

整合多元適應性雲形迴歸與 支援向量迴歸建構股價預測模式之研究

呂淑瑜¹ 呂奇傑^{2*}

摘要

預測股價變化是一項具挑戰性的任務。本研究整合支援向量迴歸（support vector regression, SVR）與多元適應性雲形迴歸(multivariate adaptive regression splines, MARS)建構股價預測模式。SVR 為近年來廣被重視與應用於各領域的預測方法之一，然而其卻不具有變數篩選的能力。由於在建構股價預測模式時，預測變數的選擇通常會影響模式的績效，因此研究使用處理速度快且具有選擇變數能力 MARS 做為 SVR 的前處理工具，提出一個有效的股價指數預測模式。本研究以台灣加權股價指數為實證對象。實證結果顯示，所提之整合 MARS 與 SVR 模式無論是在預測誤差或是趨勢預測準確度的表現上均較多元線性迴歸(MLR)、直接使用 MARS 及直接使用 SVR 模式為佳。並且在整合模式中，透過 MARS 能夠從眾多預測變數中篩選出重要的預測變數，藉由分析變數選擇的結果，可對所建構的預測模式有更深入的理解，提供投資者更多的參考資訊。

關鍵詞： 股價預測、多元適應性雲形迴歸、支援向量迴歸

¹健行科技大學工業管理系

²健行科技大學工業管理系

* 聯絡作者: jerryly@uch.edu.tw; chijie.lu@gmail.com

Combining MARS and SVR for Stock Price Forecasting

Shu-Yu Lu³, Chi-Jie Lu^{4*}

Abstract

In a financial company, stock index forecasting is one of the most important and challenging tasks. In this study, we combined multivariate adaptive regression splines (MARS) and support vector regression (SVR) for stock price forecasting. The proposed model first uses the MARS to select important forecasting variables. The obtained significant variables are then served as the inputs for building the SVR forecasting model. Experimental results from TAIEX revealed that the obtained important variables from MARS can improve the forecasting performance of the SVR models. The proposed combined model outperforms the results of using single multivariate linear regression (MLR), single SVR and single MARS models and hence provides an efficient alternative for stock index forecasting.

Keywords: Stock price forecasting, multivariate adaptive regression splines, support vector regression,

³Department of Industrial Management, ChienHsin University and Science and Technology.

⁴Department of Industrial Management, ChienHsin University and Science and Technology.

*Corresponding author: jerry.lu@uch.edu.tw; chijie.lu@gmail.com

1. 前言

隨著世界經濟與金融體系的結構轉變，金融市場的本質產生了變化，包括交易的數量與範圍，以及市場結構和運作模式，都與過往不同。國際金融交易的自由化使得人們有更多資產配置與投資的選擇，但也意味著這些金融商品身處於更複雜的全球環境，使其價格更難以預測。在變化如此快速的複雜環境下，資產的配置或轉換，需在一瞬間就做出決策，機會稍縱即逝，也因此造就財務時間序列預測的蓬勃發展。但由於金融商品的交易特性以及價格容易受政府政策、景氣循環、國際市場的漲跌及物價膨脹等眾多因素影響，使財務時間序列成為一非常複雜、極具挑戰性以及難以預測的時間序列資料。

在財務時間序列的預測技術方面，人工智慧方法 (artificial intelligence, AI) 由於不須對問題做模型假設就可以從大量的樣本資料中學習，然後對母體做出一般性推論，得出一個更接近真實的預測結果，已廣泛的應用於各種時間序列預測問題上 (Zhang et al. 1998)。常見的人工智慧技術為類神經網路 (artificial neural networks, ANN) (Nag and Mitra 2002; Chen et al. 2003; Chen and Leung 2004; Roh 2007) 與支援向量迴歸 (support vector regression, SVR) (Mohandes et al. 2004; Pai and Lin 2005; Karras and Mertzios 2004)。其中，SVR 與 ANN 最大不同在於 ANN 是藉由極小化實驗風險 (empirical risk or training error) 來尋求局部 (local) 最佳解，而 SVM 則是追求結構風險 (structural risk or generalization error) 最小化以求達到全域最佳解，因此，理論上 SVR 可以獲得比 ANN 更好的結果 (Ince and Trafalis, 2006)。是以 SVR 在近年來越來越受到矚目，本研究以 SVR 作為財務時間序列預測工具。

