

因素分析運用於股票與基金市場相關性 探討

李泰明·趙曉晴*

(收稿日期：98 年 9 月 3 日；第一次修正：99 年 2 月 23 日；
第二次修正：99 年 4 月 2 日；接受刊登日期：99 年 4 月 15 日)

摘要

本研究運用 Error Corrected Model，將財務時間序列資料間存在的關聯性，如自我相關，共整合先予以抽離，這樣使得資料做初步分析後，能排除這些特殊意義的關聯成份，讓訊息更顯明確，也讓模型的殘差序列資料穩定且合乎因素分析假設。進而對殘差序列資料做因素分析，萃取出資料短期的關聯因子。本研究運用 EM 演算技術在因素分析的參數估計上，這樣的方法相當便捷及具效率。實證的資料分析結合股票與基金市場，找出兩市場間的互動關聯因子，發現台灣股票市場及開放式股票型基金幾乎只存在一個主要的關連性，也就是所謂的短期伴隨效應，自我相關及共整合效應並不顯著。這種分析的結果，將多變數時間序列資料之間的不同關聯結構完整地分析出來。這是之前大多數的研究者所無法釐清的事，乃是本文主要之貢獻。更為此類資料分析提供一個完整的範例程序。

關鍵詞彙：因素分析，時間序列，誤差修正模型，EM 演算法

壹·前言

經濟行為可以透過實際資料來觀察，研究學者根據經驗方法找尋需要的資料或者根據經驗直接設想背後的經濟模型。我們分析個體或總體經濟所產生的傳統時間序列資料，基本上可由最簡單的向量自我迴歸模型 (Vector Autoregressive Models) 來做為開頭。而大多數的分析技巧都不易得到適當的模型。若觀察者執著於收集有代表性的資料，意味著要等待更多的時間，而一個景氣循環更可能要等上好幾年的時間。實際上，如何從大量資料中發掘有用的訊息比等待收集更具意義。我們可以在上市公司定期公開的資訊中，找到上百個變數，但分析方法跟不上資訊的膨脹，擁有大量資料的政府經濟人員，卻只判斷了簡單的經濟的狀態，暗示我們並未充份萃取出有用的訊息，參考 Heaton (2008)。

* 作者簡介：李泰明，輔仁大學統計資訊學系副教授；趙曉晴，輔仁大學應用統計研究所碩士生。

財務時間序列資料具有特殊經濟意義的結構，因此發展出貼切於財務資料本身特性的因素模型及假設，參考 Forni et al. (2000, 2004), Bernanke et al. (2005), Stock and Watson (2005), Favero et al. (2005)。在傳統的應用，多於描述經濟指標及作情勢的預測。總體經濟方面，如 Sargent and Sims (1977) 預測美國景氣循環，Bernanke and Boivin (2003) 討論貨幣政策，並形容變異與時間相關的動態關係及加入 AR (Autoregressive Models) 及 VAR (Vector Autoregressive Models) 模型來做討論。針對處理高維度的資料，Stock and Watson (2002a and 2002b) 研究表示所使用的因素模型對於某些總體經濟變數的分析可提高預測能力，Boivin and Ng (2005) 並針對其分析方法再做改善。

由 Sargent and Sims (1977) 提出 Dynamic factor model 的一般式來討論時間動態的關係開始，接續延伸將時間序列模型 AR、VAR、VARMA (Vector Autoregressive Moving-Average Models) 結合因素模型，如 Bernanke, Boivin and Elias (2004) 將 VAR 結合因素模型、Forni et al. (2004) 結合 VAR 及 VARMA、Dungey et al. (2000) 結合 GARCH (Generalized autoregressive conditional heteroskedasticity) 模型。而形容總體經濟均衡的 ECM 模型 (error-correction model) 應可以帶給我們不一樣的訊息。

本研究利用因素模型來分析財務時間序列資料，基本上當然是多變量的組合。因素分析在經濟上的應用是相當吸引人的，它是一個整合大量變數的技術，而這些變數可能受少數未觀察到的因素 (隱藏變數) 所影響，這樣的概念是所有經濟學者公認的，特別是運用在經濟循環理論和財務領域上，是一個成功分析多變量資料的工具。以台灣股市和基金市場資料做為實證分析，將能有效地展現出這個市場所獨有的特性。

總體經濟變動是在複雜的變數相互影響下所產生的，許多研究結果曾指出股市與總體經濟間變動的關係 (匯率、利率、進出口等...) 或者是基金市場與股票市場間的關連性。那麼站在結合整個市場的角度，我們如何知道這些影響變動的因子，佔整體關聯性的百分比呢？譬如在股票市場中，可能有長期 (共整合) 及短期 (伴隨效應) 的影響，也有前期與後期連動的自我相關成份，本研究嘗試運用 ECM 模型 (error-correction model) 展示時間序列間的聯連結構，ECM 模型是一種具有特殊型式的經濟模型，其主要結構由 Davidson, Hendry, Srba, and Yeo (1978) 所提出，Engle and Granger (1987) 及 Johansen (1988) 作深入探討。並對各個結構成份的權重作具體表達，以及如何利用因素分析工具分解出隱藏在背後的因子，是本研究所關心的事。

貳·文獻探討

一、財務時間序列

經濟市場的變動，存在許多不同形態的關連結構，因此發展出 AR (autoregressive model) 模型來形容時間的延續性及 MA (Moving-Average Model) 模型，MA 模型隱含了在經濟行為體系中，其結構式有 "error correction" 的特性，之後又定義了共整合效用 (Cointegration)，而有誤差修正模型 (error-correction model)。大部分的財務時間序列資料，都具條件變異數不齊一的現象，於是如何將這種會因時而異的條件變異數模型化，因而發展出 ARCH (Autoregressive Conditional Heteroscedasticity) 模型，參考 Engle (1982)。又如財務資料具高狹峰分配及波動叢聚的特性，恰可以 ARCH/GARCH (Generalized autoregressive conditional heteroskedasticity) 模型適切地描述，參考 Bollerslev (1986)。爲了避險及估計報酬，延伸出現了 ARCH-M (ARCH in mean) 模型 (Engle et al., 1987)。這些都是爲了解決實際社會所發生的問題，促使了模型的演化。

時間序列資料經常發生非定態的性質，當變數是非定態時，最大的影響不外乎是由 Granger and Newbold (1974) 發現非定態變數之間，可能會出現假迴歸的問題，使得原本毫無關聯性的變數之間，出現假象的關聯性，即實際上在迴歸模型中反應出的顯著性，並沒有真正的經濟意義。將非定態的模型拆解兩個影響成分來解釋，一爲定性趨勢 (deterministic trend)，二爲隨機趨勢 (stochastic trend)，依據 Enders (2004) 的說法，隨機趨勢就是變數中的隨機成分 (stochastic component) 對該變數所形成永久性 (permanence) 的影響。我們可從 random walk (RW) 模型來模擬這個現象，從此模型產生方式可以發現，原來 RW 模型其實就是 AR(1)模型落後一期的係數等於 1 時的自我相關特例，自然我們就可用自我相關模型來轉化這種時間序列的不平穩問題。

Engle and Granger (1987) 提出共整合 (co-integration) 理論，發現非定態 (nonstationary) 變數之間的迴歸關係如果出現共整合現象，則這樣的迴歸關係仍然有經濟意義，而使原迴歸推論性質也可以適用，參考楊奕農 (2005)。由此，利用誤差修正模型中抽出共整合及 AR 效應後，金融變數間動態關聯結構，還剩下什麼？這是我們好奇的結果。

二、傳統因素分析的難度

因素分析始於心理學上的應用，本文將之運用到財務時序資料上。因為時間序列資料本身自我相關及不平穩的特色會破壞因素分析的基本假設，所以藉由自我迴歸模型 (autoregressive model) 及 ECM 模型 (error-correction model) 的特性，解決了這個障礙。在實際操作上，又必須面臨大量資料的運用容易有遺缺值的問題。另外，是否能夠精確的估計及運算都是因素分析需要克服的困難。

在因素分析的計算過程，估計 Factor Loading 就包括主成份法 (Principal Component method)、主要因子法 (Principle Factor Method)、最大概似估計法 (ML, Maximum Likelihood)、典型相關因子法 (Canonical Factor Analysis) 等...，Turville (2000) 運用 EM 演算法 (Expectation-Maximization Algorithm) 於隨機移除並插補缺漏值，並指出沒有特定的方法能讓差補的結果完美但使用 EM 演算法通常來說卻是最有效益的。Rubin and Thayer (1982, 1983) 將 EM 演算法運用於因素負荷估計，而它所使用的方法卻被 Bentler and Tanaka (1983) 質疑，認為其所提出的 EM 代疊收斂至 MLE (Maximum Likelihood Estimation) 的過程有瑕疵，有可能無法得到其所宣稱的極大值，而本研究將 EM 演算法估計的優勢發揮，除了克服 Bentler 與 Tanaka 所質疑的問題，在本研究提出的 EM 演算公式中，還可同時估計因素分數及因素負荷，達到主成份因素法 (Principal Component Factoring)、MLE 的要求，是相當精簡的計算方法。在本研究因素負荷的估算，即 EM 演算法之 E 步驟的思考與 Roweis (1998) 所提出的想法類似，但其研究在收斂的過程並不合理，本研究加入 Bartlett (1937, 1938) 所得結果，並證實收斂的合理性。

研究中的另一主題為轉軸的最佳化過程，其轉軸的目的是希望讓因素負荷的特性更具有區別力，在使原始資料不變又得獲唯一解的條件下，如何尋找一個最佳化的直交轉軸？套裝軟體中從未交代這些最佳化的過程。本文提出馬可夫鏈蒙地卡羅法 (MCMC, Markov chain Monte Carlo) 進行隨機搜尋極值的方法。

參·研究方法

一、因素分析

本文主要是根據資料型態搭配適當模型能使殘差平穩且符合因素分析的

假設，並將因素分析表達成多變量複式迴歸 (Multivariate Linear Model)。此時，解釋變數代表隱藏變數 (Latent Variables) 的角色，表達式可寫成

$$\underset{p \times 1}{\mathbf{y}} - \underset{p \times 1}{\boldsymbol{\mu}} - \underset{p \times m}{\mathbf{A}} \underset{m \times 1}{\mathbf{f}} + \underset{p \times 1}{\boldsymbol{\varepsilon}} \quad (1)$$

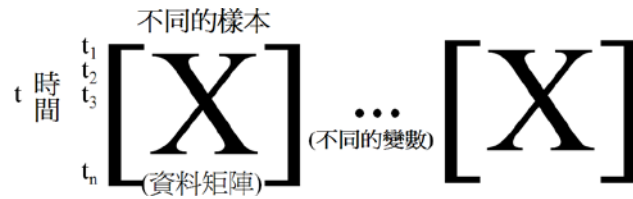
假設觀測值 \mathbf{y} 有 p 個變數，其平均數為 $\boldsymbol{\mu}$ 及共變異數矩陣為 Σ 。而在因素模型中 \mathbf{y} 與少數未觀測到的變數 \mathbf{f} 有線性相依的關係，並稱 \mathbf{f} 為共同因素 (common factors)，其它無法由共同因素表達的部分則用誤差 $\boldsymbol{\varepsilon}$ 來表示，又可稱為 specific factors，其中 \mathbf{A} 是指 factor loading 矩陣。此式是指原本的 Factor Model 有 p 個變數， m 個因子，這是指只有一筆觀測資料的情況。當收集 n 筆資料後，其矩陣表達方式可為：

$$\underset{n \times p}{\mathbf{Y}_c} = \underset{n \times m}{\mathbf{F}} \underset{m \times p}{\mathbf{B}} + \underset{n \times p}{\mathbf{E}} \quad (2)$$

