

整合式演化計算於證券投資組合之應用

陳伊伶
輔仁大學資訊管理研究所
研究生
498746062@mail.fju.edu.tw

林文修
輔仁大學資訊管理研究所
教授
wslin1949@gmail.com

摘要

基因表達規劃法(Gene Expression Programming, GEP)為一項新穎的演化式計算，結合遺傳演算法(Genetic Algorithm, GA)的表達操作效率與遺傳規劃法(Genetic Programming, GP)的樹狀結構能力之優勢，本研究將 GEP 以及 GA 兩種演算法結合並應用於股票金融市場，透過結合兩種演算法的優點，建立一個具有選股能力且分散風險的投資組合模組。

實驗一發現本研究設計的整合式演化計算投資組合模型具有穩定性，能呈現穩定的報酬率；實驗二證實本研究設計的適應值，考量了投資組合內的報酬與風險是恰當的，換言之，一個良好的投資組合，必須同時兼顧報酬與風險這兩者；實驗三，自動定義函數可以保留較好的基因，且避免產生過大的搜尋空間，實驗效果表現比沒使用自動定義函數的模組績效還要良好。

關鍵詞：遺傳演算法、基因表達規劃法、投資組合

1. 前言

自民國 81 年以後，證期會逐漸放寬投資機構，造成國內機構投資增加，投信公司增加最多，證期會企圖藉著加重機構投資人(法人)的投資比例，來穩定國內股市。三大法人買賣股票之目的都是想獲取高額報酬，但隨著投資機構的增加，機構間就會夾持強烈的競爭壓力，投信公司募集眾人的資金購買基金從中獲得手續費及股票利潤之報酬；外資公司根據股票買賣的獲利為報酬；自營商不接受顧客委託，將公

司額外的一筆費用投入於股市，盈虧自負，雖然三大法人的獲利方式不同，但可以明顯知道自營商的獲利盈虧必須自己承擔，負責壓力較重。

自營商為集合國內資金投資於台灣股市具有穩定的效果，且目前在台灣的機構沒有其他法人機構這麼多，當自營商與其他相關投資機構或投資人的利益衝突情形發生時，如何讓證券自營商真正獲取股市價值並站穩專業投資機構角色，成為一個值得研究的方向。

國內證券自營商買賣股票之方式，通常是透過自營商內部的金融領域專業人員的知識與各方消息進行股票買賣，經由這種「質化」的傳統型投資，有時候可能會因為專業人員的主觀概念或是人員判斷誤差造成公司的損失，如果有其他參考資訊配合一起使用或許能降低風險。

近年資訊科技不斷的演化，如果能將這些影響股市的因素「量化」，並經由資訊科技技術解決複雜問題，相信對證券自營商於股市上有更多的幫助。最早是使用統計的方式計算股票於某景氣出現的機率，但是股市是屬於動態市場，影響因子錯綜複雜，單純的統計技術無法克服非線性的金融市場。目前人工智慧演算法在金融領域研究中發展相當熱烈，遺傳演算法

(Genetic Algorithm, GA) 應用於證券市場中也獲得不錯之成效(侯佳利, 2001; 江吉雄, 2002), Ferreira(2001)提出的基因表達規劃法(Gene Expression Programming, GEP)為一項新穎的演化技術，結合遺傳演算法 (Genetic Algorithm, GA) 的表達操作效率與遺傳規劃法 (Genetic Programming, GP) 解決複雜能力的樹狀圖，而達到快速解決複雜問題的能力(陳奕帆, 2010; 蔡慧菊, 2010)。不同的演算法各有其優缺點，

若混合各種演算法增加研究模型之效能，以及結合不同投資策略，運用這兩種機制讓投資人有效率地挑選股票外，也提升投資人於股市之信心。

本研究擬建立一個整合式投資組合模組，提升自營商於市場上股票投資與操作之績效，評估此整合式演化式計算的適用性，且探討基因表達規劃法 (GEP) 加入自動定義函數(ADF)的演化效能，分析遺傳演算法(GA)對股票做資金配置，是否能夠降低風險。

2. 文獻探討

2.1 遺傳演算法

遺傳演算法是模擬自然界生物演化過程及運作機制，以求問題之最佳化解決方法的人工智慧技術，主要精神以達爾文的進化論為基礎，從初始族群開始，透過適應函數，遵循其「適者生存不適者淘汰」的原則，進行交配、突變的演化過程，即產生下一代，藉由生物物種的基因，在每代間進行演化，終而尋得適當問題的最佳解。陳稼興與劉慧敏(2002)與 Kin et al. (2006)指出使用基本面指標於遺傳演算法可以幫助投資人於市場上選股，研究中顯示使用遺傳演算法(GA)演化資金權重且能降低投資風險，因此本研究運用基本面指標於選股模組且納入 GA 於資金比例配重。

