

第三章 研究方法

第一節 研究樣本與期間

3.1.1 研究期間

2003 年 ~ 2004 年 8 季資料

3.1.2 研究樣本

一、資料來源及家數

本研究以台灣上市資訊電子產業以及塑膠產業公司為研究樣本；起先，由台灣經濟新報取得的資料樣本數分別為：上市資訊電子產業公司有 310 家，上市塑膠產業公司有 22 家，經過資料整理後，排除研究期間內資料不足者，實際做為本研究的研究樣本為：上市資訊電子產業公司有 246 家，上市塑膠產業公司有 20 家。

二、採用變數期間

O'Neil 所提出 CAN SLIM[®]選股策略 7 個指標中，由於 N 及 M 較不屬於數量化的指標，因此本研究僅以其它五者來做分析，其中，專業投資機構(I)則以國內外信託基金家數為代表。依據 O'Neil 提出的最原始想法，C 為當季每股純益較去年同季的成長率，因此，若研究期間為 2003 年第一季度，此時所採用的變數則為 2002 年及 2003 年第一季度資料，而 A 變數為五年每股純益複合成長率，因此採用 1998 年到 2002 年資料，而 S 及 L 變數則採用研究時間 2003 年第一季度資料，I 變數則採用 2002 年年報的資料，其餘 7 季所變數所採用的時間以此類推，整理於下表：

[表 3.1-1] 各指標變數採樣時間表

研究期間	C	A	S	L	I
2003 1Q	2002.1Q/2003.1Q	1998-2002	2003 1Q	2003 1Q	2002
2Q	2002.2Q/2003.2Q	1998-2002	2003 2Q	2003 2Q	2002
3Q	2002.3Q/2003.3Q	1998-2002	2003 3Q	2003 3Q	2002
4Q	2002.4Q/2003.4Q	1999-2003	2003 4Q	2003 4Q	2003
2004 1Q	2003.1Q/2004.1Q	1999-2003	2004 1Q	2004 1Q	2003
2Q	2003.2Q/2004.2Q	1999-2003	2004 2Q	2004 2Q	2003
3Q	2003.3Q/2004.3Q	1999-2003	2004 3Q	2004 3Q	2003
4Q	2003.4Q/2004.4Q	2000-2004	2004 4Q	2004 4Q	2004

第二節 統計方法 — 群集分析(Cluster Analysis)

3.2.1 群集分析的意義

在各種領域的研究中，若面臨到必須將所看到的資料分成幾個「有意義」的組時，則群集分析則是提供這樣「分類」的工作；主要目標是依照收集的 p 個變數($x_1 \sim x_p$)將 n 個個體分幾個群，使群內個體間離的近(或相似)，不同群的個體離的遠(或差異性大)。(群集分析方法也可以對變數做分群)。一般群集分析的分類法有「聯合分群法」(Joining)(或稱樹形分群法)、「雙向聯結法」(Two-Way Joining)(或稱集區分群法)、「 K 組平均數分群法」(K-Means Clustering)。

3.2.2 群集分析二大主要工作，定義：

一、第一步：距離量測(個體間距離)

聯合分群法(樹形分群法)在結合群體時是利用個體間的不相似性(或距離)為依據，距離可以是一維空間或多維空間的距離；對兩個個體在多維空間(多變量)，量測其距離，最容易想到的是歐氏距離(Euclidean Distance)，在 2 維或 3 維空間，此量測就是兩個體在空間上的真正幾何距離。

(一)歐氏距離(Euclidean Distance)

(二)歐氏距離平方(Squared Euclidean Distance)

(三)城市街道距離(City-Block Distance)

(四)謝比雪夫距離

(五)不一致的百分比

(六)相關係數

二、第二步：聯結法(群體間距離)

在第一步將距離近的兩個體合成一群後，還必須決定群與群結合的依據，計算兩群的分方很多，例如：

(一)單一聯結法(最近法)(Single Linkage)