\mathbf{Y}_c 表示觀測值減其平均數後的資料， \mathbf{F} 表示共同因素， \mathbf{B} 為因素負荷以及 \mathbf{E} 表示誤差部分。將多變量資料透過 Principal Component Factoring 及 EM 演算法 (Expectation-Maximization Algorithm) 來估計參數，找出影響這群資料背後的因素。利用這種多變量複式迴歸的模型去詮釋多變量時間序列資料，並直接以這樣的工具去做因素分析。當然因素分析原本的用意並不能分析時間序列資料，這是因為變數間連動的關係非常複雜，所以我們必需用某種方式處理之後再做因素分析，而這樣的因素分析如何解釋也是本文的重點。

二、多變量時間序列

時間序列 (time series) 是一組有順序的且隨時間變化的序列資料，廣泛受社會、經濟、醫學等領域運用，這是因為資料的產生及記錄多為多變量時間序列 (Multivariate time Series) 的資料型態。在醫學上，我們觀測一群糖尿病患者病患在血糖濃度上的變化，記錄每個時間點的血糖濃度量，則隨著時間 t 變化即可收集到一組多變量時間序列的資料。多變量時間序列資料矩陣的特色是其中的一個維度必需給時間軸 t ，則資料矩陣有下面二種表達的方式：



圖一 資料矩陣示意圖 (以不同樣本為例)



圖二 資料矩陣示意圖 (以不同變數為例)

以圖一舉例來說，我們想知道台灣股市中，半導體類股股價短期波動的影響因子？所以我們收集台積電、聯電、聯發科等半導體類股，各股股價資料，有收盤價、報酬率、外資及法人持股比率等變數，而資料放置的方式，看成每一個資料矩陣代表不同的變數，設第一個矩陣代表收盤價，兩個維度分別代表半導體各類股及隨著時間收集的收盤價資料。第二個矩陣代表報酬率，而兩個維度分別代表半導體各類股及隨著時間收集的報酬率資料。以此類推。所以上圖中的…的用意在於延伸另一個維度空間。

以圖二來說，若我們想知道相同的問題，則將資料放置的方式，看成每一個資料矩陣代表不同的半導體公司，設第一個矩陣代表台積電，兩個維度分別代表台積電股價變數及各股價變數隨著時間收集的資料。第二個矩陣代表聯電，而兩個維度分別代表聯電股價變數及各股價變數隨著時間收集的資料，以此類推。

三、多維隨機微分方程的模擬

本研究以自行撰寫 C++ 程式進行模擬，利用 Multidimensional Lognormal Random Walks 及 ECM (error-correction model) 模型來模擬兩種財務上時間序列資料的類型。透過資料的模擬，可以令我們充份理解模型背後的隨機性機制，以及它如何反應在資料上。

股價變動的過程符合幾何布朗運動 (geometric Brownian motion)，可用隨機漫步 (random walk) 來形容股價在一持續時間內的行為，而在特定時間點的股價機率分佈符合對數常態分配 (lognormal distribution)，因此可將股價變動的行為視為 lognormal random walk，將隨機變數 S 表達成 dX (即隨機微變量)

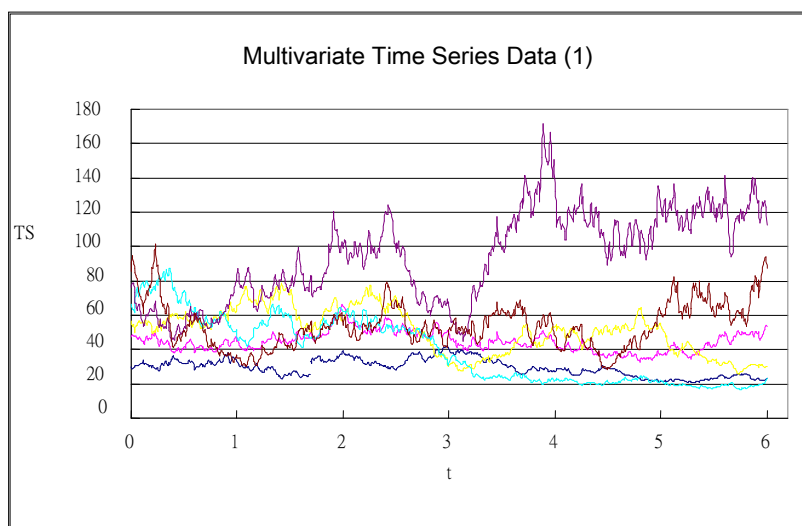
的函數，此完整的隨機過程稱之為 Ito's Stochastic Process：

$$dS = A(S, t)dX + B(S, t)dt \quad (3)$$

將上式特殊化後，即可得 $\frac{dS}{S} = \mu dt + \sigma dX$ ，證明推導 $e^{\mu dt + \sigma dX} - 1$ 及 $\mu dt + \sigma dX_i$ 之間關係，可參考 Norstad (2005)。將股價 S 的變動視為 Lognormal Random Walks。我們想知道的是多檔股票的連動關係，所以利用多變量隨機微分方程來描述這個現象，以下為表達式：

$$dS_i = \mu_i S_i dt + \sigma_i S_i dX_i \quad i = 1, \dots, k \quad (4)$$

我們可以看成同時考慮 k 檔股票，其中 S_i 為第 i 檔股票價格，則 dS_i 表示第 i 檔股票的股價變動， dt 為時間變化量，而 dX_i 是第 i 檔股票的隨機微變量： $E[dX_i] = \mathbf{0}_{k \times 1}$ ， $Var(dX_i) = \sum_{k \times k}$ 。 μ 代表股票的瞬間期望報酬率， σ 代表股票報酬率的瞬間標準差。模型特色在於只能展現出多變量之間“短期連動關係”，我們考慮 ℓ 個時間單位的落差，單獨看時間點 t 與時間點 $t + \ell$ 間的共變異數，其 $\Sigma(t, t + \ell)$ 隨著 ℓ 變大，快速收斂到 $\mathbf{0}_{k \times k}$ ，這樣的結果也相當符合直覺。隨著二個時間點的拉長，點與點之間的共變異也愈小。Tsay (2005) 稱此效應為伴隨效應 (concurrent effect)。



說明：橫軸為時間軸 t ，縱軸為股價指數

圖三 模擬六檔股票股價的時間序列圖

為利用 Multidimensional Lognormal Random Walks 的機制，模擬出六檔股票股價變動的時序序列圖。參數設定 $\sigma_1 = 0.01, \sigma_2 = 0.02, \sigma_3 = 0.03, \sigma_4 = 0.04, \sigma_5 = 0.05, \sigma_6 = 0.06$ ，

$$\text{Var}(dX_i) = \begin{pmatrix} 1 & 0.8 & \cdots & 0.8 \\ 0.8 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0.8 \\ 0.8 & \cdots & 0.8 & 1 \end{pmatrix}_{6 \times 6}, \text{ 起初股價分別為 } 40、50、60、70、80$$

及 90。其中 σ_i 為控制個股股價波動程度， $\text{Var}(dX_i)$ 代表股票之間的相關係數 Σ 。取其中一筆模擬資料如上圖三所示。從整個時間軸觀察，六檔股票股價並無長期連動趨勢，而將時間軸切割成一小段來看，股價在短期下的確存在部分的連動關係。

時間軸在時序資料中充滿了重要訊息，若將時間 t 誤解成不同的樣本，則得到下表一的共變異矩陣及此矩陣的特徵值。從此共變異矩陣的數值來看，出現不合理及異常大的變異數及共變異數。而所得出的特徵值，以凱莎 (Kaiser) 準則，保留特徵值大於 1 的標準來看，6 個中就取出 4 個因素。這不合理的結果，都是因為乎略時間軸所造成的，這也是在研究時序資料之間關聯性的難度所在。

表一 模擬六檔股票股價的變異及共變異數矩陣

Variance and Covariance Matrix =

$$\begin{bmatrix} 0.67822 & 0.43112 & -0.69946 & -0.25814 & -0.31807 & -1.45595 \\ 0.43112 & 2.21794 & -1.39044 & -0.3704 & 5.12584 & 1.91674 \\ -0.69946 & -1.39044 & 4.01868 & 1.34772 & -0.01213 & -1.05552 \\ -0.25814 & -0.3704 & 1.34772 & 3.13045 & -3.23524 & 0.15032 \\ -0.31807 & 5.12584 & -0.01213 & -3.23524 & 30.09649 & 8.25049 \\ -1.45595 & 1.91674 & -1.05552 & 0.15032 & 8.25049 & 11.7785 \end{bmatrix}$$

Eigenvalues of Variance and Covariance Matrix =

$$[34.518 \ 9.154 \ 5.298 \ 2.276 \ 0.497 \ 0.178]$$

四、運用向量自我迴歸模型初步分析

自我迴歸模型 (autoregressive model) 是形容什麼樣的經濟行為呢？簡單說就是現在的某一變數值與過去發生的值有關，可能與上一個時間點或者上二、三個時間差的變數值有關。舉例來說 AR(1) 模型可以描述某經濟變數好恰與其上一期的變數值發生某一種延續性的關係。這樣的結構可將其表達成函數

的型式並且用遞迴方式推算出未來任一期的變數值。在模型中加入誤差項也符合經濟長期均衡的意義。

因為 Multidimensional Lognormal Random Walks 模擬機制的特性，正好吻合 AR(1)的特性，所以我們用 Autoregressive Model 自我迴歸模型來做初步分析，解決時序資料非定態問題。

VAR(1)模型的表達式：

$$r_t = \varphi_0 + \Phi_1 r_{t-1} + a_t \quad (5)$$

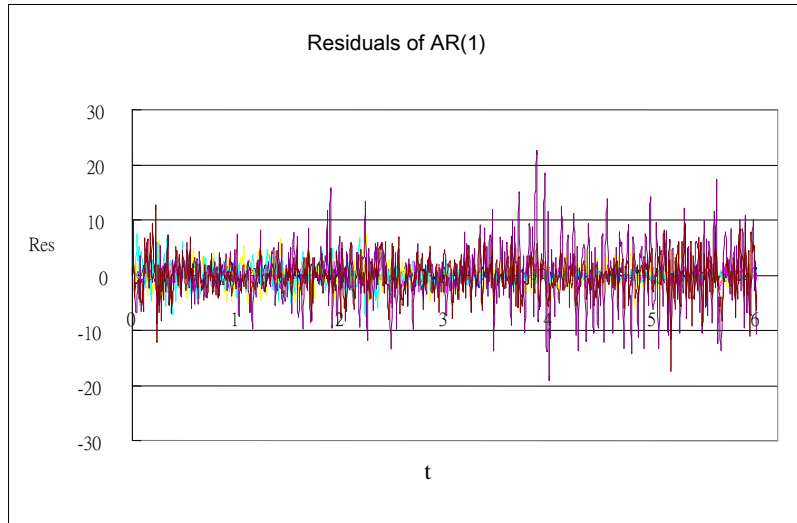
其中 r_t 指多變量的時間序列資料， φ_0 為 VAR(1)模型的截距項， Φ_1 指自我相關具一階關係時之參數。將模擬資料代入 VAR(1)模型中，求出 $\hat{\varphi}_0$ 及 $\hat{\Phi}_1$ 後，再求出 \hat{a}_t (residual) 並且計算這些殘差間的關連性，也就是 Correlation Matrix，可得下表二的結果：