一般而言，在建構財務時間序列預測模型時，首先需決定或選擇許多與預測標的有關的預測變數，如股價指數、匯率及利率等，之後再透過探討預測變數和預測標的之間的關係，建構出一最適的預測模型。但是，並非所有的預測變數皆為那麼重要，換言之，變數的重要程度是有差別的。由於預測變數的選擇結果通常會影響到預測模型建構之準確性，過多的預測變數會增加模式建構的困難度，過少的預測變數可能無法建構出有效的預測模式，如何選擇適當的預測變數來建構預測模式一直是財務時間序列預測中重要議題之一。然而，無法篩選重要預測變數是 SVR 的主要缺點之一。有鑑於此，本研究使用一個新興、處理速度較快且具有選擇變數能力的「多元適應性雲形迴歸」(multivariate adaptive regression splines, MARS) 方法做為 SVR 的前處理工具，提出整合 MARS 與 SVR 的財務時間序列預測模式。所提方法將先利用 MARS 對預測變數進行篩選，之後再利用篩選後的變數建構 SVR 預測模式。

MARS 為近年來廣泛地被使用在各種分類問題(classification)的資料探勘技術(許峻源 2001; 唐筱菁 2002; Zareipour et al. 2006; 朱慧祺 2008)，其基本概念為將非線性資料大概分成幾個區間，每一區間內的資料關係期望是線性的關係，之後在每一區間內，皆利用一基本方程式所構成的線性函數來描述資料間的關係。區間與區間的隔點稱為節點(knot)，換言之，即為每一段基本方程式斜率改變或資料行為改變之處。在將數段斜率不同的基本方

程式累加運算之後，便可以組合出一個較具彈性的非線性函數預測模型，使用者可以用這個模型解決多元且複雜的非線性資料問題(Friedman, 1991)。MARS 和其他軟性計算(soft computing)技術相比較，所得到的預測模型比較準確、誤差值也相對小許多(Abraham et al. 2001)。而且 MARS 本身具有較佳的篩選變數能力，可以藉此分析原始資料數據之變數影響程度的差異，對使用者而言，方便其對資料的解讀並且提供更高的使用價值。

MARS 雖然已常被應用於分類問題上，並獲致良好的成果(Griffin et al. 1997; Chou et al. 2004)，但由文獻中可知，其較少用於財務時間序列預測問題。因此，本研究除了探討所提的整合 MARS 與 SVR 模式的預測績效外，也會討論 MARS 用於建構財務時間序列預測模式的效果。在實證的對象及資料上，由於股票長久以來一直是人們投資理財時最常用的投資標的，股價指數的預測也是為實務界與學術界最常探討的議題，因此本研究以台灣加權股價指數作為實證的對象。而在模式結果的比較上，所提之整合 MARS 與 SVR 模式的預測結果將與直接 SVR 模式（亦即直接利用預測變數建構 SVR 預測模式，不經過變數篩選）、直接 MARS 模式及多元線性迴歸(Multiple linear regression, MLR)之預測結果進行比較，以驗證整合 MARS 與 SVR 模式的有效性，進而作為擬定相關投資及金融操作之重要參考依據。

2.文獻回顧與探討

2.1 多元適應性雲形迴歸(MARS)

MARS 係於 1990 年代早期由 Friedman 所提出(Friedman 1990; 1991)，主要是用來解決多元資料問題。其在經濟、醫學、地球科學等各領域皆普遍被應用，在商業上則主要被使用於資料探勘的分類(Classification)問題，較少被用於預測相關議題，尤其在財務時間序列預測的議題上，更為少見，以下針對 MARS 於分類與預測應用上的相關文獻整理。

De Veaux et al. (1993)運用 MARS 在地球科學的領域上，進行勘測南極洲附近的地形。Nguyen-Cong et al. (1996)則使用 MARS 進行遺傳學上做基因工程的研究。許峻源(2001)以台灣某大銀行提供之信用卡客戶為對象，運用傳統統計鑑別分析、MARS 與 ANN 建構分類模式分辨信用卡申請者的信用優良與否。其研究結果發現，在變數間關係較複雜的信用卡資料上，使用 MARS 之判別結果明顯比單純使用鑑別分析與 ANN 好；此外若把藉由 MARS 所分析出的結果，再置入 ANN 提供一個較佳的起始原點建立出來的分類模式更為精確與快速。