2.2 基因表達規劃法

由葡萄牙學者 Ferreira 提出基因表達規劃法(Gene Expression Programming, GEP)，是一種新穎的演化式計算，不同於以往遺傳演算法(Genetic Algorithm, GA)和遺傳規劃法(Genetic Programming, GP)的表達方式，GEP 將兩種演算法結合，此種基因型(genotype)表現方式用 GA 來表達，演化速度較快；表現型(phenotype)以 GP 樹狀結構表示，可應用於較複雜的問題上，結合兩種演算法的優點，其演化速度

和效果應該會有不錯的表現。

2.2.1 GEP 編碼方式

GEP 編碼包含兩種型態，基因型(genotype)與表現型(phenotype)，基因型的部分使用 GA 表示，表現型的部分使用 GP 表示。因此 GEP 將染色體設計成由頭部(head)和尾部(tail)組成，head 部分可為函數節點或終端節點；而 tail 部分只能為終端節點。函數節點通常為 { +, -, ×, ÷, <, >, =, ≠, AND, OR }，終端節點通常為真實資料的屬性集和常數。而 head 的長度通常由使用者設定，tail 長度則透過 Ferreira 提出的公式來計算。

$$t = h \times (n_{\max} - 1) + 1$$

t 代表尾部長度(tail)，h 代表頭部長度(head)， n_{\max} 代表函數集合中每一個函數所需最多的參數，通常都是 2。假設 $h=6$ ，透過公式計算得知 $t=7$ ，因此染色體長度為 $6+7=13$ 。之後透過轉換機制，將染色體轉換成表示樹，依照染色體的基因順序逐一讀取，排列成表示樹的方式為由上到下、左到右按照基因順序排列。圖 1 (a) 為基因表達規劃法編碼示意圖，透過轉換機制轉換成表示樹如圖 1(b)。

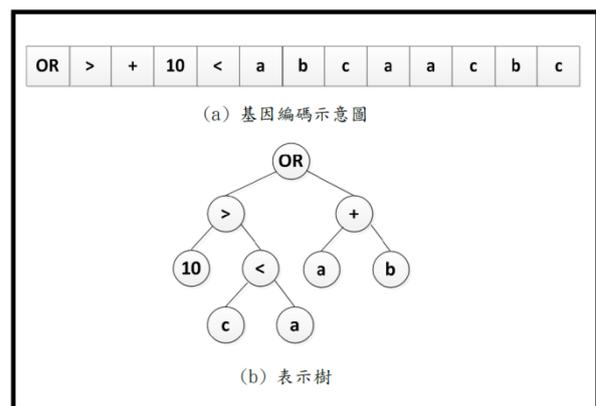


圖 1 基因表達規劃法編碼示意圖

2.2.2 GEP 相關文獻

GEP 是一個新穎的演化技術，結合遺傳演算法(GA)的表現方式和遺傳程式規劃法(GP)解決複雜問題的能力，所以 GEP 執行時其演化速度和效果比單執行這兩項演化式計算還要佳。Ferreira 還加入了遺傳規劃法(GP)創始人 Koza 教授(1992)提出的自動定義函數(Automatically Defined Function, ADF)，利用封裝的概念將演化過程中較複雜的子樹轉變成一個自動定義函數節點，此方法可以增加 GP 的效率與應用。

陳玲楨(2008)與陳奕帆(2010)研究中顯示出使用不同的 ADF 設計方式會影響其效果，但陳玲楨研究結果顯示只有在資料值域小的情況下表現才較好，其原因可能是因為此研究的 ADF_0 和 ADF_1 被限制住，能夠演化的方式較少，所以造成值域大的資料集表現效果沒那麼好。Chen et al. 結合 GEP 和免疫演算法(IP)的移動視窗預測模型於股市預測，適合非線性資料之應用(2007)；經由上述整理，本研究認為將財務觀念納入 ADF 方式是可採取的模式，將 ADF 概念加入 GEP 模型中，並建立移動視窗預測模型。

2.3 投資組合

投資組合概念緣起於馬可維茲(Harry Markowitz, 1952)之平均數-變異數投資組合模式，他認為不同的資產藉由計量與平衡後，投資人可建構出分散型的投資組合，即「相同風險下，預期報酬最大」或「相同預期報酬下，風險最小」作為篩選最佳投資組合的準則(Markowitz, 1952)。馬可維茲最先以數學模式分析投資組合的學說，精確計算投資組合風險與報酬，之後與夏普(Sharpe)一起研究，發展成現代投資組合理論。

投資組合概念興起至今，一直是金融領域中重要議題，且相關研究一直持續探討。沈峰儀(2002)與張婉蘭(2002)使用馬可維茲投資組合之概念能降低投資風險，產

生最佳報酬與風險的投資組合，且沈峰儀(2002)將馬可維茲概念納入 GA 演算法，張婉蘭(2002)是將投資組合結合景氣循環調整；楊千霽(2003)將選股策略、擇時策略、資金配置策略納入投資組合；Orito et al. (2003)於投資組合中加入 GA，可以產生多元化投資組合；Oh et al. (2005)透過排名方式建構投資組合，可見投資組合可以使用多種不同方式建構而成；經由上述整理，本研究擬採用馬可維茲投資組合之概念，並將選股策略與資金配置納入投資組合。

