應用類神經網路於國內開放式科技基金淨值預測之研究

陳香伶[1]、呂立晨、劉益萍、傅麒錦、陳昱達

南亞技術學院 財務金融系

摘要

本研究係以國內開放式、股票型、單一市場之成長型科技基金為研究對象，在考量基金成立時間之後，從中選出三支科技基金－保誠高科技基金、華南永昌前瞻科技基金、以及匯豐龍騰電子基金為研究標的，其資料範圍從民國 86 年 1 月至民國 95 年 12 月，共計 120 筆資料。此外，在蒐集 36 個可能影響基金績效及淨值的因素之後，本研究使用統計軟體 MINITAB 及 SPSS 所做的初步篩選所建立的預測模式有兩種（模式一及模式二），而模式三為未經任何篩選機制而考量所有因素的對照組。研究中針對前兩種模式進行敏感度分析，所考量之網路學習參數有：隱藏層節點數、學習循環次數、學習速率、慣性因子等。研究結果顯示，隱藏層只有一層時測試結果多為過度學習，選擇使用兩層是較有效且適當的方式；在節點數的選擇上，不論在哪一隱藏層，5 或 10 個節點皆為最佳組合；而當學習循環次數超過 1000 次之後，過度學習情況隨即產生，其較佳循環次數為 100~500 之間；再者，學習速率方面設定值為 1 時效果較佳。最後，以平均絕對誤差（MAE）及平均均方誤差（MSE）為主要評估指標，將各模式預測結果相互做一比較，證實經過篩選後的預測模式（模式一及模式二）於三支基金皆能提出較佳的淨值預測，而三支基金的模式二預測值皆為表現最佳者。

關鍵字: 共同基金、淨值預測、類神經網路、逐步迴歸分析

聯絡作者：電子郵件地址: achen@nanya.edu.tw

Abstract

This study utilizes an artificial neural network （ANN） approach to predict the net asset value of open-end Equity mutual funds that follow value, blend and growth investment styles. Three models consider a nonlinear optimizer, fund-specific historical operating characteristics and economic measures to forecast mutual funds values. A total number of 36 variables are evaluated and used here. We employ a stepwise selection method of linear regression as an initial stage of screening and build two reduced ANN models, called model one and model two; as well as, we build one ANN model without any screening mechanism for reference purpose, called model three. Results are encouraging in that the reduced ANN models outperform the reference model in predicting their net asset value. In addition, our model two performs the best among the three in predicting the net asset value.

Key words: Mututal Fund, Artificial Neural Network, NAV Forcasting, Stepwise Statistical Analysis

壹、 緒論

台灣在共同基金的發展始於1983年，相較於英國、美國的基金起步較慢，但成長卻極為迅速。截至民國96年1月，國內的基金經理公司達41家，共同基金發行數則達到509支，基金總規模為2兆2百億。正因如此，國內共同基金市場活絡的發展在近年來逐漸成為投資客的焦點，面對當今瞬息萬變的金融市場及琳瑯滿目的投資商品，投資人往往不知如何做決定，且因為政治、經濟、環境、基金本身、專業經理人等因素皆可能影響基金的績效及淨值，對於投資標的之損益更加難以預測，因此本研究以投資國內科技相關產業的基金為研究對象，此類型的基金為開放式、股票型、且單一市場之成長型基金，找出影響基金淨值變化的因子，協助準確預測基金淨值，進而獲取更高的投資報酬。

由於基金市場是一動態且開放市場，任何一套預測規則都需要有自我調整的能力以適用於多變的投資環境，而一般常見的預測分析方法多半以統計理論為基礎的基本面分析，將所有可能影響基金淨值及績效的因素納入考量，建立一預測模型，但由於統計分析法需建立較多的關係變數和蒐集大量的資料，且運用較繁雜的數學式及迴歸模型來分析各因子之間的關係並進行預測，故所需時間較長，實用性較低。有鑑於此，人工智慧方式之運用漸漸變成市場預測的主流，藉由資訊科技的協助，不斷調整模型變數與參數，並以電腦找出最適當的解答，如此可以大幅度減低運算及處理時間，更提供使用者參考資訊以迅速且正確做出適當決策，因此本研究選擇人工智慧範疇中的類神經網路為預測工具，期望藉由此一學習性的工具，準確預測出基金的趨勢性。