Abraham et al. (2001)的研究指出降雨量(rainfall)為非線性的自然現象，故藉由使用軟性計算(soft computing)技術包括 Evolving Fuzzy Neural Network (EFuNN)、Artificial Neural Network using Scaled Conjugate Gradient Algorithm (ANNSCGA)、Adaptive Basis Function Neural Network (ABFNN)、General Regression Neural Network (GRNN)以及 MARS 去進行 Kerala state 地區的降雨量預測，實驗以依據過去 87 年的歷史資料經由上述工具去做降雨量的預測，其結果發現 MARS 是較好的預測工具而且表現皆比其他軟性計算來的好，值得一提的是 MARS 的速度最快而且誤差值最小(lowest RMSE)。黃明輝(2002)運用 MARS

與 ANN 進行探勘債券型基金是否具有擇時能力,欲了解 MARS 在分類問題上是否比 ANN 提供更為精確的分類模型。其研究結果顯示,同為 MARS 具有較佳的判別能力與運算快速的優點,同時 MARS 尚具有篩選變數的功能,投資人藉此了解了某些變數為影響擇時的重要因素,提供使用者寶貴的資訊。唐筱菁(2002)以建立一整合財務比率及智慧資本相關指標的企業危機診斷模型為目的,整合 MARS 與倒傳遞神經網路(BPN)之建立兩階段分類模式。

Ko and Osei-Bryson (2004),運用 MARS 在保健產業(health care industry)中去探討 IT 投資在生產力上的衝擊(impact of information technology investments on productivity),研究指出藉由 MARS 提供了豐富的分析,包括 IT 投資和生產力間的關係,這是傳統統計迴歸所無法提供的。Zareipour et al. (2006)提到 MARS 為一適應性無母數迴歸方法(adaptive non-parametric regression approach)並將其應用在安大略湖每小時能源價格(hourly Ontario energy price, HOEP)的預測上,研究結果顯示,MARS 的提供了準確的預測,特別是在夏季高峰需求週(summer peak-demand week)。Zhou and Leung (2007)使用 MARS 進行物件導向軟體維護性做預測(object-oriented software maintainability),並且比較多元線性迴歸模型(multivariate linear regression models)、ANN、迴歸樹模型(regression tree models)、支援向量機(support vector models)的預測結果,研究發現 MARS 的預測準確度是最佳的。Kiran and Ravi (2008)藉由使用如多元線性迴歸和 MARS 以及人工智慧如 BPN、TreeNet 等方法,建構模型去準確預測軟體的可靠度,結果顯示無論是 MARS 或是 BPN 這類的工具都有相當好的表現。

由以上文獻回顧可以發現,MARS 因其快速、準確的優點,無論是在學術或是實務上之資料探勘的應用都非常的普遍與廣泛。唯目前在商業金融的資料預測上之文獻相對較少。

2.2 支援向量迴歸(SVR)

在機器學習領域中,SVM 是以統計學習理論(statistical learning theory)為基礎,運用風險最小化之原理,以有效區隔不同種類的資料。SVM 目前已成功應用在許多領域中(吳智鴻 2004)。除了應用於處理分類問題之外,Vapnik et al. (1997)也將 SVM 的觀念用於迴歸問題中,提出了支援向量迴歸模式,並且廣泛的用於各種預測問題中(Vapnik 1999; Koike 2004; 徐美珍 2004)。

Tay and Cao (2001)針對五種期貨商品,包括 S&P 500 指數期貨、聯邦政府 30 年期公債、聯邦政府 10 年期公債、德國 10 年期政府公債及法國政府指數期貨,使用 BPN 和 SVR 兩種工具建構預測模式並將結果進行比較。研究結果顯示,除了在德國十年期政府公債,SVR 有優於 BPN 之結果外,其餘皆以 BPN 之預測效果為優。Kim(2003)利用報酬率等技術性指標作為輸入變數,配合 SVR、ANN 及案例推論法(Case-based reasoning)做為預測工具,建構每日韓國綜合股價指數(KOSPI)收盤價漲跌的模型,結果顯示 SVR 有最佳之預測結果,能準確預測 57.83%的漲跌。Mohandes et al. (2004)利用 SVR 及多層感知器(Multilayer perceptrons, MLP)於風速預測,並以 RMSE 作為評估指標,結果發現 SVR