3. 研究設計

3.1 研究架構

本研究將 GEP 基因體和 GA 基因體組成一條染色體一起進行共演化，一條染色體代表一個投資組合。GEP 基因為選股基因，將 GEP 選股基因納入基本面指標並篩選出符合的股票後，GA 基因就根據符合的股票做資金配置，如圖 2 所示，因此一條染色體(即一個投資組合)，經過 GEP、GA 演化機制，將會產生出一個最適報酬率與風險，就將此染色體設為最適投資組合。

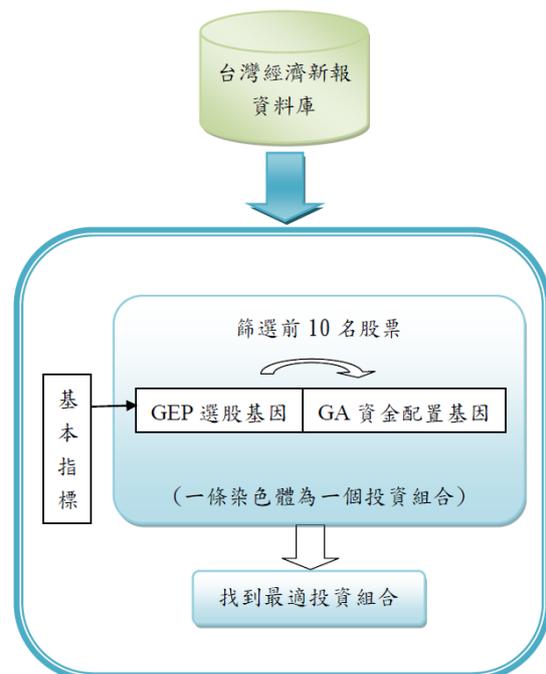


圖 2 研究架構圖

3.2 變數選擇

劉宗聖等人(2007)提到最直接影響股票價格走勢的因素，應該就是該公司之財務狀況，而基本面因子就是用來描述公司財務狀況之基本工具。最常見的基本面因子通常是由損益表、資產負債表及現金流量表計算出來的數據，投資人依數據來評估企業價值或證券預測。本研究根據杜金龍(1996)將基本面指標分為獲利能力、成長能力、財務經營、償債結構、總體經濟共五種構面，表 1 將介紹各構面包含的基本因子。

表 1 基本面指標代碼

獲利能力指標		成長能力指標	
X1	每股盈餘	X10	總資產報酬成長率
X2	資產報酬率	X11	營收成長率
X3	營業利益率	X12	營業利益成長率
X4	稅後淨利率	總體經濟指標	
X5	營業毛利率	X13	經濟成長率
財務經營能力指標		X14	失業率
X6	總負債比率	X15	消費者物價指數
X7	本益比	償債結構指標	
X8	現金股利率	X16	流動比率
X9	股東權益報酬率	X17	負債淨值比

3.2.1 獲利能力指標

每股盈餘：每股盈餘高代表公司每單位資本額的獲利能力高，這表示公司具有某種較佳的能力，使得公司可以用較少的資源創造出較高的獲利。

$$\frac{\text{盈餘}}{\text{在外流通股數}}$$

資產報酬率：資產報酬率即在衡量公司的資產是否充份利用，比率越高，表示公司的營運使整體資產的報酬運用效率越高。

$$\frac{\text{當期稅後淨利}}{\text{總資產}}$$

營業利益率：營業利益率為公司每創造 1 元的營收所能得到的獲利。

$$\frac{\text{營業收入}-\text{營業成本}-\text{營業費用}}{\text{營業收入}}$$

稅後淨利率：用以測試企業經營獲利能力的高低，淨利率越高表示企業獲利能力越好。

$$\frac{\text{本期淨利}}{\text{營業收入}}$$

營業毛利率：毛利率是一家公司獲利能力的最基本指標，因此觀察其變化將可找出公司獲利變化的趨勢。

$$\frac{\text{營業收入}-\text{營業成本}}{\text{營業收入}}$$

3.2.2 財務經營能力

總負債比率：總負債比率是用來衡量一家公司資本結構的重要指標，通常總負債比率高的公司具有較高之財務風險，獲利不如預期時，破產風險較高。

$$\frac{\text{當期總負債}}{\text{總資產}}$$

本益比：指投資人對每一元盈餘所願意付出的價格。投資人以此比率衡量股票是否值得投資。當本益比越大時，投資報酬率越低，此種公司要具有大幅成長潛力，投資人才願意投資，但此種公司風險通常較大。