以下文章結構如下：第二節中將針對相關文獻進行討論，第三節介紹所使用之類神經網路模型－倒傳遞網路的理論，第四節則為實例驗證，而第五節中將針對本研究之重點及結果做一總結。

貳、 文獻探討

由於共同基金商品有如雨後春筍般出現，許多學者紛紛開始從事基金方面的研究，試圖從中發展出一些有效的基金績效評估及淨值預測方法，以下是國內外相關文獻之探討。

對於共同基金的績效評估，所依據的指標很多，譬如：Treynor指標、Sharpe指標、Jensen指標、風險值（VaR）及報酬率等。Elton等學者（1993）的研究報告中引進市場投資組合的股價指數與債券市場指數，藉此計算出調整後的異常報酬率。研究中也發現基金週轉率與總費用率愈高，其基金會有較低的異常報酬率。Christopherson等學者（1998及1999）的研究文獻中則指出共同基金所承擔的系統性風險及所獲得之異常

報酬會受到總體經濟因素的影響而有所調整，因此建議有必要將這些總體經濟資訊有效的考量在CAPM模型中。Bollen與Busse（2000）研究中運用Treynor及Mazuy模型與Henriksson及Merton模型做為衡量基金績效的依據，並且指出在選取不同期間長短的基金淨值報酬率資料，會影響投資績效的評估結果。

而國內相關研究文獻中，劉濠葦（2000）利用倒傳遞類神經網路建構基金績效分類系統。張振魁（2000）使用過去台灣股票實際交易的個股歷史資料及相關的財經資料，利用類神經網路發展一套可以提供投資人做股市交易參考的決策支援系統。楊孟龍（2000）以台灣證券交易市場部份歷史資料，應用類神經網路於選股和預測個股股價在波段中漲跌走勢，輔以資金配置及交易策略實驗投資報酬。黃綺年（2004）以96檔國內開放式股票型基金做為研究樣本，並利用淨值、週轉率、夏普指標、貝他係數、崔納指標五個指標做系統分類，及使用統計方法與自我組織映射類神經網路應用於投資績效分類與投資報酬率預測。游建欣（2006）運用適應性網路模糊推論系統於台灣股票加權指數預測之研究。由文獻中不難發現，類神經網路多應用於股票市場之股價預測，對共同基金市場而言，相關文獻仍屬有限，因此，以下本研究將以類神經網路進行基金淨值之預測。

參、 研究方法

本研究在蒐集所有可能影響基金績效及淨值之各種因素後，首先進行因素篩選的程序，亦即透過迴歸分析等相關統計分析，挑選出具有決定性影響之因素做為第二階段類神經網路之輸入因子，而類神經網路模式則透過敏感度分析調整網路參數，如：隱藏層節點數、學習循環次數、學習速率、慣性因子等，以找尋最佳之網路參數組合。以下將針對因素篩選程序及類神經網路之理論進行詳細說明。

一、輸入因子之決定

在研究過程中，往往會面臨到需要從許多的預測因素中，選出具有影響的因素，而逐步迴歸分析法（Stepwise Regression Analysis）經常使用於變數選擇（Variable Selection），其主要架構為逐次加入或刪除一個變數於迴歸模式中，以探討其對解釋模型之影響是否顯著，至於要加入或刪除哪個變數是基於該變數能否改善模式而定。由於預測變數之選取關係著迴歸模式之好壞，若自變數選取得當，能讓判定係數（R^2）愈大，且平均均方誤差（MSE）愈小，則模式愈好。因此逐步迴歸模式之主要目的，係選取較佳的自變數組合，使得模式達到最佳的狀態。本研究所採用之逐步迴歸法為順向選擇法與反向淘汰法的綜合，可視為順向選擇的改良。首先模式中不包含任何預