(二)完全聯結法(最遠法)(Complete Linkage)

(三)平均法(Average Linkage)

(四)中心法(Centroid)

(五)華德法(或華德(Ward)最小變異數法)：以變異數分析方法來定義群間距離，也就是在每一個分群步驟試圖找出二群，使此兩群的組內變異平方和最小；此法被認為是很好的分群法，通常會得到較少的群組。

3.2.3 群集分析三大型式

一、分層法(Hierarchical)

(一)凝聚分層法(Agglomerative)

開始時每一個為一群，然後最近的兩個體合成一群，依次結合，使群組愈變愈少，最後所有個體結成一群。特性是兩個體一旦在同群內，則以後的步驟中必然在同一群內，此法最為普遍使用。

(二)分離分層法(Divisive)

開始時所有個體為一群，然後分成二群、三群，直到每一個體為獨立一群為止，方向與凝聚分層法相反。此法則較少使用。

優點：可以合理判斷應分群數目

缺點：分群結果常受到偏遠樣本的影響，對已歸併錯誤之樣本無法加以適當調整

二、非分層法(Nonhierarchical)

以 K 組平均法(K-Means)為代表：

假如已知要將所有個體分成 K 群，指定 K 群的中心值，依各個體到各中心點距離遠近重新移動個體到最近的群體，並算出各群體新的中心點，然後繼續再移動各個體到最接近的群，如此重覆進行直到個體不能再移動為止。

優點：可避免偏遠樣本誤差對整體分群影響

缺點：需要先給定群集數目始能分析

三、二階段法

結合分層法與非分層法二種方式，M. Anderbeg 在 1973 年提出，以任何一種分層群集方法求得理想群集數目，再利用非分層方法進行群集分析。例如：

第一階段：以華德法做分群，決定群組個數

第二階段：以 K-組平均法進行群集

目的是由於第一階段華德法是屬於分層法群集分析，它們的共同缺點是二個個體一旦被分在一群，之後就會永遠在同一群內，而第二階段的 K-組平均法屬於非分層法，可以彌補這種缺點。

3.2.4 群集數評量指標

一、RMSSTD (Root – Mean – Square Standard Deviation)

新群體中所有變數的綜合標準差，值愈小表群體內個體的相似性愈高

$$RMSSTD = \sqrt{\frac{\sum_{i=1}^p S_i^2}{p}}$$

S_i 為第 i 個變數的各群內的標準差之和， p 為變數個數

二、 R^2 (R - Square)

當 R^2 愈大表示群體間的相異性愈高，也間接顯示分群效果愈好

三、SPR (Semipartial R – Square)

以原來兩個群體組成新體，其所損失群內相似性比例為指標；值愈小，表示形成新結合群體時，損失的群內相似性比例愈小，也就是新群體的群內相似性愈高。

$$\text{SPR} = \frac{\text{新結合群體的群內變異平方和} - \text{原本未結合前群內變異平方和}}{\text{所有樣本的組內變異平方和}}$$

四、群集係數值

值愈小表示該群集成員愈同質，反之則成員相異性愈高。(張紹勳 2003)

五、樹狀圖

切點的選擇可以找距離較遠的位置當切點，這樣可以防止把分得較開的兩群錯誤合併；也可先決定群數再找切點。

3.2.5 與其它分類分析不同之處

— 群集分析(Cluster Analysis)

群集分析與其它分類分析(如區別分析)不同的地方在於群集分析的各群組特性事先完全未知，而分組完全是由資料所導出，群集分析方法的目標是希望群集出在同一群組內的個體在某種意義上有相近的性質，而不同的群組間差異性大。

— 區別分析(Discriminant Analysis)