之表現優於 MLP。Pai and Lin (2005) 利用 SVR、SARIMA (Seasonal ARIMA) 及廣義回歸神經網路 (General regression neural network, GRNN) 預測具有季節性的工業機械總產值，結果顯示 SVR 優於 GRNN 及 SARIMA 模型。

Karras and Mertzios (2004) 利用 SVR、MLP 及 VAR 建構超音波圖片中心內膜邊界之時間序列預測模式，實驗結果顯示在短期動態中，三種預測模型表現相當；然而在長期動態中，SVR 之表現最佳。許書瑤 (2007) 結合粒子群最佳演算法 (particle swarm optimization, PSO) 與 SVR 於財務時間序列之預測，並以日經 225 指數以及台灣加權股價指數做實證研究。結果顯示以 PSO 找出 SVR 中的最適參數組合，可以提高預測模型的準確度、穩健性以及降低參數搜尋時間，與單純 BPN 和單純 SVR 相比，有較佳的預測表現。Ince and Trafalis (2006) 提出兩階段的匯率預測模式，結合有母數的 (parametric) 工具像是 ARIMA 與 VAR (vector autoregressive) 與無母數的工具像是 ANN 與 SVR，首先利用有母數的工具篩選引入到無母數模型的變數，然後以無母數的工具進行預測。結果顯示混合模式優於單純使用 ARIMA 或 VAR，而結合 SVR 的混合模式較結合 ANN 的模式為佳。

3. 研究方法

3.1 多元適應性雲形迴歸

多元適應性雲形迴歸 (multivariate adaptive regression splines, MARS) 是一個多變量、非線性及無母數迴歸處理方法，其基本概念是運用數段基本方程式 (basis function, BF) 加總組合出一個較具彈性的 MARS 預測模型 (global MARS model，見式(1)) (Friedman 1991)。

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m B_m(x) \quad (1)$$

其中變數 a_0 與 a_m 皆為參數值，其主要是賦予類似迴歸係數的功能， M 為 BF 的個數，藉由評估準則決定； $B_m(x)$ 乃通用模型中後段累乘的部分，亦即基本方程式 (BF)。

MARS 的主要目的是欲結合遞迴分割 (recursive partitioning) 與雲形適應 (spline fitting) 兩項互補性分析技術。遞迴分割最大的優點乃由於遞迴分割其區域變數 (local variables) 子集合的選擇策略，使其具有適應性 (adaptability) 以及高度動態的計算能力，然而其模型缺乏連貫性 (continuity) 以及無法捕捉線性或低維度資料間的簡單函數關係為其缺點；另一方面，雲形適應的優點為可以產生連續型衍生變數與連續模型，但可惜的是在處理函數關係相對簡單的高維度資料時，須使用大量的運算方程式，以至於分析效率較差，也就是所謂的「維度詛咒 (curse-of-dimensionality)」 (Friedman 1991)。而突破此兩項工具的限制並結合其優點的正是多元適應性雲形迴歸。

上述提到，MARS 的基本概念為將非線性資料大概分成幾個區間，並且期望每一區間內的資料關係是線性的關係，之後在每一區間內，皆利用一基本方程式所構成的線性函數來描述區間內資料間的關係。區間與區間的隔點稱為節點 (knot)，換言之，即為每一段基本方程式斜率改變或資料行為改變之處。在將數段斜率不同的基本方程式累加運算之後，

便可以組合出一個較具彈性的非線性函數預測模型，使用者將以此解決多元且複雜的非線性資料問題(Friedman 1991; Steinberg et al. 1999)；此外，MARS 要求以最適合的解釋變數來建立各個區間的基本方程式，即為盡可能採取具有足夠貢獻性的 BF(Steinberg et al. 1999)。

3.2 支援向量迴歸

支援向量機(Support vector machine, SVM)是以統計學習理論(statistical learning theory)為基礎，運用結構風險最小化原理(structural risk minimization inductive principle)以及最大化在群體間之點的邊界(margin)所提出的機器學習系統(Vapnik et al. 1997)。SVM 主要用來解決線性二次規劃(quadratic programming)問題，並且已經廣泛的應用於解決分類的問題。而在處理預測問題上，Vapnik et al. (1997) 利用 ε -insensitivity 損失函數將 SVM 的技術運用在迴歸問題中，而提出支援向量迴歸(support vector machine regression)模型。