$$\frac{\text{每股市價}}{\text{每股盈餘}}$$

現金股利率：高股利的股票具有穩定的盈餘品質，長期具有穩定產生收益的能力。

$$\frac{\text{每股股利}}{\text{股票價格}}$$

3.2.3 成長能力指標

股東權益成長率：由股東權益成長率可以看出，該公司的經營階層的目標是否與股東目標一致，是以股東權益為主要考量。

$$\frac{\text{淨利}}{\text{股東權益總額}}$$

總資產報酬成長率：營收擴張如果未能使利潤同幅增加，就成了做白工，用此指標來確認成長性。

$$\frac{\text{稅前息前折舊前淨利變動額}}{\text{前期總資產}}$$

營收成長率：營收成長率越高之公司，代表公司的業務證此於擴張時期。

$$\frac{\text{單月營收}-\text{去年同期營收}}{\text{去年單月營收}}$$

營業利益成長率：不同於營收成長率，營業利益成長率是與去年同期的營業額計算比例，看是否比去年同時期為佳。

$$\frac{\text{單月營業利益}-\text{去年同期營業利益}}{\text{去年單月營業利益}}$$

3.2.4 總體經濟指標

經濟成長率：此比率衡量整體經濟環境之現況，當經濟成長率高時，出口導向的經濟體中，外銷類股會有較佳之表現，而內需型經濟體中，內銷類股會有較佳之表現。

消費者物價指數：在經濟學上，是反映與居民生活有關的產品及勞務價格統計出來的物價變動指標，以百分比變化為表達形式。

它是衡量通貨膨脹的主要指標之一。

失業率：為失業人口佔勞動人口的比例，目的為衡量閒置中的勞動產能。在台灣於每月 23 日由主計處行政院公布，失業數據的月份變動可適當反應經濟發展，大多數資料都經過季節性調整，失業率通常被視為落後指標。

3.2.5 償債結構指標

流動比率：表示每 1 元流動負債即有多少元的流動資產來保障。流動比率越高表示短期償債能力越強，營業週期越短者，其流動比率越低。

$$\frac{\text{流動資產}}{\text{流動負債}}$$

負債淨值比：表示債權資金佔權益資金的比重大小，此數值越大，表示負債程度越高，企業還本付息的壓力愈大。

$$\frac{\text{負債總額}}{\text{股東權益總額}}$$

3.3 模組架構流程

本研究模組運作流程請參閱圖 3 以及下面之流程說明。

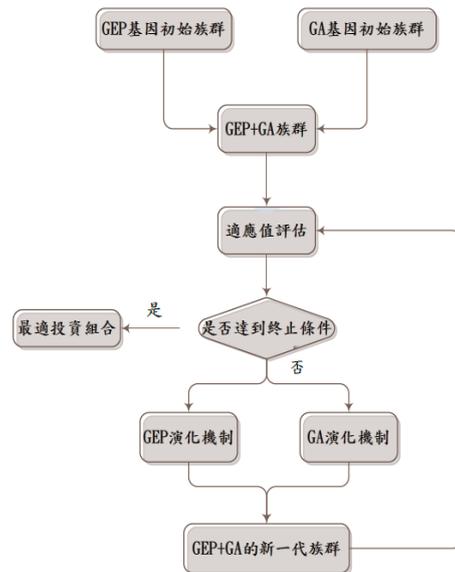


圖 3 GEP-GA 共演化投資組合之架構流程圖

流程說明：

- GEP 基因初始族群：系統在第一代隨機產生由五個 ADF 組成的 GEP 選股基因。
- GA 基因初始族群：系統依照前述 GEP 選股基因，篩選出前 10 名股票，對這 10 名股票隨機產生變數。
- GEP+GA 族群：計算 10 支股票的資金配置比例後，將 GEP 和 GA 基因組合起來，成為一個染色體，一個染色體代表一個投資組合。
- 適應值評估：評估 GEP+GA 染色體的績效，透過適應函數計算每個染色體的適應值。
- GEP 演化機制
 - 選擇：從原族群中選擇適應值較高的染色體至下一代。
 - 突變：隨機選取染色體中的基因作突變。
 - 重組：隨機選取兩個染色體的基因區段或單基因互相交換。
 - 轉位：複製染色體中的某基因區段，開頭必須包含運算函數，將其區段插入染色體頭部的第一個元素，其餘基因往後移動，移動範圍只限在頭部長度。
 - 反轉：選擇染色體的區段，將其序列反轉取代原染色體。
- GA 演化機制
 - 選擇：從原族群中選擇適應值較高的染色體至下一代。
 - 交配：隨機選取兩個染色體的基因互相交換。
 - 突變：隨機選取染色體中的基因作突變。
- GEP+GA 新一代族群：透過染色體演化產生新一代族群。
- 終止條件：當代數達到 200 代結束演化過程。
- 透過步驟四到七不停的演化過程，到達終止條件為止，產生最佳投資組合。