區別分析是已知每個觀察個體所屬的群體，最主要是要先找到有解釋能力的區別變數；在做區別分析時，所有的區別變數都可列入，

再利用統計上的變異數分析(ANOVA 或 MANOVA)來作檢定，看這些區別變數的區別能力對於分類是否有顯著，或也可利用逐步區別分析(Stepwise Discriminant)選取有用的預測變數，再利用區別函數建立區別規則；區別分析方法可以指出哪個變數對分類較有貢獻。

— 主成份分析(Principal Component Analysis)

主成份分析是用來對一組內部相關的變數做簡化的描述，目的在於將許多變數予以減少，使其變為少數幾個互相獨立線性組合變數(即主成份)，這些主成份保有原變數最多的資訊(即有最大變異數)，受測者會在這些成份上顯現出最大的個別差異；是一種將原來變數轉換成新的沒有相關的變數的方法。

— 因素分析(Factor Analysis)

因素分析是主成份分析的擴展，這二種方法都是針對內部相關性高的變數做資料簡化的工作；然而主成份分析是以變異數為導向，而因素分析是以共變異數為導向，關心每個變數與其它變數共同享有(Share)部分的大小，另外，主成份主要是選擇一組成份(Component)盡可能解釋原變數的變異數，而因素分析主要是選取因素，它能解釋原變數之間的相關情形。

由於本研究是希望能夠將樣本資料做最合適的分群後再進行之後的觀察，而各群組的特性在事先是未知的，且目標是希望能將有同樣特性、同樣有潛力成為飆股的樣本分在同一群，如此正符合群集分析法的目的，而不同於區別分析主要在於找到有解釋能力的區別變數、主成份分析和因素分析法主要針對內部相關性高的變數做簡化的工作；因此，本研究最後決定採用群集分析法做為本研究的分群統計方法。

第三節 實驗設計

此節主要針對本研究實驗設計中的群集分析過程做詳細的描述，讓整個研究實驗的流程更為清晰。第一部分為資料處理的過程，由於自台灣經濟新報所取得的原始資料，有些並非本研究欲討論的指標，因此必須先經過計算整理後，才可進行分析；在經過資料整理後，整個實驗的流程與執行過程則於第二部分做說明。

3.3.1 資料整理

本研究對於所取得的變數資料分為三個階段的整理步驟。第一階段為計算研究樣本各項變數的值，第二階段則針對無法計算的資料進行資料清除(data clean)，並計算 L(股價相對報酬強度)的值，第三階段則針對複合成長率公式再做修正，將有問題的 A 值再做修正。

第一階段資料整理：計算研究樣本的各项指標

C：當年度季 EPS 較前一年度同季的成長率

$$\frac{EPS_t - EPS_{t-1}}{|EPS_{t-1}|}$$

A：計算五年度 EPS 的年複合成長率

$$\sqrt[4]{\frac{EPS_t}{EPS_{t-4}}} - 1$$

S：找出研究期間各樣本流通在外股數(資料可直接取自 TEJ)

I：找出研究期間各樣本專業投資機構認同的家數

(本研究使用信託基金家數做為專業投資機構，將自 TEJ 取得的國內和僑外信託基金家數加總)

L：先經過第二階段 data clean 的資料整理後，確定可做為當季研究的

樣本後，最後計算 RPS。

$$RPS = \left[\frac{(1 - \text{漲幅排名})}{\text{全部股票}} \right] * 100$$

第二階段資料整理：進行 data clean

C：透過上述成長率公式，當研究樣本去年同季的每股純益(即 EPS_{t-1}) 為零時，將造成無法計算的問題，因此，若發生分母為零的問題，在本研究中則將此公司從當季的研究樣本中刪除。

A：透過上述複合成長率公式，當研究樣本期初或期末的 EPS ，有一為負時，將使根號內的數字為負而無法計算，因此，若發生根號內數字為負的情況，則將此公司從當季的研究樣本中刪除。