為說明 SVR 模式，首先考慮一個典型的迴歸模式，其目的是估計一個未知方程式 $r(x)$ 來預測一個未知的值 q ，兩者的關係為， $q = r(x) + \delta$ ，其中 δ 是隨機獨立並且平均數為零的隨機誤差 (random error)， x 是多變數的輸入變數， q 是單一 (scalar) 的輸出值。這個方程式可以用有限個數的訓練樣本來建立： $(x_i, q_i), (i = 1, \dots, n)$ 。在迴歸問題中可分為線性迴歸與非線性迴歸問題，線性迴歸問題是較易處理，並且已有相當多的技術被提出來與成功的應用；相對而言，非線性迴歸則較不容易處理。SVR 主要就是被發展來處理非線性迴歸的議題的工具，其主要概念是將一個在低維度輸入空間 (input space) 中非線性迴歸之問題，轉換成在高維度特徵空間 (feature space) 中的線性迴歸問題。因此在 SVR 中，輸入變數 x 首先被轉換 (map) 到一個高維度的特徵空間 (F) 中，在這特徵空間中，輸入變數可以被修正為與輸出變數有線性的關係。以數學模式表示，SVR 模式可以表示為 (Vapnik 1999)

$$f(x) = (\omega \cdot \Phi(x)) + b \quad (2)$$

其中， ω 為權重向量， b 是為偏差值 (bias)， $\Phi(x)$ 為一轉換函數，用以將輸入變數 x 線性的方式轉換到高維度的特徵空間中， $(\omega \cdot \Phi(x))$ 則用來描述在特徵空間 F 中的值。

在迴歸問題中，可以在訓練資料用最小化實驗風險 (minimization of empirical risk) 的準則，估計迴歸模式中的參數。典型被用在最小化實驗風險中的損失函數包含有絕對值誤差及平方誤差等，但在 SVR 中，Vapnik et al. (1997) 則使用了一個稱為 ε -insensitivity 的損失函數(L_ε)

$$L_\varepsilon(f(x) - q) = \begin{cases} |f(x) - q| - \varepsilon & \text{for } |f(x) - q| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

其中， ε 代表一個參數，用來定義一個圍繞在迴歸方程式 $f(x)$ 周圍的管狀區域。這些管狀區域就是所謂的 ε -insensitivity 區域，當預測值落在管狀區域內時其損失為零，亦即當預測誤差小於 ε 時，其損失成本會相當於零；反之，當預測值落在區域外時，其損失相當於預測值跟區域邊界的差，可以加入寬鬆變數 (Slack variables, ξ, ξ^*) 來表示之，

$$|q_i - f(x_i)| - \varepsilon = \xi_i, \quad |q_i - f(x_i)| - \varepsilon = \xi_i^*, i = 1, \dots, n$$

其中 ξ 為高於區域邊界的訓練誤差值， ξ^* 為低區域邊界的於訓練誤差值。

4.實驗結果

為了驗證本研究所提之結合 MARS 與 SVR 模式(簡稱 MARS-SVR)的預測能力，在實驗部分將以台灣加權股價指數(Taiwan stock exchange capitalization weighted stock index, TAIEX)作為研究標的，對 TAIEX 收盤價格進行預測。並且將 MARS-SVR 模式之預測結果，與直接 MARS 模式、直接 SVR 模式及多元線性迴歸(MLR)模式之結果進行比較，以驗證所提模式的有效性。

在台灣加權股價指數(即 TAIEX)之收盤資料方面，所蒐集的資料期間為 2004 年 1 月 27 日至 2008 年 7 月 23 日共 1113 個交易日。其中，從 2004 年 1 月 27 日至 2007 年 8 月 27 日共 890 筆資料(佔資料期間 80%)作為訓練樣本；另外保留 2007 年 8 月 28 日至 2008 年 7 月 23 日共 223 筆資料(佔資料期間 20%)作為測試樣本。

在預測變數方面，技術分析是根據市場過去的趨勢特徵或循環軌跡以探索未來的變動，回顧股價預測之相關文獻，技術指標為常用的預測變數(Balachandher et al. 2002 ; Leigh et al., 2005)，故本研究將使用如表 1 所示之 11 個預測變數建構 TAIEX 收盤價之預測模型。相關之公式請參閱 Balachandher et al.(2002)及 Leigh et al. (2005)。

