int_m^\infty F(y) dF(y) \\ &= \frac{3}{8} \end{aligned}$$

最後再把兩個部份的機率加起來，就可以得到四分之三的結論。Andrew(2009)中有證明即使是非常數變異數(Nonconstant Variance);還有序列相依性;或非常態分配(Nonnormal Distribution)等等…都試用這個定律。



DOI:10.6814/NCCU201900124

# 附錄二　OU 過程殘差弱定態證明

證明$e_t$為弱定態過程，$e_t \sim N(\mu_e, \frac{\sigma^2}{2\theta-\theta^2})$

證明:

$$
\begin{aligned}
E(e_t) &= a + bE(e_{t-1}) \\
&= a + ab + b^2E(e_{t-2}) \\
&= a + ab + ab^2 + \cdots + b^tE(e_0) \\
&= a\frac{1}{1-b}, (b < 1, t \to \infty) \\
&= \frac{a}{1-b} = \frac{\theta\mu_e}{1-(1-\theta)} = \mu_e \\
Var(e_t) &= Var(a + be_{t-1} + \varepsilon_t) \\
&= b^2Var(e_{t-1}) + \sigma^2 \\
&= b^2\mathrm{Var}(\mathrm{a} + \mathrm{b}e_{t-2} + \varepsilon_{t-1}) + \sigma^2 \\
&= b^2(b^2Var(e_{t-2}) + \sigma^2) + \sigma^2, (b < 1, t \to \infty) \\
&= \sigma^2 + b^2\sigma^2 + \cdots + (b^2)^{\mathrm{t}}\mathrm{Var}(e_0) \\
&= \frac{\sigma^2}{1-b^2} = \frac{\sigma^2}{2\theta-\theta^2} \\
Cov(e_t, e_{t-h}) &= E[e_t - E(e_t)][e_{t-h} - E(e_{t-h})] \\
&= E(e_te_{t-h}) - E(e_t)\mathrm{E}(e_{t-h}) \\
&= E[(a + be_{t+h-1})e_t] - {\mu_e}^2 \\
&= E[(a + b(a + be_{t+h-2})e_t] - {\mu_e}^2 \\
&= E[(a + ab + ab^2 + \cdots + ab^{h-1} + b^he_t)e_t] - {\mu_e}^2 \\
&= E\left[\left(\alpha\frac{1-b^h}{1-b} + b^he_t\right)e_t\right] - {\mu_e}^2 \\
&= \frac{\alpha}{1-b}(1-b^h)E(e_t) + b^hE({e_t}^2) - {\mu_e}^2
\end{aligned}
$$



DOI:10.6814/NCCU201900124

$$= {\mu_e}^2(1-(1-\theta)^h) + (1-\theta)^h(\frac{\sigma^2}{2\theta-\theta^2} + {\mu_e}^2) - {\mu_e}^2$$

$$= (1-\theta)^h(\frac{\sigma^2}{2\theta-\theta^2})$$





DOI:10.6814/NCCU201900124