測變項，然後採順向選擇法，每次只容許一個預測因素進入迴歸公式，根據對模式的貢獻最大者，挑選預測因素進入迴歸模式中。而在接下來的每一步驟中，已被納入模式的預測變項則必須再經過反向淘汰法的考驗，以決定該變項要被淘汰亦或留下。換言之，當所放入迴歸方程式中之預測因素對模式不再具有顯著的貢獻時，則會被淘汰。在逐步迴歸各階段所產生之模式，將藉由R-Sq值高於0.95以上及C-p值36以下之門檻進行篩選，之後所選出的模式將以類神經網路加以訓練與測試，決定最佳者為模式一，但由於此法之缺點可能存有共線性之疑慮，因此。模式一決定之後，遂計算變異數影響因子（Variance Inflation Factor，簡稱VIF），當有任何輸入因子之VIF值大於10，將自模式中移除，如此之步驟重複直到無任何影響因子的VIF值大於10為止，如此決定之模式名為模式二，將與模式一共同利用類神經網路進行敏感度分析，挑選最佳之網路參數組合。

二、類神經網路

類神經網路最早於 1957 年，由 Rosenblatt 提出第一種類神經網路模式－感知機（Preceptor），主要用於理論研究與樣本識別，為類神經網路的發展建立基礎。所謂類神經網路，是模仿生物神經網路所衍生而成的，在現代智慧型控制的領域裡，類神經網路已成為主流，而類神經網路發展至今已具有許多不同形式及功能，其分類方式也有許多種。本研究是運用監督式學習之類神經網路－倒傳遞類神經網路（Backpropagation Network），適用於診斷、預測等方面。倒傳遞類神經網路模式可說是目前類神經網路學習模式中最具代表性、應用最普遍的模式。

倒傳遞類神經網路之基本原理是利用最陡坡降法（The Gradient Steepest Descent Method）的觀念，將誤差函數予以最小化，其架構可區分為輸入層、隱藏層與輸出層。輸入層（Input Layer）為網路之輸入變數，即訓練範例之輸入向量，其處理單元數則依研究目的之需要而訂定；隱藏層（Hidden Layer）表現輸入處理單元間之交互影響，其處理單元數目並無標準，須以試誤（trial and error）方式決定其最佳數目；輸出層（Output Layer）表現網路之輸出變數，其處理單元數目亦依問題而定。

倒傳遞類神經網路之學習過程中，首先需設定網路參數；再者，根據均勻分佈隨機亂數設定加權值矩陣W_xh與W_hy，以及偏權值向量θ_h 與θ_y之初始值；接著，輸入一個訓練範例之輸入向量 X，與目標輸出向量 T；最後，分別根據公式（一）及公式（二）計算隱藏層輸出向量H_k及推論輸出向量Y_j。

$$net_h = \sum_i W_xh_{ih} \times X_i - \theta_h_h \quad 及 \quad net_j = \sum_h W_hy_{hj} \times H_h - \theta_y_j \qquad 公式（一）$$

$H_h = \frac{1}{1+\exp(net_h)}$ 及 $Y_j = \frac{1}{1+\exp(-net_j)}$ 公式（二）

在求得隱藏層輸出向量H_k及推論輸出向量之值Y_j後，從公式(三)及公式(四)計算輸出層差距量δ_j及隱藏層差距量δ_k。

$\delta_j = Y_j(1-Y_j)(T_j - Y_j)$ 公式（三）

$\delta_h = H_h(1-H_h)\sum_j W_hy_{hj}\delta_j$ 公式（四）

之後可計算輸出層（隱藏層）之加權值矩陣修正量$\Delta W_hy(\Delta W_xh)$，和偏權值向量修正量$\Delta\theta_y(\Delta\theta_h)$

$\Delta W_hy_{hj} = \eta\delta_j H_h - hy_{hj}$ 及 $\Delta W_xh_{ih} = \eta\delta_h X_i$ 公式（五）

$\Delta\theta_y_j = -\eta\delta_j - y_j$ 及 $\Delta\theta_h_h = -\eta\delta_h$ 公式（六）

求得修正量後，更新輸出層（隱藏層）加權值矩陣$W_hy(W_xh)$，及偏權值向量$\theta_y(\theta_h)$。

$W_hy_{hj} = W_hy_{hj} + \Delta W_hy_{hj}$ 及$W_xh_{ih} = W_xh_{ih} + \Delta W_xh_{ih}$ 公式（七）

$\theta_y_j = \theta_y_j + \Delta\theta_y_j$ 及 $\theta_h_h = \theta_h_h + \Delta\theta_h_h$ 公式（八）