表二 根據殘差 \hat{a}_t 計算變異及共變異數矩陣

Correlation Matrix of residual=

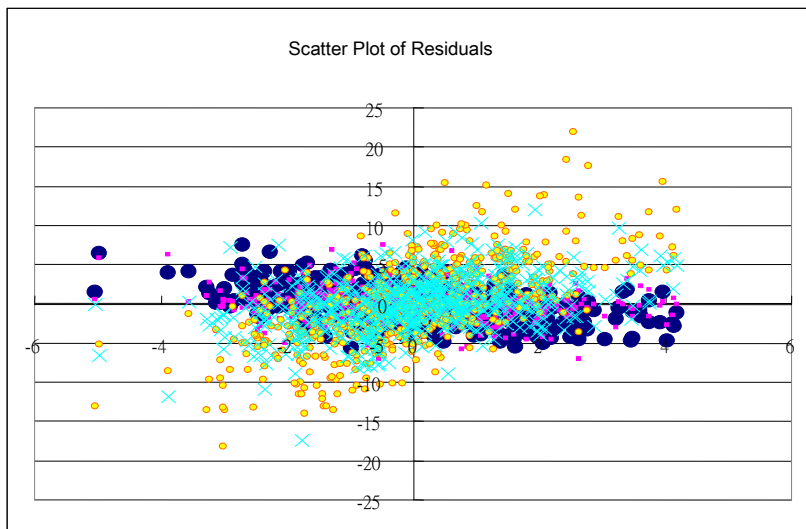
1.00000	0.35151	-0.42368	-0.17716	-0.07040	-0.51513
0.35151	1.00000	-0.46573	-0.14057	0.62738	0.37501
-0.42368	-0.46573	1.00000	0.37997	-0.00110	-0.15342
-0.17716	-0.14057	0.37997	1.00000	-0.33331	0.02475
-0.07040	0.62738	-0.00110	-0.33331	1.00000	0.43820
-0.51513	0.37501	-0.15342	0.02475	0.43820	1.00000

由表二變異及共變異數矩陣的上三角或下三角部分即伴隨效應 (concurrent effect)，我們畫出模擬資料的殘差序列圖，如下圖四所示：

圖四 殘差 \hat{a}_t 之序列圖

我們觀察此殘差序列圖，很明顯其變異大致介於正負 10 之間，有趣的是可以看到波動性小的殘差被波動性大的所遮蔽。此殘差符合因素分析的假設 (weak stationary)，拿這些殘差做因素分析意義較強烈，意義上不會被扭曲，所以因素分析只能取到伴隨效應，不應該將它與自我相關及共整合現象混淆不清。

接下來，我們畫出殘差 \hat{a}_t 兩兩之間的散佈圖，由此圖上我們觀察殘差之間的關聯結構，約略看出圓點“○”資料之弱正相關程度最高，符號“×”資料次之。而圓點“●”資料略成弱的負相關。

圖五 殘差 \hat{a}_t 二維散佈圖

接著我們運算這個相關係數矩陣的特徵值如下：

表三 共變異矩陣計算出特徵值
Engenvalues of Variance and Covariance Matrix =

$$\begin{bmatrix} 2.255 & 1.755 & 0.906 & 0.820 & 0.151 & 0.113 \end{bmatrix}$$

由表三共變異矩陣計算出的特徵值可知，大於 1 的特徵值有 2 個，表示背後大約有 2 個共同因子。而不是表一所看到的 4 個特徵值的結果，所以從這樣分析的過程可知，怎麼樣能明確區分出在長期的關聯性中有那些影響的因素、短期關連性中又有那些影響因素或者另外有我們不知道的影響因素呢？

五、共整合與誤差修正模型

討論多變量時間序列長期的連動關係，借由 ECM (error-correction model) 模型來模擬財務上時間序列資料，此種模擬機制所產生的資料具有共整合效應、伴隨效應及 AR 效應的特色。以下是它產生的機制：

$$\Delta x_t = \Pi x_{t-1} + \Phi_1^* x_{t-1} + a_t \quad (6)$$

其中 Π 表示共整合效應， Φ_1^* 表示 AR 效應， a_t 之間的關聯性則為伴隨效應。而用差分資料 Δx 來表達數列平穩的結構。在經濟上的意義，共整合常被詮釋為『經濟變數間具有長期均衡關係』，當變數具有共整合關係時，隱含這些變數長期而言，具有往『均衡方向調整』的特性。即在短期時，變數間存在偏離長期均衡的現象，也會逐漸縮小，就是所謂的誤差修正機能（楊奕農，2005）。

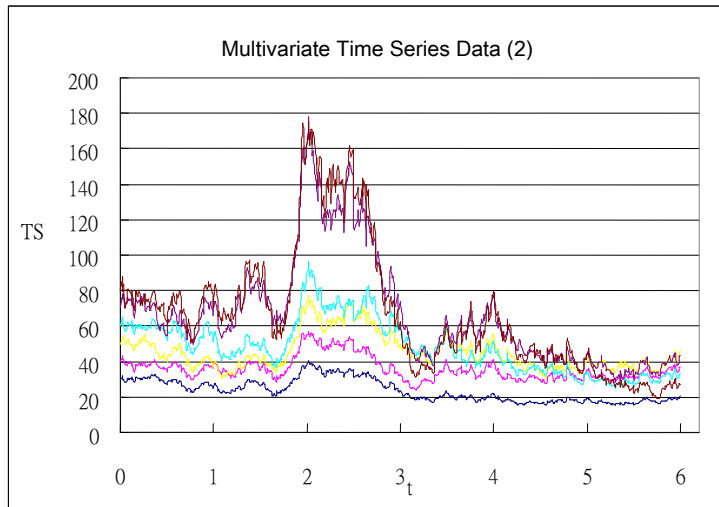
在這裡由另一筆模擬資料來表現六檔股票的股價行為，設定參數為

$$\Pi_{6 \times 6} = \begin{bmatrix} 0.3 & 0.2 & 0.1 & 0.3 & 0.3 & 0.1 \\ 0.4 & 0.3 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.1 & 0.1 & 0.3 & 0.1 & 0.1 & 0.3 \\ 0.2 & 0.2 & 0.6 & 0.1 & 0.1 & 0.6 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.45 & 0.45 \end{bmatrix}, \text{ 所以 } Rank(\Pi) = 4, \Phi_1^* = \mathbf{I}_{6 \times 6},$$

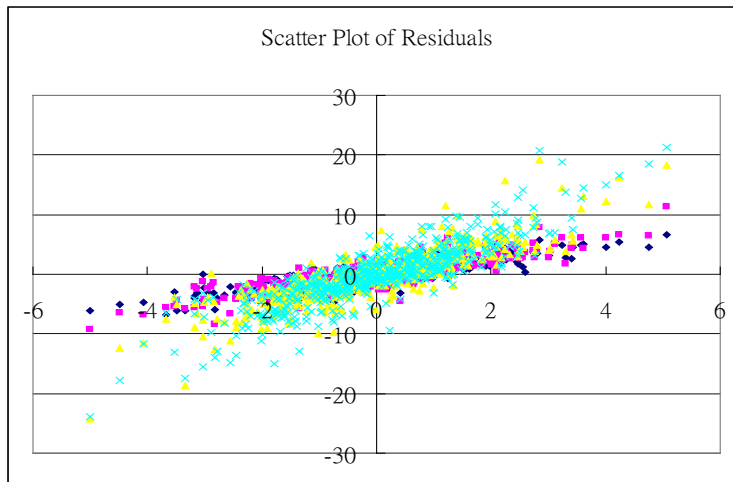
$$a_t = \begin{bmatrix} 1 & 0.4 & \cdots & 0.4 \\ 0.4 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0.4 \\ 0.4 & \cdots & 0.4 & 1 \end{bmatrix}_{6 \times 6}, \text{ 起始股價分別爲 30、40、50、60、70 及 80,}$$

繪出股價之時間序列圖如圖六所示：

從整個時間軸來看，六檔股票的股價波動有明顯的長期連動趨勢。從圖形上觀察，可以推論共整合的效應可能佔掉大部分可解釋的變異，而剩下的變異就由殘差來解釋。另外，透過時序資料做一次差分後的行為即可說明共整合關係不容易透過差分的效用而消除。



圖六 模擬六檔股票股價的時間序列圖



圖七 殘差二維散佈圖

從殘差兩兩散佈圖，我們可以觀察殘差之間的關聯結構，很明顯看出所有資料都呈現正相關的訊息，“×”符號資料正相關程度最高，“▲”符號次之。經過以上殘差間關聯性的探討，的確可說明資料經過 ECM 的調整，不但解決了資料不平穩的問題，移除長期的連動關係，更孤立出變數間短期的關連結構，為因素分析延伸 ECM 模型，將 Δx_i 視為反應變數，共整合效應為水準 A，AR(1)效應為水準 B 的 Two-Way 模型，假設 A、B 之間無交互作用，則 sums of squares 可表達為 $SST = SS_{\text{共整合}} + SS_{\text{我相關殘差}} + SSE$ ，運用矩陣 Trace 的特性，可計算出各效用佔整體的比重。

若用線性迴歸模型來表達這樣的想法，從 Multiple Regression 出發，假設模型：

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad i=1,2,\dots,n \quad (7)$$

滿足 $\varepsilon_i \sim N(0, \sigma^2)$ ，那麼用 Sum of squares 表達自變數及誤差占整體變異的解釋比例應是很適當的。將(7)化成矩陣型式：

$$Y_c = X \beta + \varepsilon \quad (8)$$

$\begin{matrix} n \times 1 & n \times 3 & 3 \times 1 & n \times 1 \end{matrix}$

其中 Y_c 為對平均數平移後的資料，其被解釋變數轉換成 Y_c 是為後面 EM 演算法的需求作準備，可使 EM 演算法收斂的動作，它並不會影響 SST、SSR 及 SSE 的值。將 SSE 表達為 $(Y - \hat{Y})(Y - \hat{Y})$ ，根據最小平方方法 Min SSE 可得 $\hat{\beta} = (X'X)^{-1} X'Y_c$ ，則可求出最佳的直線迴歸方程式 $\hat{Y} = X\hat{\beta}$ ，則 $SST = SSR + SSE$ ，其中：

$$(Y - \bar{Y})'(Y - \bar{Y}) = (Y - \hat{Y})'(Y - \hat{Y}) + (Y - \bar{Y})'(Y - \bar{Y}) \quad (9)$$

$\begin{matrix} (SST) & (SSE) & (SSR) \end{matrix}$

擴展成多變量複式迴歸 (Multivariate Linear Model)，以兩個反應變數為例，可表達為：

$$Y_c = X B + \Xi \quad (10)$$

$\begin{matrix} n \times 2 & n \times m & m \times 2 & n \times 2 \end{matrix}$

延申至有 p 個反應變數與 m 個解釋變數的一般式為：

$$Y_c = X B + \Xi \quad (11)$$

$\begin{matrix} n \times p & n \times m & m \times p & n \times p \end{matrix}$

將矩陣 $SST = SSR + SSE$ 分解後，利用矩陣 Trace 的特性 $tr(SST) = tr(SSR) + tr(SSE)$ ，其中由 $tr(SSR) / tr(SST)$ 來表達共整合及 AR(1) 效應占整體變異的解釋比例以及 $tr(SSE) / tr(SST)$ 為殘差部分占整體變異的解釋比例。由多變量複式迴歸的一般情況，可將上述表達延伸為 $SST = SSRA + SSRB + SSRC + \dots + SSE$ ，其中的 A、B、C... 代表多變量複式迴歸中具有 A、B、C... 組反應變數。當然，在矩陣分解後也可計算 $tr(SST) = tr(SSRA) + tr(SSRB) + tr(SSRC) + \dots + tr(SSE)$ 並且表達各個變數占整體變異可以解釋的比例。我們運用因素分析的特性，接續將殘差部分再做拆解。隱藏在殘差之中的訊息是接下來討論的重點。