第三階段資料整理：複合成長率的修正公式

若研究樣本期初和期末的 EPS 皆為負，不論期末表現較期初差，如期初為 -1，期末為 -4，或期末表現較好，原公式計算的結果將正好相反，因此，將公式修正，以符合常理。

$$\text{修正公式} = - \left[\sqrt[4]{\frac{EPS_t}{EPS_{t-4}}} - 1 \right]$$

3.3.2 實驗流程

一、研究樣本分群

本實證首先將研究樣本利用多變量方法中的群集分析做分群；在第二節已經說明有關群集分析的类型，由於分層法(Hierarchical)和非分層法(Nonhierarchical)各有優缺點，因此，本研究採用 M.Anderbeg 在 1973 年提出結合二者的想法，也就是群集分析中的二階段法；在第一階段時先以分層法中的華德法(Ward)決定群組個數，再利用非分層法進行第二階段群集

分析。

其中，在進行群集分析時，一開始所使用的分群資料”A”為 O’Neil 所提出的「五年」每股純益複合成長率；然而，以台灣快速變化的市場來看，五年似乎太長，無法確實反應台灣市場特性；因此，本實證也另外考慮以「三年」每股純益複合成長率做為分群資料，即利用各研究樣本的三年每股純益複合成長率取代原本的五年每股純益複合成長率再進行分群與比較。

下列為二階段分群法：

(一)第一階段分群：

採用分層法中的華德法(Ward)。將研究樣本分為台灣上市資訊電子產業和塑膠產業，截取 CAN SLIM[®]選股策略中的五項指標：C(每當季每股純益)、A(年度每股純益成長)、S(流通在外股數)、L(股價相對強弱指標)、I(專業投資機構認同)做為分群的變數；研究期間為 2003-2004 年的季資料，因此，分別以華德法進行群集分析，共執行 32 次(二個產業在不同的”A”資料情況下)，並以統計學家提出的三個指標 RMSSTD、 R^2 、SPR 做為群集數的評量指標，最後決定此階段的分組群數。

(二)第二階段分群：

此階段為非分層法分群，本研究以 K-Means 為代表；在第一階段華德法決定分組群數後，利用 K-Means 方法和所得到的分群數再進行一次群集分析，得到最終研究樣本分群結果，即各季各群組的成份公司。

二、與類指報酬做比較

在研究樣本利用群集分析法決定最後分群結果後，則將各群組在投資期間內的平均投資報酬與類指報酬做比較，在此部分主要針對各指標的門檻值及投資期間的長短做敏感度分析。

(一)各變數門檻值：

在此節的第一部分－「選股策略」中已描述 O'Neil 所提出 CAN SLIM®各指標的原始門檻值；因此，首先即整理出各季各群組能夠符合此原始門檻的數目。其中，本研究是以各群組中各個指標的中位數做代表，計算各群在當季符合指標門檻值的數目。觀察是否符合指標門檻值愈多的群組，其未來投資期間表現愈可能為當季群組中表現最好的一群，並且也超越類指表現。

另外，由於 O'Neil 認為有潛力成為飆股的條件必須符合其所提出的各指標門檻值；然而，這樣的標準卻未必適用於台灣股市，因此，本實證再以不同的門檻值來做敏感度分析，觀察在不同的門檻值下是否與原始門檻所造成的結果有所不同，進而找出較適合台灣股市的標準。

(二)投資期間：

除了在各指標門檻值上做敏感度分析，本實證也針對 1 個月、3 個月和 6 個月的投資期間做探討，觀察 CAN SLIM®選股策略較適合長(或短)期投資策略。以研究期間為 2003 年第一季為例，投資買進的時點為第一季最後一個交易日，若投資期為一個月，賣出時點為下個月的最後一個交易日，若投資期為三個月，賣出時點為第二季最後一個交易日，若投資期為六個月，賣出時點為六個月後的最後一個交易日；其它研究期間各個群組的投資買賣時點則以此類推。

(三)實驗流程圖

[圖 3.3-1 實驗流程圖]