3.4 編碼方式

GEP-GA 共演化投資組合模組，將基因表達規劃法(GEP)基因和遺傳演算法(GA)基因結合為一個染色體，一起進行共演化。首先 GEP 裡面的基因是由頭部和尾部所組成，頭部包括函數節點與終端節點，尾部只有終端節點，本研究將 GEP 基因納入基本面指標，因此函數節點設定為運算符號，如「<、>、=、AND、OR」，終端節點為研究目標的屬性與常數，如「每股盈餘、本益比」。頭部長度 h 與尾部長度 t 的關係，經由 Ferreira 設計，請參照下列公式。

$$t=h*(n_{max} - 1) + 1$$

n_{max} 為函數集中每一函數所需最多的參數，在 GEP 基因中使用二元樹，因此一棵樹的節點最多會使用兩個參數，所以 n_{max} 設為 2。本研究將基本面指標分為五種構面，每一個構面代表一個 ADF 基因片段，因此本研究的 GEP 基因由有五個 ADF 基因片段所組成。每個基因片段的頭部長度設為 5，透過公式運算尾部長度為 6，一個 ADF 基因片段長度為 $5+6=11$ ，因此本研究的染色體總長度為 $11 * 5 = 55$ ，ADF 基因片段之間用 AND(連結函數)結合。GEP 基因設計完之後須轉換成表示樹，再將資料庫內的股票一筆一筆放進去，符合此樹的規則就當成候選股票，候選股票經由報酬率排名，依法人的要求選擇前 10 名，因此 GA 基因就設計為 10 個 bits 當作資金配置比例。第一個 10 為股票代碼，第二個 10 為資金配置比例，圖 4、圖 5 以簡單方式說明其運作方式。

圖 4 的 GEP 基因片段頭部長度為 2， n_{max} 為 2，透過公式運算尾部長度為 3，基因片段長度為 $2+3=5$ ，本研究採用五種基本構面，一個構面代表一個基因片段，所以 GEP 基因長度 $5*5=25$ 。一個基因片段將採用自動定義函數(ADF)的方式建置，ADF 基因片段就代表一顆子樹，因此這個染色體是由五個 ADF 基因片段組成，染色

體會轉換成五個子樹，如圖 5 所示。子樹間透過一個連結函數(And)結合，符合此樹規則之股票設為候選股票，再選取候選股票報酬率為前三名者當作投資組合內的股票，因此將 GA 基因長度設為 3，分別代表資金配重比例，所以染色體總長度為 $25+3=28$ 。GA 資金配置比例計算方式為 $\frac{15}{15+10+30} = 0.27$ ，因此第一支股票的資金配重比例為 27%，第二支比重為 18%，第三支比重為 55%。

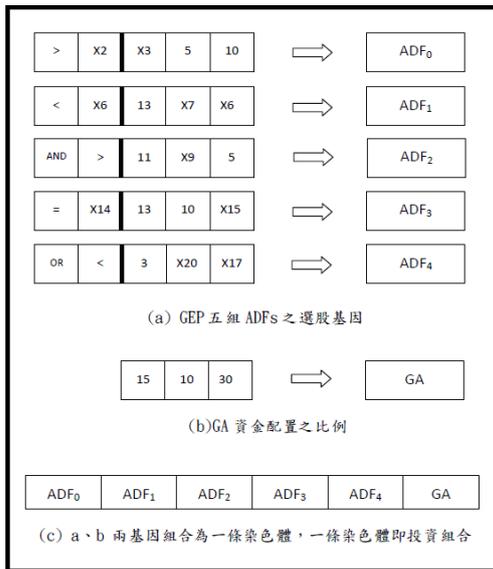


圖 4 GEP-GA 共演化投資組合編碼方式

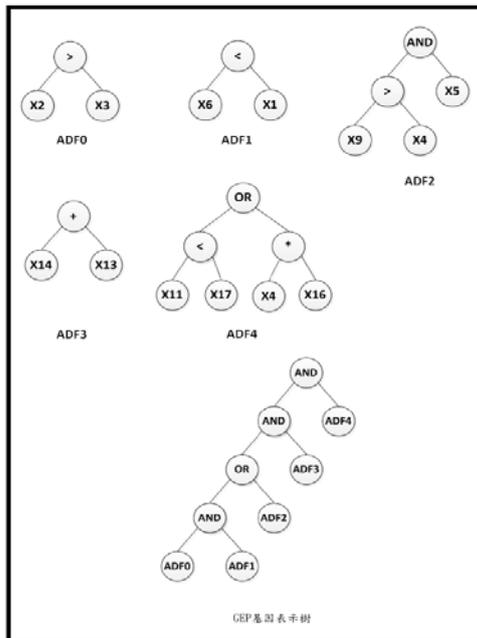


圖 5 GEP-GA 共演化投資組合表示樹

3.5 適應值與參數設定

3.5.1 適應值

本研究使用馬可維茲概念(1952)，投資組合整體標準差不是由單一資產的標準差決定，各資產之間的相關性才是重要角色。因此 $E(R_p)$ 為期望報酬率，用來評估投資組合所能獲取的利潤， σ_p 為投資組合之標準差，用來評估投資組合內股票兩兩之間在市場上的風險。