接下來，我們表達共整合效應在整體變異所佔的百分比例，如下表四所示：

表四 模擬共整合效應的影響比例表

Effect	Trace	共整合效應	
		值	比例
trace(SSR)		14672.6746	69.70%
trace(SSE)		6379.2111	30.30%
trace(SST)		21051.8857	100%

由表四可知，此模擬資料可由共整合效應解釋的成份占整體變異的 69.7%，剩餘的殘差解釋能力有 30.3%。而這 30.3% 的變異就是指短期的伴隨效應所造成的影響。接下來我們觀察式子(6)中， a_t 的部分即指 ECM 模型中的殘差。

六、殘差因素分析的參數估計

參考 Tsay (2005) 的見解，我們拿殘差來做因素分析。此時，將式子(2)視為多變量反應線性模式。本文運用 EM 演算法 (Expectation-Maximization Algorithm) 於因素分析的參數估計上。EM 演算估計方法在於利用可觀測到的資料來估計其它未知的部份，主要可為二個步驟，一為 E 步驟 (expectation step)，另一為 M 步驟 (maximization step)，並且反覆操作 E 步驟及 M 步驟讓估計參數的過程能不斷修正直到結果合於穩定收斂的狀態。因此 EM 演算技術被廣泛運用於不完整資料的估計上。接下來說明因素分析的參數估計過程，首先假設 B 為因素負荷 (Factor loadings)，F 為因素分數 (Factor Scores)，先以光

譜分解 (spectral Decomposition) 估計 B ，如果知道 B (Loadings) 則 Factor Scores 的期望值為：

$$E(F) = YQB'(BQB')^{-1} \quad (12)$$

其中

$$Q = [Diag(S - B'B)]^{-1} \quad (13)$$

上述(12)(13)式為 EM 演算法 E 的步驟，此乃根據 Harman (1976) 的結果，其(13)式之 Q 表達權重的意義，也就是 Weighted Least Squares。若知道 F (Factor Scores) 則 B 的 Maximum Likelihood Estimates (under Normal) 的估計為

$$B = (F'F)^{-1} F'Y_c \quad (14)$$

此公式(14)為 EM 演算法 M 的步驟。這樣的過程可同時滿足主成份因素法及 MLE。這樣的演算法同時獲得因素負荷與因素分數 (factor scores)，同時也達到最大概似估計值。

由 Principal Component Method 所得 factor loading 當作起始值，根據式子(1)可直覺的猜想，若將變數 Y 、 F 固定，即假設 Y 、 F 已知，則可求得 $B = (F'F)^{-1} F'Y_c$ 。若變數 Y 、 B 固定，我們將(2)式轉置： $Y_c' = B'F' + E'$ ，用相同角度解 F' ，則 $F' = (BB')^{-1} BY'$ ，整理後可得 $F = YB'(BB')^{-1}$ ，而這樣思考也可在 Roweis (1998) 中看到類似的運算。再加入權重後可表示為： $F = YQB'(BQB')^{-1}$ ，本研究採用 Bartlett (1937, 1938) 所得結果，在因素負荷的估算上，Bartlett 建議用 Weighted Least Squares 計算參數，由式子(1)， $\varepsilon' = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p]$ 為殘差， $Var(\varepsilon_i) = \Psi_i, i = 1, 2, \dots, p$ ，以 $Var(\varepsilon_i)$ 的倒數為權重，則 SSE 為

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\Psi_i} = \varepsilon' \Psi^{-1} \varepsilon = (y - \mu - \Lambda f)' \Psi^{-1} (y - \mu - \Lambda f) \quad (15)$$

求上式(15)之極小，可得 $F = Y \hat{\Psi}^{-1} \hat{\Lambda} (\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda})^{-1}$ ，即為公式(12)。

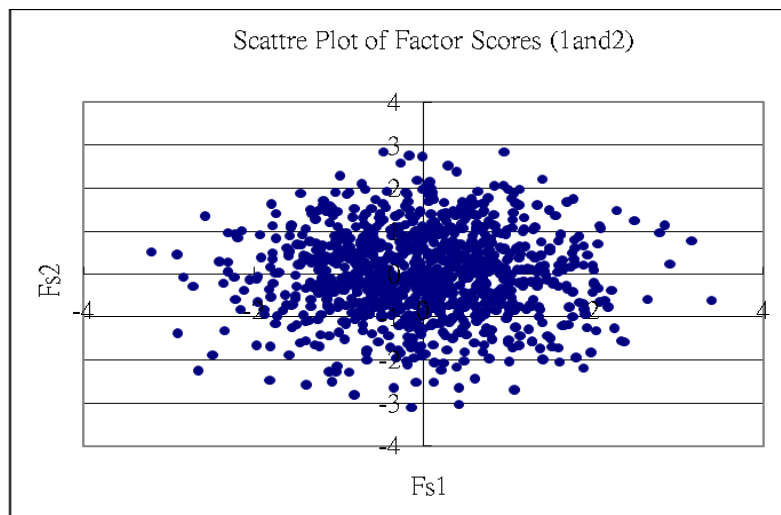
簡單用數學式來表達 EM 演算法運用的過程，首先因素分數 F 為未觀測到的訊息，給定一起始值 B_o 為 PC 所得的因素負荷，並以上標表第 i 次代疊，則 \hat{B}^i 為第 i 次代疊所得因素負荷值，用以下兩式交互代疊， $i = 0, 1, 2, \dots$

$$F^{i+1} \leftarrow E[F | B_o, \hat{B}^i] \quad (16)$$

$$\hat{B}^{i+1} \leftarrow \arg[\max_B L(B; B_o, F^{i+1})] \quad (17)$$

以(16)式將未觀測到的因素分數 F 以條件期望值進行估計，再將所得的估計值，代入概似函數中並求出極大(17)式之 \hat{B} 值，將(17)式中所估的 \hat{B} 值代回(16)式，求算出下一回代疊的估計值 F^{i+2} ，反覆進行此動作，當 $i \rightarrow \infty$ 時，這一串參數的估計值 $(\hat{B}^\infty, B^1, B^2, \dots)$ 將收斂至 B 的一個最大概似估計值， \hat{B}_{MLE} 。此處的「 \leftarrow 」表示以右邊的運算過程之結果給予左邊的隨機變數一個值，而「 \leftarrow 」表示使右邊極大的數值賦予左邊。

EM 演算法是屬於定值收斂，可求出點估計值，不同於 EM 演算法衍伸出來的蒙地卡羅馬可夫鏈方法 (MCMC, Markov chain Monte Carlo)，其過程是於隨機收斂。MCMC 也會在隨後的因素轉軸中運用到。接下來，我們實際驗證模擬資料是否可執行因素分析，以下使用另一筆模擬 10 檔時間序列資料做因素分析，當然先由前述的 ECM 模型取得殘差。首先，對此因素模型做偵斷，判斷殘差是否符合模型的基本假設，如下圖八所示：



圖八 Factor score 1 與 2 的散佈圖

我們將兩個因素分數散佈在二度空間上，從圖八可明顯看出 Factor score1 與 Factor score2 獨立，散佈點大約介於 (3, -3) 之間，像似二元常態分佈，符合 Factor score 的假設。接著我們計算 Y 的共變異矩陣 S ，並嘗試與殘差矩陣比較收斂的情況。我們利用 $\hat{\Psi} \cong S - \Lambda'\Lambda$ 計算出殘差矩陣，可比較 S 與 $\hat{\Psi}$ 的上三角矩陣部分，藉由表五的結果觀察殘差矩陣的數值的確比共變異矩陣的數值小很多，且許多數值都已收斂到 0，未收斂到 0 的數值部分相對來說也小了許多。這說明了此次資料分析結果，的確達到因素分析的意義。

表五 因素分析的殘差矩陣 $\hat{\Psi}$

Residual Matrix=

0.29	-0.11	-0.09	-0.10	-0.01	0.00	0.00	0.02	0.01	-0.02
-0.11	0.30	-0.10	-0.09	0.00	0.01	0.02	0.00	-0.03	-0.01
-0.09	-0.10	0.29	-0.10	-0.03	0.01	0.00	0.00	0.02	0.00
-0.10	-0.09	-0.10	0.29	-0.02	0.00	0.00	0.01	0.00	0.02
-0.01	0.00	-0.03	-0.02	1.29	-0.22	-0.23	-0.31	-0.20	-0.32
0.00	0.01	0.01	0.00	-0.22	0.13	0.01	0.31	-0.02	-0.20
0.00	0.02	0.00	0.00	-0.23	0.01	0.83	-0.05	-0.74	0.20
0.02	0.00	0.00	0.01	-0.31	0.31	-0.05	0.76	-0.07	-0.60
0.01	-0.03	0.02	0.00	-0.20	-0.02	-0.74	-0.07	0.87	0.14
-0.02	-0.01	0.00	0.02	-0.32	-0.20	0.20	-0.60	0.14	0.77

七、因素分析的轉軸問題

在因素分析中，另外一個嚴重的“不確定”問題就是在因素負荷估算出來後，可以有不同的轉軸方法，都不會影響到這個模型的正確性，因而造成無所適從的結果。轉軸方法又包括變異最大旋轉法 (Varimax Rotation)、四方最大法 (Quartimax Rotation) 及 Procrustes Rotation Problem，而套裝軟體從未交待這些 Optimization 的演算法。在轉軸技術中，提出馬可夫鏈蒙地卡羅法，利用其隨機尋找的過程，發現這種應用技術相當可行。

馬可夫鏈蒙地卡羅法最早是應用在貝氏統計學上的演算技術，本研究將這方法帶到古典統計學中，使得統計學中這兩大派別，有其相似之共通處，雖然他們在原始的認定上是截然不同的。這種異中求同的方式，在傳統與貝氏的選擇上，有助於消除其之間的困擾。至於這種隨機尋找的演算邏輯，在 Spall (2003) 有完整的說明。

首先，因素分析的轉軸問題要如何獲得唯一解呢？我們列出共變矩陣與位移之後的資料矩陣，如何表達在不同的正交轉軸上，為之下公式(18)(19)

$$\begin{aligned}
\Sigma &= \Lambda \Lambda' + \Psi \\
&= \Lambda P P' \Lambda' + \Psi \\
&= \Lambda^* \Lambda^{*'} + \Psi
\end{aligned} \tag{18}$$

$$\begin{aligned}
Y_c &= F \Lambda' + E \\
&= F P P' \Lambda' + E \\
&= F^* \Lambda^{*'} + E
\end{aligned} \tag{19}$$

其中 P 為正交轉軸矩陣，轉軸後 Σ 及 Y_c 完全不變。在(18)式中， Λ 表示經過因素分析後所選擇的因素負荷 (factor loading) 矩陣，將共變異矩陣 Σ 表達成由因素負荷吸收掉的部分加上殘差矩陣 Ψ ，由 Ψ 表達剩餘無法由因素負荷解釋的部分。轉軸問題所引發的不確定性可以從兩個觀點去看待，一個是透過正交轉軸後共變矩陣 Σ 並沒有改變，另一個觀點是觀測到的 Y_c 矩陣也沒有改變。這就是說 Λ 決定之後，還要決定正交轉軸矩陣 P ，才能將這結果圓滿化。