重覆公式 3 至公式 7，直至收斂（誤差不再有明顯變化）或執行一定數目的學習循環為止。接著，進入倒傳遞類神經網路之回想過程；首先，進行設定網路參數。讀入加權矩陣W_xh與W_hy，以及偏權值向量θ_h與θ_y後，輸入一個測試範例之輸入向量 X。計算隱藏層輸入向量H_k及推論輸出向量Y_j。

$net_h = \sum_i W_xh_{ih} \times X_i - \theta_h_h$ 及 $net_j = \sum_h W_hy_{hj} \times H_h - \theta_y_j$ 公式（九）

$H_h = \frac{1}{1+\exp(-net_h)}$ 及 $Y_j = \frac{1}{1+\exp(-net_j)}$ 公式（十）

肆、 結果與分析

本研究首先由科技類基金中選出基金樣本，並根據基金成立期間、基金持股狀況、基金投資標的、投資區域相似度較高的三支科技基金－保誠高科技基金、華南永昌前瞻科技基金、匯豐龍騰電子基金為研究標的。資料範圍從民國 86 年 1 月至民國 95 年 12 月，共 120 筆月資料。所有可能影響基金績效及淨值之各種因素，大致分為幾大類：基金評估指標，其項目包括基金淨值標準差、資訊率（大、細)、夏普、貝塔值、崔納等，景氣訊號燈，貨幣金融指標如 M1A、M1B、M2 等，此外，工業與電子業生產指數如電子、電機、機械、資訊外銷訂單等共計 36 個變數，各因素之中英文對照表如表一所示。

表一、影響因素名稱對照表

中文	English	中文	English
標準差	STD	實際國內生產毛額	GDP
資訊率(大)	IR_L	外銷訂單總額	ET
資訊率(細)	IR_S	電子產品外銷訂單	EET
夏普	Sharpe	電機產品外銷訂單	EMET
崔納	TREYNOR	機械產品外銷訂單	MET
貝塔值	Beta	資訊與通訊產品外銷訂單	ICET
同時指標綜合指數	GI	製造業生產指數	MPI
景氣燈號	ECI	工業生產總指數年增率	IP_YEAR
通貨淨額(日)	M1A_DAY	外匯存底	FR
通貨淨額(年)	M1A_YEAR	當期國民生產毛額	GNP
狹義貨幣供應量(日)	M1B_DAY	躉售物價指數	WPI
狹義貨幣供應量(年)	M1B_YEAR	消費者物價指數年增率	CP_YEAR
廣義貨幣供應量(日)	M2_DAY	進口物價指數年增率	IM_YEAR
廣義貨幣供應量(年)	M2_YEAR	出口物價指數年增率	EX_YEAR
流動性負債年增率	CL	領先指標綜合指數	ML_INDEX
股票市場股價指數	STOCK_INDX	物價總指數	PI
台股加權指數	TSE	店頭市場加權指數	OTC
躉售物價指數變動率	WP_DELTA	躉售物價指數年增率	WP_YEAR

表二、三支基金之逐步迴歸模式摘要(以%為單位)

保誠高科技基金				華南永昌前瞻科技基金				匯豐龍騰電子基金			
模式	R-Sq	R-Sq（adj）	C-p	模式	R-Sq	R-Sq（adj）	C-p	模式	R-Sq	R-Sq（adj）	C-p
18	95.72	95.24	36.0	17	97.26	96.93	31.8	18	95.38	94.76	32.6
19	95.87	95.36	33.4	18	97.37	97.02	29.1	19	95.31	94.74	32.3
20	96.03	95.50	30.5	19	97.45	97.08	27.4	20	95.68	95.1	24.6
21	96.25	95.71	26.0	20	97.61	97.23	22.3	21	95.94	95.35	19.6
22	96.42	95.86	22.7	21	97.57	97.22	21.9	22	96.04	95.42	19.1
23	96.52	95.94	21.7	22	97.65	97.29	20.1				
24	96.49	95.95	20.5	23	97.72	97.34	19.2				
25	96.46	95.95	19.4	24	97.83	97.45	16.0				
26	96.57	96.03	18.3	25	97.79	97.42	15.9				
27	96.54	96.05	16.9	26	97.86	97.48	14.6				
28	96.68	96.16	14.9	27	97.94	97.55	13.2				
29	96.78	96.24	13.8	28	97.98	97.58	13.2				
30	96.85	96.29	13.5								