當染色體計算好資金配置比例後，就計算整個投資組合的期望報酬 $E(R_p)$ ，並計算投資組合之標準差 σ_p 。本研究希望獲取的利潤越高越好且風險越小越好，因此本研究採用以下的適應值評估投資組合之績效，最後選擇適應值較高的染色體當作最佳投資組合。公式如下：

$$\text{適應函數} = \frac{E(R_p)}{\sigma_p}$$

$$E(R_p) = \sum_{i=1}^n W_i * R_i$$

$$\sigma_p = \sqrt{\sum_{i=1}^n W_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{j=1}^n W_i W_j \sigma_{ij}}$$

$$\sigma_{ij} = \frac{\sum_{a=1}^n (R_{ia} - R_{iave}) * (R_{ja} - R_{jave})}{n - 1}$$

其中， $E(R_p)$: 投資組合之期望報酬

σ_p : 投資組合之標準差

R_i : 第 i 支股票的當月報酬率

W_i : 第 i 支股票的權重

σ_{ij} : 第 i 與第 j 支股票之報酬共變異數

R_{ia} : i 股票在第 a 月的報酬率

R_{ja} : j 股票在第 a 月的報酬率

R_{iave} : i 股票的平均報酬率

R_{jave} : j 股票的平均報酬率

3.5.2 參數設計

本研究基因表達規劃法(GEP)與遺傳演算法(GA)的選擇方式採用菁英法以及 Goldberg(1989)提出的輪盤法，菁英法是保留前五名個體至下一代。GEP 演化方式是對個別基因體(ADF)分別做突變和重組。相關參數設定請參閱表 2。

表 2 參數設定表

類型	控制項目	基因表達規劃法(GEP)	遺傳演算法(GA)
終端節點	資料屬性	X1-X17	-
函數節點	算術運算	+ , - , × , ÷ , < , > , = , AND, OR	-
	連結函數	AND , OR	-
演化參數	族群大小	100	
	終止條件	200 代	
	基因長度	55	10
	選擇方式	菁英法、輪盤法	
	突變方式	多點突變	多點突變
	突變率	0.3	0.01
	重組方式	基因重組、雙點重組	雙點交配
	基因重組率	0.3	-
	雙點重組率	0.15	0.7

4. 實驗設計

4.1 資料來源

實驗資料取自台灣經濟新報資料庫(TEJ)，樣本資料為台灣中型 100 指數成分股，代表台灣市場具有成長潛力之中型股票。研究時間為 2009 年 1 月 1 日起至 2010 年 11 月 30 日，抓取個股月資料，共 23 個月。實驗時間採用「月」的時間移動序列進行，移動視窗長度為七個月，前六個月為模型的訓練期，最後一個月為測試期，每次移動視窗長度為一個月，總共有 17 個

移動區間視窗。

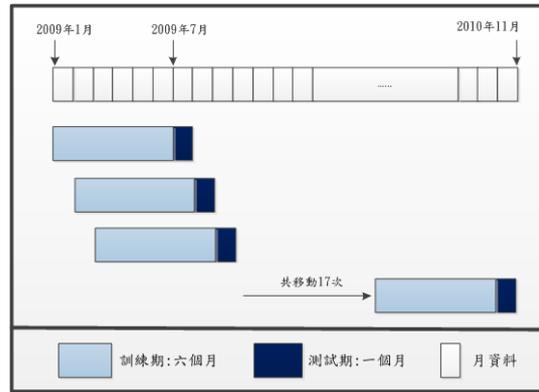


圖 6 移動視窗示意圖

4.2 模組績效評估

實驗一：穩定度測試

為了證實本研究之穩定性，實驗一將模組執行五次，且將每個移動視窗的最佳適應值呈現出來，觀察本模組在同個期間執行的適應值是否趨於一致。圖 7 顯示執行五次的最佳投資組合的適應值，雖然在 2009/08 以及 2010/07 適應值沒有達到一致的趨勢，但是其他移動視窗的適應值幾乎呈現相近的趨勢走動。表 3 呈現五次最佳投資組合訓練期與測試期的報酬率，五次執行結果很明顯地訓練期的報酬率都高於測試期報酬率，其結果可能是因為實驗時間處於股市上漲的期間。由圖 8 大盤走勢可以看出，2009/01-2010/11 台灣中型 100 指數呈現成長趨勢，因此在測試期為一個月的的情況下，擁有股票的時間太短，無法獲得更高的報酬，但是根據表 3 測試期報酬率的標準差來看，執行五次的報酬率標準差不高，綜合上述實驗一可證實本研究具有基本的穩定性。