表 1 預測變數表—TAIEX

1. 前一日最高指數	7. 10 日移動平均線
2. 前一日最低指數	8. 5 日相對強弱指標 (5 days-relative strength indicator, RSI5)
3. 當日開盤指數	9. 10 日相對強弱指標 (10 days-relative strength indicator, RSI10)
4. 加權指數	10. 5 日心理線指標 (PSY)
5. 台指期貨收盤價	11. 5 日乖離率 (BIAS)
6. 5 日移動平均線	

在預測衡量績效的衡量準則方面，本研究藉均方誤差(Mean Square Error, MSE)、均方根誤差(Root Mean Square Error, RMSE)、平均絕對百分比誤差(Mean Absolute Percentage Error, MAPE)、方向對稱性(Directional Symmetry, DS)等 4 個常用之指標來評估預測模式的預測結果準確度。其中 MSE、RMSE 與 MAPE 為衡量預測值與實際值之間的誤差(Deviation)，此 3 個指標均為愈小表示預測值與實際值愈接近；而 DS 則用來衡量整體預測的漲跌方向與實際方向的準確率。

4.1、TAIEX 之預測結果

首先在直接 MARS 模式部分，表 2 為 11 個預測變數經由 MARS 變數篩選的結果，由表中可知，經由 MARS 的分析後，原先 11 個預測變數被篩選出較具顯著性之四個變數，

分別為 X6「5 日移動平均線」、X11「5 日乖離率」、X5「台指期貨收盤價」、X8「5 日相對強弱指標」，此四個變數影響預測當日收盤價的重要程度較大，其中又以「5 日移動平均線」之相對重要性達 100% 為最高。其推論為 5 日移動平均線是採五天的股價資料去做平均，而台股每周的交易日恰好為五天，可以藉由一整周的資料來推估未來股價的變動，故 5 日移動平均線的相對重要性最高。

表 2 MARS 之變數篩選結果

變數代碼	變數名稱	Std. dev.	-GCV 值	重要性(%)
X6	5 日移動平均線	0.555	0.001	100.000
X11	5 日乖離率	0.101	0.000864	78.520
X5	台指期貨收盤價	0.044	6.24E-005	6.951
X8	5 日相對強弱指標	0.002	5.64E-005	1.571

表 3 顯示 MARS 共建構出六條基本方程式，經篩選後較具顯著性之基本方程式共四條，並且據以發展出 MARS 最適模型方程式。從最適模型方程式中可發現，使用 MARS 預測時，與變數 X5「台指期貨收盤價」、變數 X11「5 日乖離率」、X6「5 日移動平均線」有正向的關係，且與變數 X8「5 日相對強弱指標」具有反向的關係。

表 3 MARS 模式基本方程式及最適模型方程式

基本方程式：

$$BF1 = \text{MAX}(0, X5 + 0.147);$$

$$BF4 = \text{MAX}(0, X11 + 0.059);$$

$$BF5 = \text{MAX}(0, X6 + 0.071);$$

$$BF6 = \text{MAX}(0, X8 - .379337E-07);$$

MARS 最適模型方程式：

$$Y = -0.499 + 0.077*BF1 + 7.135*BF4 + 0.929*BF5 - 0.008*BF6$$

在直接 SVR 的部分，輸入變數方面，研究中採用 11 個變數作為輸入變數，其在不同參數組合下的訓練及測試結果列於表 4 中，由表 4 中可知當 ϵ 為 2^{-9} 與 C 為 27 時有最小的測試 RMSE 值，是直接 SVR 模式的最佳模式。

表 4 直接 SVR 不同參數組合之結果

ϵ	C	訓練 RMSE	測試 RMSE
2^{-9}	23	0.007357	0.014209
	25	0.007350	0.014491
	27	0.007207	0.013837
	29	0.007350	0.015004
2^{-11}	23	0.007415	0.014106
	25	0.007262	0.013443

2^{-13}	27	0.007561	0.015794
	23	0.007325	0.014040
	25	0.007256	0.013739
	27	0.007262	0.013965

而在 MARS-SVR 模式部分。輸入變數方面，以 MARS 所篩選出的變數 X6「5 日移動平均線」、變數 X11「5 日乖離率」、變數 X5「台指期貨收盤價」以及變數 X8「5 日相對強弱指標」，此四項變數作為輸入變數。在參數選擇部分，則同樣透過試誤法決定。表 5 為不同參數組合下的 MARS-SVR 模式的訓練及測試結果。由表 5 可知，當 $\epsilon=2^{-11}$ 、 $C=217$ 時有最小的測試 RMSE 為 0.008267，為最佳模式。