接著，我們界定轉軸的目標函數，

$$OBJ(P | \Lambda, Y_c) \tag{20}$$

$P_{m \times m}$ 的維度為 $m(m-1)/2$ ，參考 Tanner and Thisted (1982)。要討論轉軸問題，我們可以先固定 Λ 和 Y_c ，並界定上述的 OBJ 函數，其輸入為 $m(m-1)/2$ 維度的 P 轉軸矩陣，而輸出則以不同的準則界定之。Varimax 和 Quartimax 準則為大家所熟悉的，另外的 Procrustes Rotation，參考 Mulaik (1972)。其目標為面對兩個不同的 Λ_1 和 Λ_2 透過轉軸讓它們最為接近。當我們使用 Varimax 準則來處理轉軸問題時， OBJ 函數為 ΛP 之行變異總合，當使用 Quartimax 準則時， OBJ 函數為 ΛP 之列變異總合，另外當使用 Procrustes Rotation，則 OBJ 函數為 $trace(\Lambda_1 - \Lambda_2 P)'(\Lambda_1 - \Lambda_2 P)$ 。

運用馬可夫鏈蒙地卡羅法 (MCMC, Markov chain Monte Carlo) 進行隨機尋找 (Random Search) 目標函數的極值。利用 MCMC 的技術隨機尋找極值，此法視為 EM 演算法 (Expectation-Maximization Algorithm) 的延伸，我們的目的是要找出第(20)式的極值，也就是在數值上操作如下第(21)及(22)式

$$\max_P OBJ(P | \Lambda, Y_c) \tag{21}$$

or

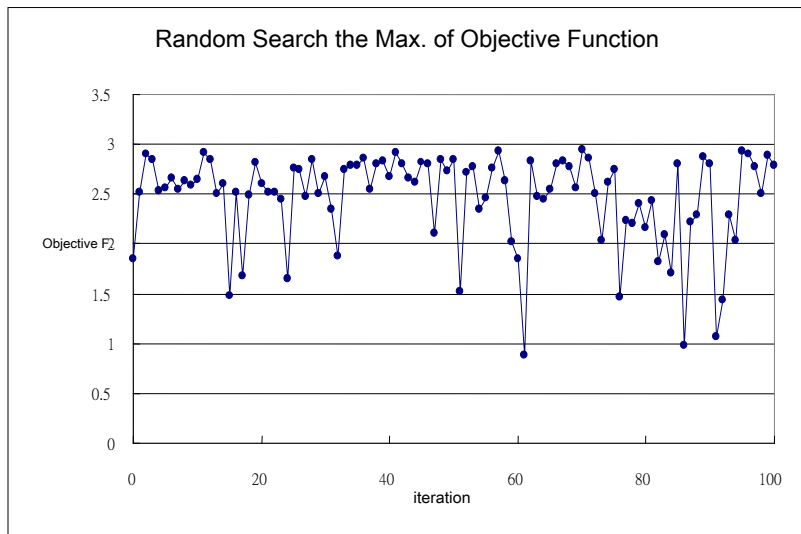
$$\min_p OBJ(P|\Lambda, Y_c) \quad (22)$$

運算方法我們用 MCMC 產生在 $m(m-1)/2$ 維度的隨機正交矩陣 P ，代入 OBJ 函數，逐步求得極值。馬可夫鏈蒙地卡羅法的原因事實上是源自於 EM 演算法，MCMC 只是將 EM 演算法的定值收斂改變成隨機收斂，去獲得某一個機率分佈的大量觀測向量。把這樣的想法當成一個隨機尋找的過程。我們的目的是要求得目標函數 OBJ 的極值，只是將輸入的正交矩陣 P 做成有效率的隨機值輸入，並運用電腦大量運算的能力去求得極值。

操作方法為在 m 個座標中，每次隨機挑 2 個軸做轉軸，旋轉一次當作一次代疊，而隨機轉軸角度為 θ 。也就是之下的(23)代疊式

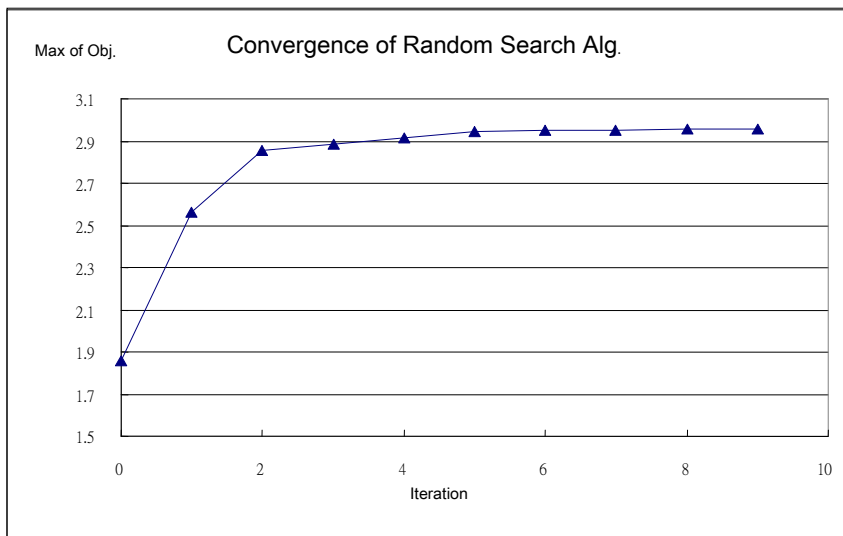
$$P_{i+1} = P_i * \begin{pmatrix} \cdot & \cdots & \cdots & \cdot \\ \vdots & \cos \theta & \cdots & \sin \theta \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & -\sin \theta & \cdots & \cos \theta \end{pmatrix} \quad (23)$$

根據 MCMC 的性質，此馬可夫鏈的極限分佈為 $m(m-1)/2$ 維度上均勻分佈的隨機正交矩陣，將每次代疊的 P 代入 OBJ 函數。MCMC 的方法雖然原創者是為了給貝氏統計學更方便的演算法，但是這套技術絕對不是只能用在貝氏統計學，例如這個隨機正交矩陣的大量產生，就可以利用 MCMC 所產生的極限分佈去獲得，因為每次只轉 2 個軸，成爲一個很有效率的演算法。比起獨立產生隨機正交矩陣，在電腦運算上節省相當多的時間。下圖九爲我們模擬隨機尋找極值的過程圖，橫軸爲第 i 次代疊，縱軸爲目標函數值



圖九 隨機尋找極值的過程圖

雖然在 $m(m-1)/2$ 的維度上是均勻分佈，但是經過 OBJ 函數轉到 1 度空間時，是“集中火力”在 OBJ 較高的值。上面這個圖九是利用前述 MCMC 的每次移轉後的正交矩陣代入目標函數，做成時間序列圖，我們可以清楚地看出這是一個非獨立式的抽樣方法，產生的目標函數值存在有自我相關，但是又較集中在 OBJ 較高的值，在尋找極大值的現象上，結果是令人滿意的。



圖十 隨機尋找的收斂圖

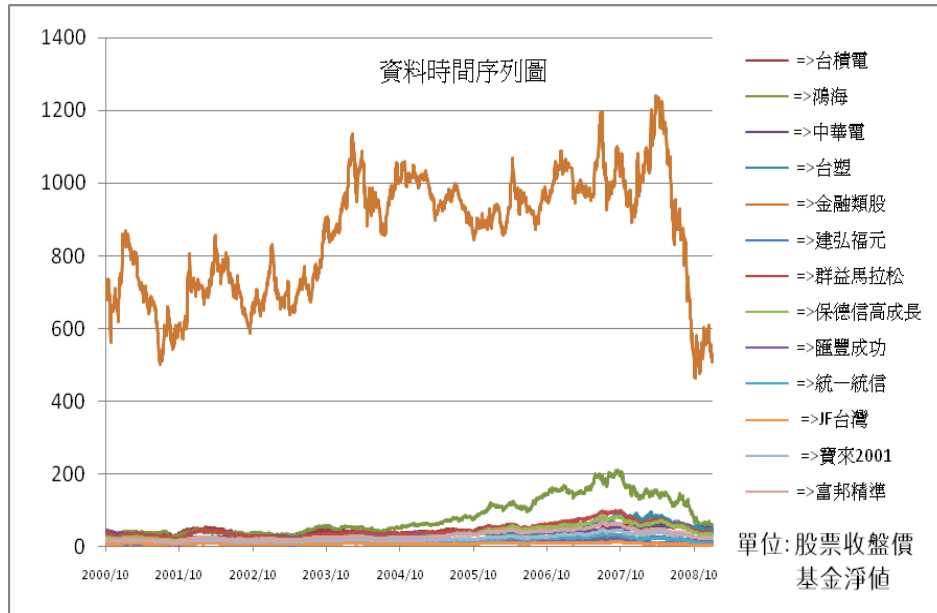
圖十所示，我們接著再定義另一個隨機過程為以之前的過程中所找到的最大之目標函數，則我們可以清楚地看到它收斂到約 2.94。而且對任何目標函數而言，這樣的演算法永遠行得通。

肆·實證結果分析

一、實證資料的介紹

資料來源為臺灣經濟新報資料庫，選擇 2009 年 2 月 5 日台灣 50 成份股中所占比重最高的四檔股票，包括台積電、鴻海、中華電信及台塑及金融類股股票之除權息後的收盤價。另外，由臺灣經濟新報資料庫股票型基金評等資料，選擇自成立以來至 2009 年 1 月之報酬率排名最高的八檔開放式一般股票型基金，包括建弘福元、群益馬拉松、保德信高成長、匯豐成功、統一統信、JF 台灣、寶來 2001 及富邦精準之基金淨值，資料期間為 2000 年十月至 2009 年一月共 2047 筆日資料。為什麼選擇這些變數來討論股票與基金市場？理由很直接，就是針對我們有興趣的標的做選入。

當然，若要討論更全面的市場結構，包含的變數也需要具全面性。由於因素分析本身的機制可以消化訊息雜亂的問題，所以不用擔心放入因素模型的資料過於複雜。因素分析的體制就是根據投入的變數，捉出其變數之間看不見的影響因子。我們將這 13 檔金融市場資料畫成時間序列圖，觀察其走勢。初步對資料的觀察可以看到其中 12 檔金融資料，概略呈現一致的波動，相對來說，金融類股是波動劇烈的。

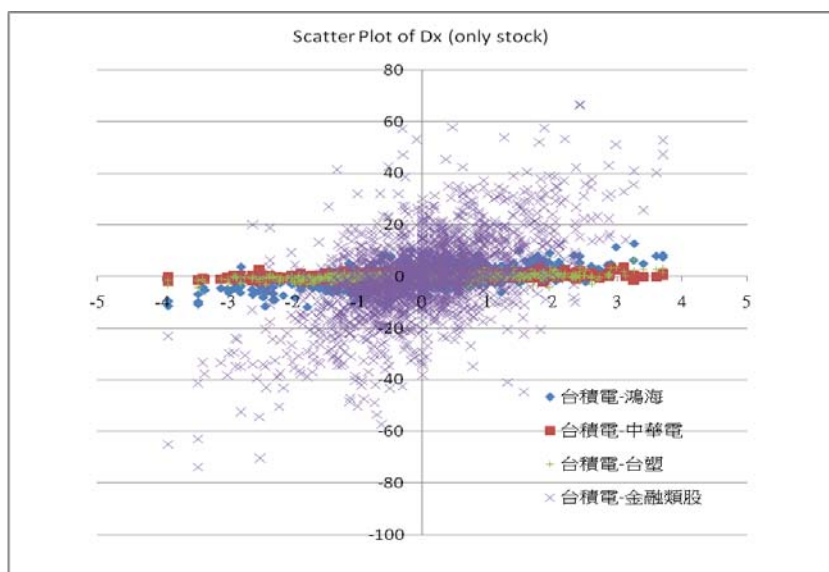


圖十一 實證資料的時間序列圖

將這 13 檔時間資料求相關係數矩陣及其特徵值，結果第一個特徵值佔了大部分的資訊有 10.672，第二個快速落到 1.315 接下來的差距都 0.3 之內，所得到的訊息其實並不明確，為什麼這樣說呢？我們直覺假設事件的波動主要受長期、短期及伴隨效應或者其它未知因素所影響，如果要將整體波動的因素分成這幾部分來觀察，那麼重要因素至少也有 2 個以上。從時序圖上看到基金的淨值波動都相當的一致，只有金融類股及鴻海二檔股票波動特別大，站在觀察者的角度會認為長期穩定的因子是主要的影響成份，將資料切成小的時間區段來看彼此間的關係更是強烈，由此可知這樣的結果並未將資料的訊息真正表現出來。