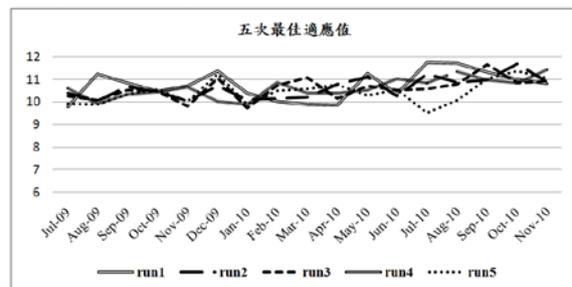


圖 7 五次最佳投資組合之適應值

表 3 五次最佳投資組合之報酬率

執行次數	訓練期報酬率(%)	測試期報酬率(%)
1	8.260544	0.77309
2	5.408335	0.721719
3	4.281187	-0.02949
4	2.65352	1.741062
5	5.423334	0.121944
平均	5.205384	0.665665
標準差	2.04871	0.698218

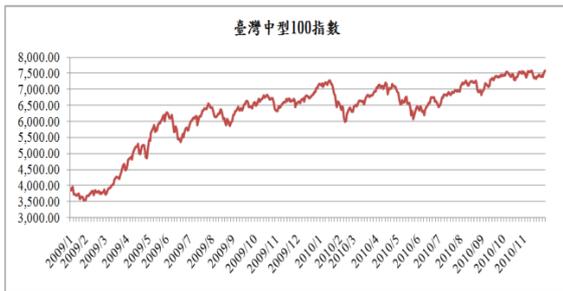


圖 8 大盤走勢圖

實驗二：前五名投資組合效能

本研究採用菁英法，將每代前五名投資組合保留至下一代，並且依照投資組合的期望報酬與標準差之比值來評估投資組合，是否達到最適報酬與風險的效能，也就是將五次執行的前五名投資組合報酬率進行平均。圖 9 顯示前五名投資組合的訓練期與測試期之報酬率。實驗一已經說明造成訓練期與測試期報酬率之差異，其原因是因為實驗期間處於股市上漲時期。首先圖 9 可以觀察出前五名投資組合於測試期的報酬率呈現遞減，雖然第四第五名投資組合報酬率為負值，但是負值很小。另外可以看出雖然第一名訓練期的報酬率不及二三四名，但是測試期報酬率卻比較好，這可以證實本研究在設計適應值的部分，考量到投資組合內的報酬與風險是恰當適宜的。一個良好的投資組合，必須同時兼顧報酬與風險這兩者因素，且根據表 4 前五名投資組合的適應值來看，前五名適應值數據相差不大。

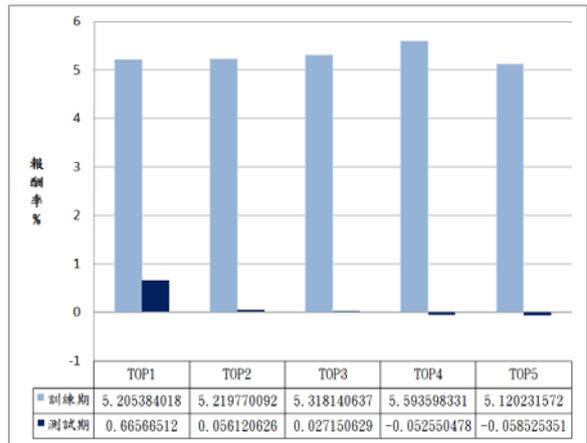


圖 9 前五名投資組合報酬率

表 4 前五名投資組合適應值

	TOP1	TOP2	TOP3	TOP4	TOP5
run1	10.75797	10.75769	10.75765	10.74832	10.74827
run2	10.62545	10.61776	10.61757	10.61737	10.61737
run3	10.5863	10.58593	10.5858	10.58577	10.58553
run4	10.62153	10.6128	10.60465	10.60452	10.60361
run5	10.44842	10.44829	10.44828	10.44828	10.44826
Average	10.60794	10.60449	10.60279	10.60085	10.60061

實驗三：自動定義函數測試

本模組在基因表達規劃法(GEP)選股基因使用自動定義函數(ADF)，目的是避免好的基因在演化過程中被破壞，以及避免產生過大的搜尋空間，與未使用 ADF 的模組比較適應值與報酬率，其目的為檢視本研究使用自動定義函數是否能保留較好的基因以及搜尋空間是否會影響效果。圖 10 呈現使用 ADF 與未使用 ADF 的訓練期與測試期報酬率進行比較，結果發現使用自動定義函數(ADF)的訓練期與測試期報酬率皆比未使用的報酬率還要高，其因為可能是使用 ADF 能夠保留住好的基因體，不會因為演化過程導致良好基因體被破壞。圖 11 選取 2009/02-2009/09 移動視窗期間，呈現使用 ADF 與未使用 ADF 的適應值演化趨勢，圖中可以發現有使用 ADF 的最佳解比未使用的還要高，而且當個體演化到某個期間後，未使用 ADF 的模組最佳解就