表 5 結合 MARS 與 SVR 之不同參數組合結果

隱藏層節點數	學習率	訓練 RMSE	測試 RMSE
2^{-9}	23	0.007148	0.013332
	25	0.007029	0.013122
	27	0.004072	0.008324
	29	0.005308	0.010029
2^{-11}	23	0.006210	0.012401
	25	0.008393	0.014177
	27	0.004066	0.008267
	29	0.004896	0.009849
2^{-13}	23	0.007312	0.013697
	25	0.006805	0.012668
	27	0.006739	0.013032
	29	0.005443	0.010902

表 6 為本研究所提之 MARS-SVR、直接 MARS、直接 SVR 及 MLR 之預測模式的預測結果彙總表。由表 6 可知，本研究所提模式的 RMSE 及 MAPE 值分別為 18.35、0.15%，均比 MARS、SVR 及 MLR 模式小。除此之外，相較於 MARS、SVR 及 MLR，所提之 MARS-SVR 模式亦有最高的 DS 值 96.40%，顯示其在股價趨勢的預測上亦較其他三個方法準確。

表 6 各模式於 TAIEX 收盤價的預測結果彙總表

Metrics Models	MSE	RMSE	MAPE	DS
MARS+SVR	336.77	18.35	0.15%	96.40%
MLR	7887.65	88.81	0.71%	83.78%
MARS	2815.89	53.06	0.43%	89.64%

SVR	1066.03	32.65	0.29%	92.79%
-----	---------	-------	-------	--------

4.2 穩健性評估

本研究藉由不同的訓練樣本與測試樣本比率進行分析，以了解在各種不同的訓練與測試組合下各預測模式之穩健性，由表 7 可以發現，在台股加權股票指數之資料下，不管是採用何種訓練與測試比率，皆以 **MARS-SVR** 模式的測試誤差(Testing RMSE)值為最小，且方向正確性(DS)為最高，代表所提預測模式的績效最佳。

表 7 台股加權股票指數之穩健性評估

Relative ratio(%)	Models	TAIEX Testing RMSE	TAIEX Testing DS(%)
60	MARS-SVR	46.71	95.27
	MLR	374.06	86.49
	MARS	430.14	91.22
	SVR	33.83	93.69
70	MARS-SVR	22.40	97.30
	MLR	297.05	84.68
	MARS	316.16	88.89
	SVR	36.87	93.39
80	MARS-SVR	18.35	96.40
	MLR	88.81	83.78
	MARS	53.06	89.64
	SVR	32.65	92.79
90	MARS-SVR	20.12	99.09
	MLR	45.17	95.45
	MARS	17.56	99.09
	SVR	20.77	98.18

另外，關於預測結果的檢定部分。魏克森符號檢定(nonparametric wilcoxon signed-rank test)經常被用在測試相關的兩母體間是否存有顯著差異。因此本研究採用無母數魏克森符號等級檢定，來測試本研究所提之 **MARS-SVR** 模式之預測結果與其它預測模式之結果是否有顯著差異。

表 8 為在台股資料中，各種不同的訓練與樣本組合下，針對不同的預測方法所得的檢定結果。從表 8 可以看出，本研究所提之 **MARS-SVR** 模式，無論是在任何一種訓練與樣本組合下，和其它三種模式相比，其檢定結果相當一致且皆達統計上顯著差異($p < 0.05$)，此表示所提模式的預測績效顯著的較其他方法為佳。

表 8. MARS-SVR 預測結果與其它方法比較之檢定結果表

Models	Relative ratio (%)	TAIEX		
		MLR	MARS	SVR
MARS+SVR	60	-17.071 (.000)	-17.458 (.000)	-12.030 (.000)
	70	-15.486 (.000)	-15.504 (.000)	-14.121 (.000)
	80	-10.725 (.000)	-9.412 (.000)	-8.812 (.000)
	90	-7.438 (.000)	-4.628 (.000)	-4.172 (.000)

*括弧中的數字為相對應的 p-values.