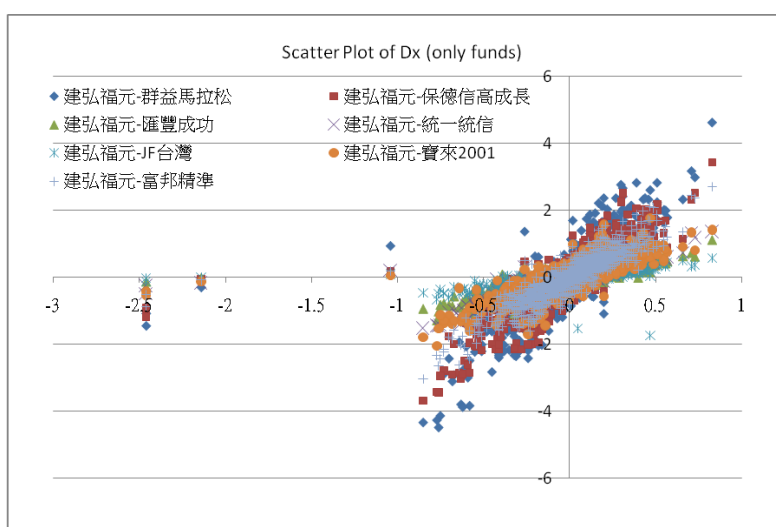
二、實證資料分析

嘗試將原始資料取一次差分，拿差分轉換的訊息來看股票及基金之間的相關性。圖十二表示 5 檔股票間有弱正向的關聯性，其中變異最大的金融類股為圖上散佈面積最大的數列，以符號“×”表示。



圖十二 股票間差分後相關散佈圖

圖十三顯示 8 檔基金間有高度正向的相關性，其中在 橫軸 2.5 處有明顯整齊的離群值出現，應是某個特定時間點發生特定事件所導致。已知在股票及基金中本身有關聯性，本研究更想了解兩者相互關係及影響的程度。所以嘗試將其中的關係一個一個抽出。

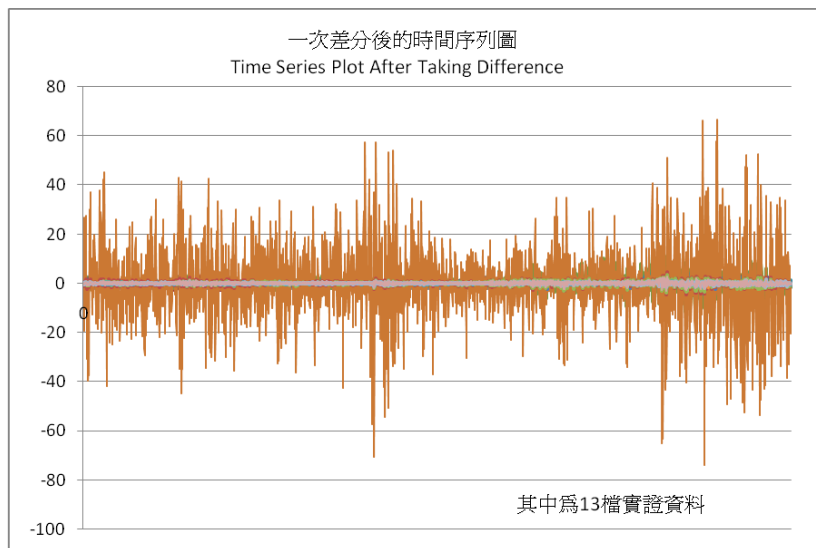


圖十三 基金間差分後相關散佈圖

根據 ECM (error-correction model) 模型產生的機制 $\Delta x_t = \alpha(\beta - x_{t-1}) + a_t$ ，將其中共整合效應及 AR 效應分別抽出。在資料中究竟是 AR 效應較強還是共整合效應較強烈呢？先做以下的思考，假設 AR

效應對資料的影響較為強烈，則可將 $\Delta \Phi \Delta x_t = \alpha_1 x_{t-1} + a_t$ 視為原模型的縮減模式，而完整的模型為 $\Delta \Phi \Delta x_t = \alpha_1 x_{t-1} + \Pi x_{t-1} + a_t$ ，若共整合效應為主要成份，則縮減模型表達為 $\Delta x_t = \Pi x_{t-1} + a_t$ 。

將 13 檔時間序列資料取一次差分後作時間序列圖，如圖十四所示。很明顯的，其中的 12 筆資料經過一次差分後已呈現穩定的狀態，這同時說明資料中受到共整合效應及 AR(1)的影響是很微弱的。若拿具共整合效用資料與此筆資料比較可知，資料經過差分後傳達相當不同的訊息。



圖十四 實證資料取一次差分後之時間序列圖

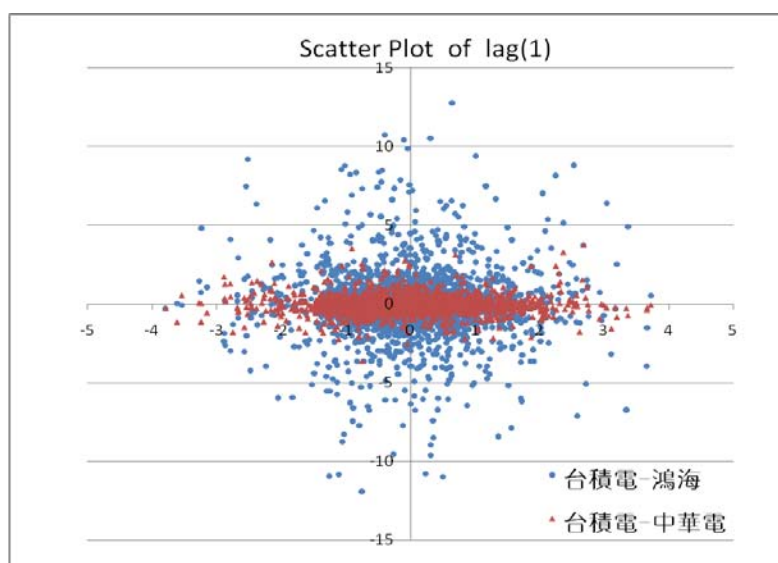
我們想知道這些影響的效果究竟分掉了整體多少的比例，從表六可看到一個有趣的結果：

表六 共整合效應及 AR(1)的影響比例表

Effect \ Trace	先加入共整合效應		加入共整合及 AR(1)效應後	
	值	比例	值	比例
trace(SSR)	6958.0639897480	1.42%	9123.8619260378	1.86%
trace(SSE)	484729.9344087390	98.58%	482216.2736057330	98.14%
trace(SST)	491687.9983984830	100%	491340.1355317670	100%

我們先帶入共整合的效果。如果將整體波動的影響成分，分成幾個部分解釋，共整合效應只佔整體的 1.42%，在同時加入共整合及 AR(1)效應後，總共也只佔整體的 1.86%。顯示實際上資料受到此二者效果的影響非常的微小。

將近 98% 的解釋部分由 SSE 所表示，究竟在其中隱含了什麼樣的波動因子，正是本文要探討的重點。



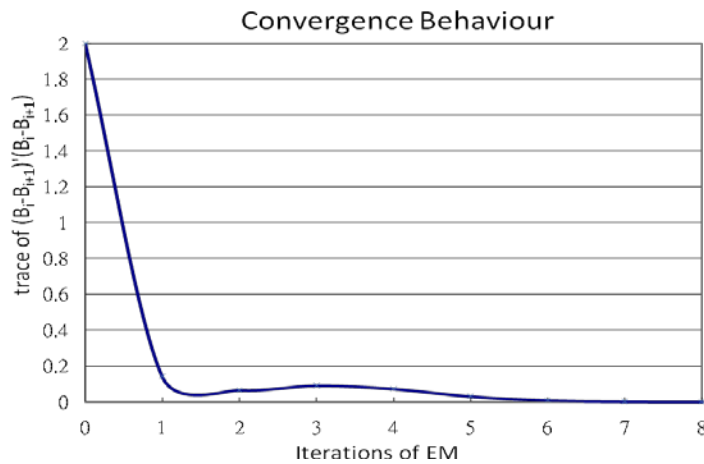
圖十五 Cross-Correlation 落後期數 1 之散佈圖

在全球化及資訊快速流通的市場，價格的變動行為是直接且易於散佈的，了解市場間動態的關連結構，對於決策者是相當重要的資訊。在時間序列資料中，時間點與點之間的相關性是重要的訊息，因此，我們討論 Cross-Correlation 時間落差一單位，在此實證資料中佔了多少的影響成份，資料是一次差分後的資料，由上圖十五將台積電時間 t 點分別對鴻海及中華電信時間 $t+1$ 點做散佈圖，圖上兩兩並沒有呈現任何趨勢，初步可知此資料沒有時間落差的關係。若資料具有一期落差的關係，就應再討論有無落差二期或三期以上的關係。這是因為單獨看時間點 t 與時間點 $t+l$ 間的相關性，會隨著 l 變大，點與點之間的相關性或變異會愈來愈小。

由上例可知，在台灣的股票市場，鴻海或者中華電信的股價波動受台積電股價影響非常小，甚至沒有所謂的價格延續性，這在穩定成熟的經濟市場中是個奇特的現象，也意味著投資者做短期操作時，無法參考同業及大環境所給的即時訊息判斷未來的狀況。這時再檢視表七的結果，是吻合資料所給的訊息的。

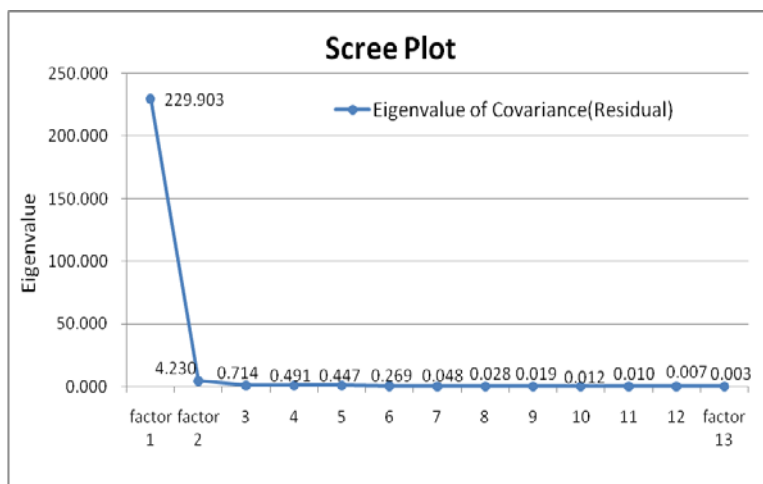
實證資料透過 ECM，抽離 AR 及共整合效應，剩下將殘差拿來做因素分析。下圖為實際操作 EM 演算法，因素負荷矩陣收斂的表示圖，本研究將疊代