于逐步迴歸中採用 *R-Sq* 值高於 0.95 以上及 *C-p* 值 36 以下之門檻值進行篩選後，保誠高科技基金共有 13 個模式、華南永昌前瞻科技基金共有 12 個模式、匯豐龍騰電子基金則有 5 個模式，三支基金的最適迴歸模式組合皆從模式 17 或 18 開始，且當 *C-p* 值愈低時 *R-Sq*（*adj*）值愈高，逐漸形成趨於穩定的狀態（如表二所示）。此外，從此表亦可看出，三支基金的 *C-p* 值與 *R-Sq* 值並不一定隨模式指標而增加。

當 VIF 值大於 10 時，可能所表示共線性在某些程度上影響了最小平方值（Least Squares Estimates）的計算，因此本研究根據第三節所描述的方式，將 VIF 大於 10 的變數刪除再進行迴歸分析，直到該模式共線性為最低，且能夠保有模式之顯著性。表三為此兩種模式在經過本研究之變數篩選後的結果，其中 TSE（台股加權指數）、IM_YEAR（進口物價指數年增率）、EX_YEAR（出口物價指數年增率）為三家基金公司共同具有之變數。

表三、模式一與模式二之變數組合

保誠高科技		華南永昌前瞻科技		匯豐龍騰電子	
模式一	模式二	模式一	模式二	模式一	模式二
BETA	BETA	BETA	CL	CP_YEAR	EET
CL	CL	CL	CP_YEAR	EET	EMET
EC	EMET	CP_YEAR	EX_YEAR	EMET	ET
EET	EX_YEAR	EMET	FR	ET	EX_YEAR
EMET	GDP	EX_YEAR	GDP	EX_YEAR	GI
EX_YEAR	ICET	FR	IM_YEAR	GDP	IM_YEAR
GDP	IM_YEAR	GDP	M1A_YEAR	GI	IR_S
ICET	IR_L	GNP	M2_YEAR	IM_YEAR	M2_DAY
IM_YEAR	IR_S	ICET	MET	IR_S	ML_INDEX
IP_YEAR	M2_YEAR	IM_YEAR	MPI	M1A_DAY	STD
IR_L	ML_INDEX	M1A_DAY	PI	M2_DAY	STOCK_IND
IR_S	MPI	M1A_YEAR	Sharpe	M2_YEAR	TSE
M1A_YEAR	STD	M2_DAY	TREYNOR	ML_INDEX	WP_YEAR
M2_DAY	TSE	M2_YEAR	TSE	PI	WPI
M2_YEAR	WP_DELTA	MET		STD	
ML_INDEX		MPI		STOCK_IND	
MPI		PI		TSE	
PI		Sharpe		WP_DELTA	
STD		TREYNOR		WP_YEAR	
TREYNOR		TSE		WPI	
TSE		WP_DELTA			
WP_DELTA		WP_YEAR			
WPI					

表四為針對類神經網路的隱藏層節點數進行敏感度分析之結果摘要，隱藏層數為二時表現較單一隱藏層為佳，而在三支基金當中，以華南永昌前瞻科技基金在第一及第二層所使用的節點數最少。

表四、隱藏層節點數比較表

	模式一		模式二		模式三	
基金名稱	隱藏層 1	隱藏層 2	隱藏層 1	隱藏層 2	隱藏層 1	隱藏層 2
保誠	5	10	10	10	20	5
華南	5	5	10	5	5	5
匯豐	20	10	20	10	5	10

在針對學習循環次數進行分析時發現，三支基金的循環次數超過 1000 次之後，即有過度學習情況發生。所以從表五中可看出，保誠高科技基金之模式一及模式二循環次數分別達 300 和 500 次時為效果最佳；華南永昌前瞻科技基金之循環次數在模式一及模式二之循環次數皆為 500 次時為效果最佳；匯豐龍騰電子基金之循環次數在 100 次時則為最佳狀況。表六為學習速率的敏感度分析，發現三家基金中僅匯豐龍騰電子基金在兩個模式的學習速率稍有差異，其他兩家學習速率皆為一致。而慣性因子之敏感度分析結果則如表七所示。