無法繼續成長下去。不過以兩模組的平均適應值來看，未使用 ADF 在每一代的平均值都比有使用 ADF 的還要高，其原因有可能是在未使用 ADF 的情況下，搜尋空間比較大，所以整個族群的適應值都偏高，但是又因為演化過程中沒有將良好的個體保留下來，才會造成最後最佳適應值無法比使用 ADF 還要高，且訓練期與測試期的報酬率效果也較不好。

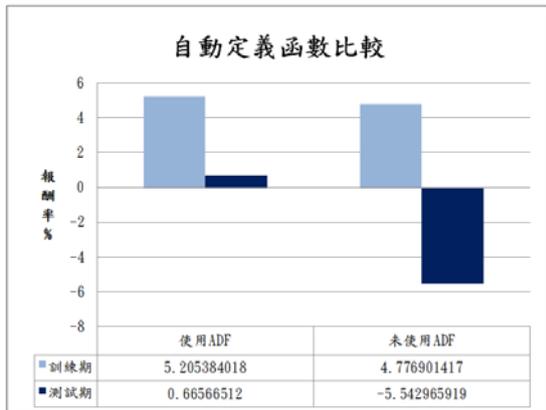


圖 10 未使用自動定義函數(ADF)報酬率比較圖

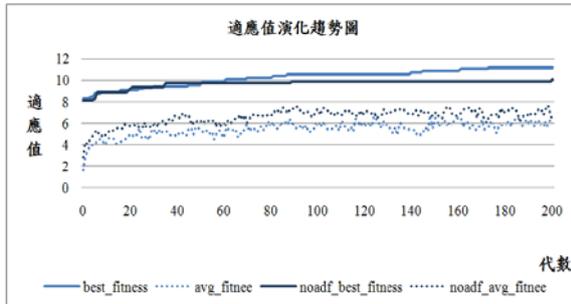


圖 11 自動定義函數適應值演化趨勢圖

5. 結論

本研究結合基因表達規劃法以及遺傳演算法於證券投資之應用，其目的是希望提供自營商在買賣股票時有更客觀的系統做參考以及更簡單易懂的選股概念。因此研究中將重要的基本面因子納入基因表達規劃法，且使用自動定義函數建構選股規則的基因，並且設計遺傳演算法當作候選股票的投資權重，將兩者演算法的基因結合成一個染色體，即一個投資組合，讓自營商在投資股票時，可以參考選股基因建

構出來的規則以及買賣權重比例，適度地調整投資金額。

研究結果發現：第一，本研究設計的整合式演化計算之投資組合模型具有穩定性，能呈現穩定的報酬率(實驗一)；第二，證實本研究設計的適應值，考量了投資組合內的報酬與風險是恰當的(實驗二)，換言之，一個良好的投資組合，必須同時兼顧報酬與風險這兩者；第三，自動定義函數可以保留較好的基因，且避免產生過大的搜尋空間，實驗效果表現比沒使用自動定義函數的模組績效還要良好(實驗三)。

參考文獻

- [1]江吉雄，遺傳演算法於股市選股與擇時策略之研究，國立中央大學資訊管理學系碩士論文，2002。
- [2]杜金龍，基本分析在台灣股市應用的訣竅，金錢文化，1996。
- [3]侯佳利，組合編碼遺傳演算法於投資組合及資金分配之應用，國立中央大學資訊管理學系碩士論文，2001
- [4]許家榮、粘瑞益、張力文、劉宗聖，計量化投資策略：資產管理的創新與應用，財訊出版社，2007。
- [5]陳玲穎，以 GEP 為基礎的分類規則產生器，元智大學資訊管理學系碩士論文，2008。
- [6]陳稼興，劉慧敏，多目標遺傳演算法於基本面選股策略之應用，第十三屆國際資訊管理學術研討會論文集(II)，2002/05，pp.365-372。
- [7]陳奕帆，演化式計算在股票型共同基金最適資產配置之研究，輔仁大學資訊管理系碩士論文，2010。
- [8]蔡慧菊，基因表示規畫法於台股期貨價格發現之研究，輔仁大學資訊管理學系碩士論文，2010。
- [9]Ferreira, C., Gene Expression Programming: A New Adaptive Algorithm for Solving Problem, *Complex System*, Vol.13, 2001, pp.87-129.

- [10] Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, California: Addison-Wesley, 1989.
- [11] Holland, J. H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.
- [12] Kin Keung Lai, Lean Yu, Shouyang Wang, and Chengxiong Zhou, A Double-Stage Genetic Optimization Algorithm for Portfolio Selection, *Lecture Notes in Computer Science*, Vol. 4234/2006, Springer Berlin / Heidelberg, 2006.
- [13] Koza, J.R., *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Cambridge, MA:MIT Press,1992.
- [14] Markowitz, Harry M., Portfolio Selection, *The Journal of Finance*, Vol.7, 1952, pp.71-91.
- [15] Yuehui Chen, Qiang Wu, and Feng Chen, An IP and GEP Based Dynamic Decision Model for Stock Market Forecasting, *Springer-Verlag Berlin Heidelberg*, 2007, pp.473-479.