次數設定為 20 次，由圖十六所示，橫軸為代疊次數，縱軸為 $\text{trace}(B_i - B_{i+1})'(B_i - B_{i+1})$ ， $i = 1, 2, \dots, 20$ ，其中 B_i 為第 i 次代疊的因素負荷矩陣，結果是收斂的，代疊到 7 次左右，因素負荷矩陣就已穩定不再變動。



圖十六 EM 演算法估計參數的收斂行為

由 EM 演算法代疊收斂的結果，同時得到滿足 PC 及 MLE 法的因素負荷及因素分數矩陣，由上圖十六可知，EM 演算法成功的運用於因素分析的參數估計中，這樣收斂的結果是相當合理的。



圖十七 實證資料之因素陡坡圖

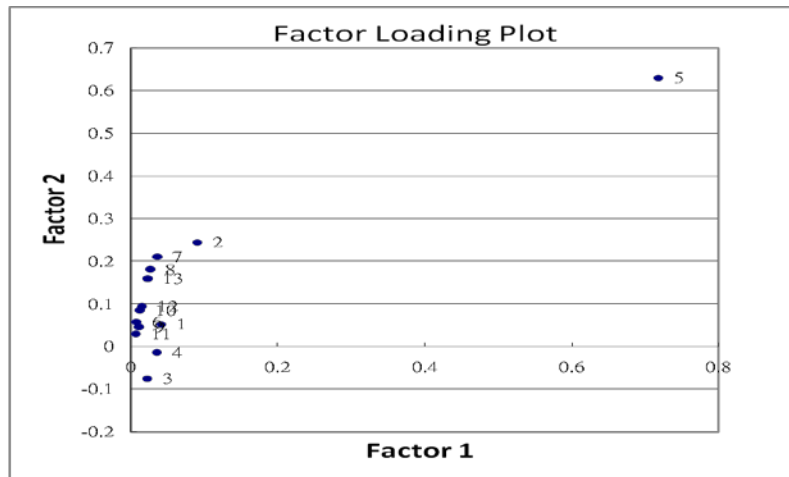
由因素陡坡圖可知，第一個因子由高點迅速掉落到低點，從第二個因子之後解釋能力不到整體變動的 1%，因此，取其前 2 個因子為此實證資料背後

波動的主要影響因素。第一因子的特徵值為 229.903，占整體的 97.34%，第二個因子的特徵值為 4.230，占整體的 1.79%。接下來，將殘差變異的解釋成份由因素分析的因子來呈現，在殘差的因素分析中，取出的 2 個因子，因子一可解釋殘差變異的 97.34%、因子二可解釋殘差變異的 1.79%，其餘因子僅占殘差變異的 0.87%。

表七 將殘差做因素分析後之影響比例表

Effect	Trace	加入共整合及 AR(1)效應後		將殘差做因素分析後	
		值	比例	成份	比例
trace(SSR)	9123.861926		1.86%	SSR	1.86%
trace(SSE)	482216.2736		98.14%	Factor 1	95.53%
				Factor 2	1.76%
				Factor 3~Factor 13	0.85%
trace(SST)	491340.1355		100%	SST	100%

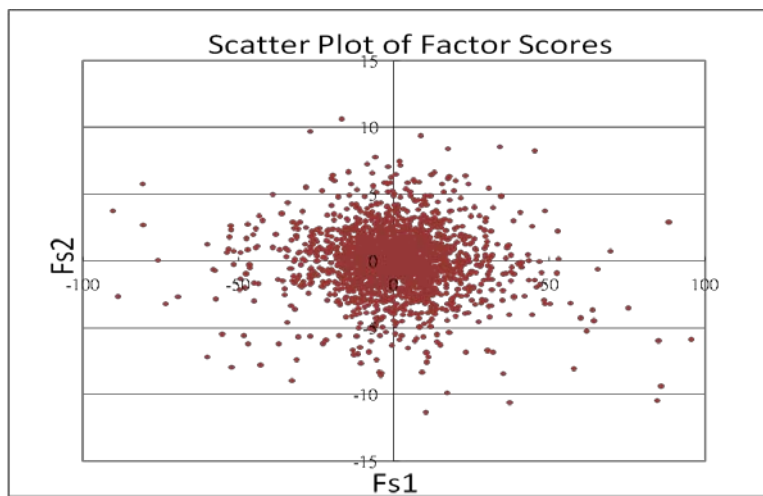
由表七對整體變異來做表示，可由因子一解釋整體變異的 95.53%、因子二解釋整體變異的 1.76%，其餘因子僅占整體變異的 0.85%。接下來，我們分析二個因子的特性，解釋波動的特殊關係。將因子一及因子二的因素負荷值畫散佈圖，由下圖十八轉軸前的因素負荷圖所示：



圖十八 轉軸前之因素負荷圖 (Factor 1 vs. Factor 2)

因素負荷圖中，共有十三的圓點，依序由數字 1~13 來表示第 1 到第 13 個變數，變數由 1 至 13 依序代表(1)台積電、(2)鴻海、(3)中華電信、(4)台塑及(5)金融類股股票。接下來為基金部分為(6)建弘福元、(7)群益馬拉松、(8)保德信高成長、(9)匯豐成功、(10)統一統信、(11)JF 台灣、(12)寶來 2001 及(13)富邦精準。在此因素負荷圖中，表示變數分別在因子一及因子二的影響重要成份，其中，金融類股股票對於二個因子都佔相當重要的影響比例。從因子一的特色來看，金融類股與其它十二檔資料的相對距離與值都是差距相當大的，且其餘十二檔資料在因子一中占相當小的比例，這表示金融類股的波動對整體的波動影響力是很大的，深具代表因子一的行為與波動特色，也就是所謂的伴隨效應，伴隨可以解釋為一個事件的發生會陪同另一現象的產生。

由因子二來觀察，第五個變數依然位居最高點，將近 0.65 個單位高，而其餘變數分佈於 0.3 至-0.1 個單位高，轉換原單位比例來看，因子二約為 60% 的整體表現受金融類股的影響，表示金融類股的波動依然具有相當的代表性，不同於因子一，約為 30% 的整體表現受其餘的股票及基金影響。所以因子的意義可以解釋為整體波動受金融類股的引導。當然，這樣的現象與基金、股票的整體組合有關。

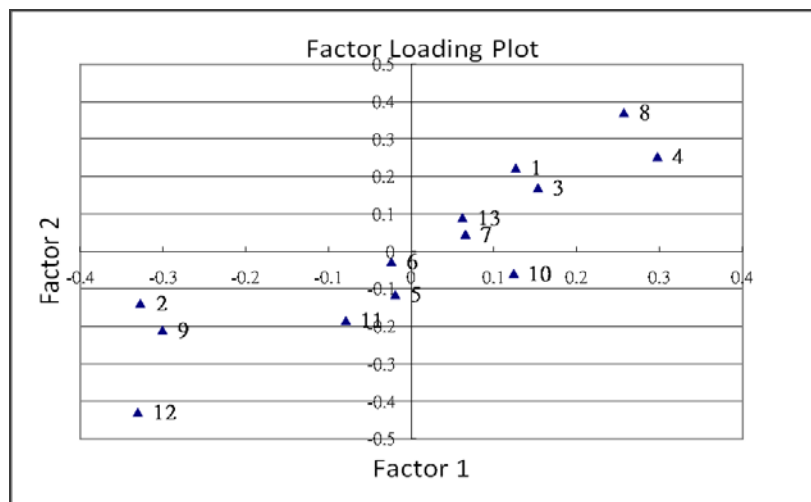


圖十九 Factor score 1 與 2 的散佈圖

在因素模型診斷部分，由圖十九，因子一與因子二的因素分數的散佈圖可知，實證資料做差分的處理，相當符合因素分析的基本假設。在未轉軸前，因素負荷圖雖然可明顯看出，金融類股與其它股票及基金變數自成一類，但對因素特色的解釋不太明朗，希望能借由轉軸，將變數間的特色表現出來。接下

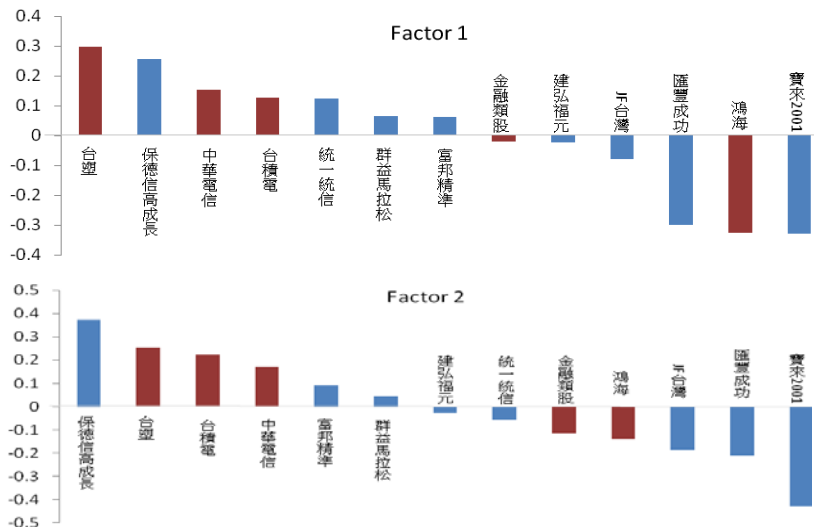
來運用馬可夫鏈蒙地卡羅法進行隨機尋找目標函數的極值，以 Varimax 轉軸判斷準則為例。模擬隨機正交轉軸矩陣 10000 次，以 Varimax 判斷準則找出目標函數極值 $\max OBJ(P | \Lambda, Y_c)$ 約為 0.0932。

由轉軸後之因素負荷散佈圖中，首先尋找最能代表因子一或因子二特色的變數，觀察變數之間散佈的相對位置，並未有靠近 橫軸或縱軸且表現值大的點。以因子的影響能力來說，台塑與保德信高成長在因子一、二都具高影響力。在轉軸前，金融類股表現在因子中具相當高的代表性，但轉軸後金融類股反而掉落到軸的中心，也就是說我們仍然維持變數間的變異重要性。換個角度來解釋變數之間的相對關係，或解釋為此筆基金與股票的整體波動受台塑及保德信高成長兩檔行為的牽引，次強力量為台積電及中華電信。我們同意股市變化會影響基金的波動，則站在伴隨效應的角度來說明因子意義，保德信高成長有強烈跟隨台塑的現象，其次為台積電及中華電信。雖然我們可由公開資訊得知基金的持股成份，卻可透過因素分析直接捉出變數之間的關係，是相當有效率的方法。



圖二十 轉軸後之因素負荷圖 (Factor 1 vs. Factor 2)

在因子一中，主要受台塑股價影響，其次為保德信高成長、中華電、台積電，其中台塑及保德信高成長行為相似，中華電信及台積電行為相近。反觀鴻海股價的行為在二因子中，雖然都對因子解釋能力小，差異於因子二較受鴻海股價的影響。分別對因子一及因子二表達出在因子中各變數的重要性，將因素負荷由高到低排序並畫直條圖，如下圖二十一所示：