表五、循環次數結果

基金名稱	模式一	模式二
保誠	300	500
華南	500	500
匯豐	100	100

表六、學習速率結果

基金名稱	模式一	模式二
保誠	1	1
華南	1	1
匯豐	0.5	1

這三支基金根據其敏感度分析之結果，分別設定類神經網路的最適參數組合，之後透過各模式各自求得出三支基金的預測值，且加以計算出 20 筆測試資料之絕對誤差率，其結果如表八所示。模式一與模式二之表現極為接近，對保誠高科技基金及匯豐龍騰電子基金而言，模式二的表現略勝一籌，而華南永昌前瞻科技基金則是以模式一表現略佳。至於對照組之模式三表現皆不盡理想，尤其在匯豐龍騰電子基金時，平均誤差率更高達 55.75%。

表七、慣性因子結果

基金名稱	模式一	模式二
保誠	0.1	0.1
華南	0.5	0.5
匯豐	0.1	0.3

表八、絕對誤差率比較表

保誠高科技			
	模式一	模式二	模式三
平均	10.55%	10.18%	16.40%
最大值	29.76%	29.22%	47.52%
最小值	0.96%	0.44%	0.27%
華南永昌前瞻科技			
	模式一	模式二	模式三
平均	6.21%	8.27%	18.00%
最大值	19.74%	25.50%	31.41%
最小值	0.32%	0.63%	3.00%
匯豐龍騰電子			
	模式一	模式二	模式三
平均	13.63%	13.04%	55.75%
最大值	33.81%	34.41%	97.59%
最小值	1.21%	1.09%	4.86%

伍、 結論

本研究針對兩種模式之學習參數做敏感度分析以增進類神經網路模式之準確度。在隱藏層節點數的分析結果以使用兩層為有效的方式，其節點數又以華南永昌前瞻科技基金所使用的節點數最少。在學習循環次數分析結果得知，三支基金之學習循環次數超過 1000 次後，皆會有過度學習情況發生。而循環次數結果以匯豐龍騰電子為佳。針對學習速率分析之結果顯示保誠及華南兩支基金的結果可看出學習速率設定為 1 是最佳的選擇。針對慣性因子分析之結果得知保誠高科技基金之慣性因子值比較低，其次是匯豐龍騰電子基金，最後是華南永昌前瞻科技基金。最後，本研究將三支基金之實際值與測試值以絕對誤差率做一比較，發現華南永昌前瞻科技基金在平均值、最大值、最小值皆較其他兩家基金偏低。由於，該研究乃藉由過去

歷史資料、數據做預測分析評估，由此可知，華南永昌前瞻科技基金之預測為準確度最高，亦明顯發現匯豐龍騰電子基金在模型預測上，顯示與其它兩家基金差異最大。

陸、 參考文獻

1. 張振魁，以類神經網路提高股票單日交易策略之獲利，國立中央大學資訊管理學系碩士班，2000
2. 陳景堂，統計分析SPSS for Windows入門與應用第四版，儒林圖書公司，2003
3. 游建欣，運用適應性網路模糊推論系統於台灣股票加權指數預測之研究，東吳大學經濟學系碩士論文，2006
4. 黃綺年，統計方法與類神經網路應用於國內開放式股票型基金投資績效分類及投資報酬率預測之研究，國立成功大學統計學研究所，2004
5. 楊孟龍，類神經網路於股價波段預測及選股對應用，國立中央大學資訊管理學系碩士班，2000
6. 劉濠葦，運用類神經網路於共同基金績效之研究，國立東華大學企業管理系碩士班，2000
7. Bollen, N. and J. A. Busse, 2000, On the timing ability of mutual fund managers. Unpublished working paper, University of Utah, Salt Lake City, UT
8. Christopherson, J. A., W. E. Ferson, and D. A. Glassman, 1998, “Conditional measures of performance and persistence for pension funds,” Research in Finance 16, 1-46.
9. Elton, E. J., M. J. Gruber, S. Das. and M. Hlavka, 1993, “Efficiency with costly information: A reinterpretation of evidence from managed portfolios,” Review of Financial Studies 6, 1-22.
10. FundDJ基智網：http://www.funddj.com/y/yFund.htm
11. Gerald Keller 著，陳相如 譯，統計學，東華書局，2006