圖二十一 二因子之因素負荷直方圖

因素負荷愈高代表變數在因子中愈具代表性。基金市場的操作會影響股市的漲跌嗎？有一說法，相對於股票市場進出的交易金額，基金比例極小，頂多跟著消費者信心，有助漲助跌及具催化的角色而以。換個角度來想，因素分析卻可以觀察到基金操作長期穩定的趨勢，即那幾檔股票在基金經理人心中是可長期持有的。我們大膽假設，台塑、中華電信及台積電在保德信高成長的投資組合中是長期持有的。

結合二因子的特性來觀察，構成二因子最大的變數組合都是由台塑、保德信高成長、中華電信及台積電所組成。另外，對二因子影響最小的變數也同樣由 JF 台灣、匯豐成功、鴻海及寶來 2001 所組成，這形成一個有趣的現象，由因子一來解釋整體變異的百分比高達 95.53，因子二只占整體的 1.76%，而二個因子的特性卻是如此相似，這樣的現象值得深入討論。究竟存在於股票市場與開放式股票型基金間的伴隨效應有什麼樣的互動關係？

表八 因素分析之轉軸前後因素負荷變化表

Variable	PC & ML Estimates of		Rotated	
	Factor Loadings		Factor Loadings	
	f_1	f_2	f_1^*	f_2^*
金融類股	0.71742	0.62995	-0.01943	-0.114634
鴻海	0.089754	0.244291	-0.326869	-0.138248
台積電	0.041208	0.051699	0.125906	0.223096
群益馬拉松	0.035527	0.210163	0.065275	0.04658
台塑	0.035171	-0.014063	0.296778	0.253982
保德信高成長	0.025959	0.180898	0.256297	0.370877
富邦精準	0.022478	0.159339	0.061599	0.090761
中華電信	0.021869	-0.075296	0.152739	0.170476
寶來 2001	0.014939	0.094195	-0.3299	-0.42916
統一統信	0.011937	0.084985	0.123459	-0.058266
匯豐成功	0.010761	0.04589	-0.300103	-0.20973
建弘福元	0.007137	0.057519	-0.024098	-0.027055
JF 台灣	0.006631	0.029437	-0.078972	-0.184035

未做轉軸前，因子一之因素負荷量排序後大致與各變數的帳面價值由高至低的順序相同，暫使命名為“市場淨值影響”因素。由因子中變數的組合，判斷此二因子有非常相似的特性。

伍·結論與建議

在實際的經濟變動中，包含了上百個可能影響的變數資訊。相對在方法的限制下，因素分析能夠指出隱藏在大量資料下有用的訊息，應該是相當具指標性的。本文以實際的電腦運作，分析多變量時間序列資料，經由模擬方法了解機率模型及統計模式背後的真實意義，再運用於實際資料，證明這樣的技術是相當有價值的。從研究過程中可發現，模擬資料與真實市場資料的分析結果非常相似，這也可說明模型的界定方法相當完好，即驗證機率模型的可行性是極具參考性的。經過模型的檢驗後，只要將真實資料套入即得到所有的分析結

果。本文更嘗試以各種不同的繪圖方式，呈現這個主題的背後思考，同時也將 EM 演算法適當的應用上去，這應該是本文主要的貢獻。

而本文更藉由實證的操作，分析出股票與基金之間的連動關係只存在短期的伴隨效應，並不具有長期的共整合效應。這使得在做因素分析時增加很多的方便性。這樣的新發現對投資人所言的財務意義是有很大的幫助的。例如短期投資組合所造成的避險效果是非常有限的，因為短期的正向連動。但對長期而言，因為不具有共整合效應，投資組合反而容易達成降低風險的作用。這樣的發現對投資者而言，當然可以令他們在做決策時有更深入的思想幫助。但這只是一種思考上的輔助，眾所週知，絕對“賺錢”的決策法則是不可能存在的。

從本文實證資料所選擇的變數組合，討論股票與基金市場間的相關性，可得知台塑、保德信高成長、中華電信及台積電間，具有極高相似的行為及伴隨特色。而針對伴隨效應來解釋，保德信高成長有強烈跟隨台塑的現象，其次為台積電及中華電信。本文受只做了一種變數組合的限制，無法只由本研究所選擇的十三個變數就得知整個市場的面貌，針對後續研究所給的建議可以對伴隨效應的特色做探討。嘗試不同的投資組合，是否具有不同特色的伴隨因子。

在林建義 (2008) 的模擬討論中，認為不同的資料共變結構無法單純由股價走勢判斷其特色。所以可根據本研究的想法，嘗試描繪出伴隨效應的輪廓。另外，可加入總體經濟變數將研究的範圍拉大，嘗試解釋不同市場間的互動關係。對於因素分析在財務資料的運用，經常遭遇資料結構破壞因素分析的假設，尤其在因素個數的選取上，可供後續做深入討論。

參考文獻

林建義，「股價相關性共整合分析研究-以台灣股市與國際股市為例」，天主教輔仁大學金融研究所在職專班碩士論文，2008 年。

楊奕農，「時間序列分析-經濟與財務上之應用」，台北市：雙葉書廊，2005 年。

Bartlett, M. S., "The Statistical Conception of Mental Factors", *British Journal of Psychology*, Vol 28, 1937, pp.97-104.

Bartlett, M. S., "Further Aspects of the Theory of Multiple Regression", *Proceedings of the Cambridge Philosophical Society*, Vol 34(1), 1938, pp.33-40.

Bentler, P. M. & Tanaka, J. S., "Problems with EM for ML Factor Analysis", *Psychometrika*, Vol. 48, 1983, pp.247-257.

Bernanke, B. S. & Boivin, J., "Monetary Policy in a Data-rich Environment", *Journal of Monetary Economics*, Vol.50(3), 2003, pp.525-546.

- Bernanke, B. S., Boivin, J., & Eliasziw, P., "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach", *National Bureau of Economic Research working paper*, No. W10220, 2004.
- Bernanke, B. S., Boivin, J., & Eliasziw, P., "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach", *The Quarterly Journal of Economics*, Vol.120(1), 2005, pp.384-422.
- Boivin, J., & Ng, S., "Understanding and Comparing Factor-based Forecasts", *National Bureau of Economic Research working paper*, No. W11285, 2005.
- Bollerslev, T., "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, Vol.31, 1986, pp.307-327.
- Davidson, J. E. H., Hendry, D. F., Srba, F. & Yeo, S., "Econometric Modelling of the Aggregate Time-Series Relationship between Consumers' Expenditure and Income in the United Kingdom", *The Economic Journal*, Vol.88(352), 1978, pp.661-692.
- Dungey, M., Martin, V. L. and Pagan, A. R., "A Multivariate Latent Factor Decomposition of International Bond Yield Spreads", *Journal of Applied Econometrics*, Vol. 15(6), 2000, pp.697-715.
- Enders, W. , *"Applied Econometric Time Series"*, New York: John Wiley & Sons, Inc., 2004.
- Engle, R. F., "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, Vol.50, 1982, pp.987-1007.
- Engle, R. F. & Granger, C. W. J., "Co-integration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, Vol.55, 1987, pp. 251-276.
- Engle, R. F., Lilien, D. M. & Robins R. P., "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M model", *Econometrica*, Vol.55, 1982, pp. 391-407.
- Favero, C. A., Marcellino, M. & Neglia, F., "Principal Components at Work: The Empirical Analysis of Monetary Policy with Large Datasets", *Journal of Applied Econometrics*, Vol.20(5), 2005, pp.603-620.
- Forni, M., Hallin, M., Lippi, F. & Reichlin L., "The Generalized Dynamic Factor Model: Identification and Estimation", *The Review of Economics and Statistics*, Vol.82(4), 2000, pp.540-554.
- Forni, M., Lippi, M. & Reichlin, L., "Opening the Black Box: Structural Factor Models versus Structural VARs", *Universite Libre de Bruxelles*, Mimeo, 2004.
- Granger, C. W. J. & P. Newbold., "Spurious Regressions in Economics", *Journal of Econometrics*, Vol.2, 1974, pp.111-120.
- Harman, H. H., *"Modern Factor Analysis"*, 3th ed., University of Chicago Press, Chicago, IL., 1976.
- Heaton, C., *"Factor Analysis of High Dimensional Time Series"*, Unpublished doctoral dissertation, University of New South Wales. School of Economics, Sydney, Australia, 2008.
- Johansen, S., "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics & Control*, Vol.12(2), 1988, pp.231-254.

- Mulaik, S. A., "*The Foundations of Factor Analysis*", New York: McGraw-Hill Book Company., 1972.
- Norstad, J., "Random Walks", Retrieved June 29, 2009, from John Norstad's Finance Page Web site, 2005, <http://homepage.mac.com/j.norstad/finance/#portopt1>
- Roweis, S., "EM Algorithms for PCA and SPCA", *Neural Information Processing Systems 10* (NIPS'97), 1998, pp.626-632.
- Rubin, D. B. & Thayer, D. T., "EM Algorithms for ML Factor Analysis", *Psychometrika*, Vol.47, 1982, pp.69-76.
- Rubin, D. B. & Thayer, D. T., "More on EM for ML Factor Analysis", *Psychometrika*, Vol.48, 1983, pp.253-257.
- Sargent, T. J. & Sims, C. A., "Business Cycle Modeling without Pretending to Have too much a Priori Economic Theory. New Methods in Business Cycle Research", C. Sims. Minneapolis: Federal Reserve Bank of Minneapolis., 1977.
- Spall, J. C., "*Introduction To Stochastic Search And Optimization*", Advisory edition, John Wiley & Sons, Inc., 2003.
- Stock, J. H. & Watson, M. W., "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, Vol.20(2), 2002a, pp.147-162.
- Stock, J. H. & Watson, M. W., "Forecasting Using Principal Components from a Large Number of Predictors", *Journal of the American Statistical Association*, Vol.97(460), 2002b, pp.1167-1179.
- Stock, J. H. & Watson, M. W., "Implications of Dynamic Factor Models for VAR Analysis", *National Bureau of Economic Research working paper* No. W11467, 2005.
- Tanner, M. A. & Thisted, R. A., "A Remark on AS127. Generation of Random Orthogonal Matrices", *Applied Statistics*, Vol.31, 1982, pp.190-192.
- Tsay, R. S., "*Analysis of Financial Time Series*", 2th ed., John Wiley & Sons, Inc., 2005.
- Turville, C., "*Techniques to Handle Missing Values in a Factor Analysis*", Unpublished doctoral dissertation, University of Western Sydney, Macarthur., 2000.

Factor Analysis of Multivariate Time Series Data on Stocks and Funds Market

TAI-MING LEE, CHAU-CHING CHAO *

ABSTRACT

This thesis presents the process of research into the use of factor models for multivariate time series data on stocks and funds Market. Time series data of financial exist distinctive behavior including autoregression and cointegration, to apply Error Corrected Model (ECM) to extract the correlation to gain more simple information. Using the process is quite applicable to agree the hypothesis of factor analysis about the residual series. And through the residual to analyze the correlation of short run movement for multivariate time series. This thesis present an expectation-maximization (EM) algorithm to estimate parameters .It is computationally very efficient in space and time. The real data of this research shows that almost one relationship between the stocks market and funds market is concurrent effect. To express more completely the multivariate time series data about correlation structures is main contribution by this thesis and provides a complete procedure for this kind of data analysis.

Keywords: factor analysis, time series, Error-Correction model, EM algorithm

* Tai-Ming LEE, Associate Professor, Department of Statistics and Information Science, Fu Jen Catholic University. Chau-Ching CHAO, MBA Student, Graduate Institute of Applied Statistics, Fu Jen Catholic University.