最後在應用 MARS 篩選重要預測變數的部分，從表 9 可以發現，在不同的樣本切割組合下，變數 X6「5 日移動平均線」與變數 X11「5 日乖離率」皆被篩選出來，且重要性相對為高，可以得知在進行台股加權收盤指數預測時，此兩變數相對比較重要；而變數 X1「前一日最高指數」、X5「台指期貨收盤價」、X8「5 日相對強弱指標」、X9「10 日相對強弱指標」在重要程度上雖然相對較低，但也可提供當作預測時的重要參考變數。

表 9 各種樣本切割組合下 MARS 篩選變數彙整表

切割比率	變數	重要性(%)	切割比率	變數	重要性(%)
90:10	X11	100.000	80:20	X6	100.000
	X6	85.771		X11	78.520
	X1	1.306		X5	6.951
				X8	1.571
切割比率	變數	重要性(%)	切割比率	變數	重要性(%)
70:30	X6	100.000	60:40	X6	100.000
	X11	38.899		X11	42.561
				X9	0.276

5 結論

本研究整合具有篩選變數能力多元適應性雲形迴歸以及近來廣受重視的 SVR 建構股價預測模型，並藉由 MARS 篩選變數之能力，洞察到影響較大的預測變數。在以台股加權指數為實證資料下，所提之 MARS-SVR 可以從眾多預測變數中找出重要的預測變數，分別為「5 日移動平均線」、「5 日乖離率」、「台指期貨收盤價」、「5 日相對強弱指標」。由上述結果可以得知，在進行預測模式建構時，並非每個變數都很重要，可以透過幾個重要程度較高的變數進行預測即可。如此，不論對學術研究或是投資人而言皆可以大幅地節省

所需耗費的時間與成本。而在預測結果部分，實證結果顯示，所提方法無論是在預測誤差或是預測準確度上的表現皆較直接 MARS、直接 SVR 以及 MLR 模式為佳，顯示本研究所提之 MARS-SVR 模式的有效性。最後，在未來研究方向上，可以本研究之成果搭配其他投資組合，並且以投資績效為預測結果之衡量，期望進一步獲得增加研究結果之實用性。

參考文獻

1. 朱慧祺(2008)。資料探勘乳部腫瘤存活分析模式之建構。天主教輔仁大學管理學研究所未出版碩士論文。
2. 唐筱菁(2002)。整合財務比率與智慧資本指標建構企業危機預警系統-MARS 與類神經網路之應用。天主教輔仁大學金融研究所未出版碩士論文。
3. 徐美珍，2004，「企業財務危機之預測」，政治大學統計研究所未出版碩士論文。
4. 許書瑤，2007，「結合粒子群最佳演算法與支援向量迴歸於財務時間序列預測模式之建構－以日經 225 指數及台灣加權指數為例」，天主教輔仁大學管理學研究所未出版碩士論文。
5. 許峻源(2001)。類神經網路與 MARS 於資料探勘分類模式之應用。天主教輔仁大學應用統計研究所未出版碩士論文。
6. 黃明輝(2002)。資料探勘在財務領域的應用－以債券型基金之績效評估為例。天主教輔仁大學金融研究所未出版碩士論文。
7. Abraham, A., Steinberg, D., & Philip, N. S. (2001). Rainfall forecasting using soft computing models and multivariate adaptive regression splines. IEEE SMC Transactions: Special, 2001, 1-12.
8. Balachandher, K. G., Fauzias, M. N., and Lai, M. M., (2002), An examination of the random walk model and technique trading rules in the Malaysian stock market, Quarterly Journal of Business & Economics, 41, pp. 81-104.
9. Cao, L. & F. E. H. Tay, (2001), Financial forecasting using support vector machines, Neural Computing & Applications, 10, pp. 184-192
10. Chen, A.-S. and M.T. Leung, 2004, "Regression neural network for error correction in foreign exchange forecasting and trading," Computers and Operations Research, 31, pp. 1049– 1068.
11. Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural network to an emerging financial market: Forecasting and trading the Taiwan stock index. Computers & Operations Research, 30, 901-923.
12. Chou, S. M., Lee, T. S., Shao, Y. E., & Chen, I. F., (2004). Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, 27(1), 133-142.

13. De Veaux, R. D., Gordon, A. L., Comiso, J. C., & Bacherer, N. E. (1993). Modeling of topographic effects on Antarctic sea ice using multivariate adaptive regression splines. *Journal of Geophysical Research*, 98(C11), 20307-20319.
14. Friedman, J. H. (1990). *Multivariate Adaptive Regression Splines*. Stanford University, Technical Report, 102 Rev.
15. Friedman, J. H. (1991). Multivariate Adaptive Regression Splines (with discussion). *The Annals of Statistics*, 19, 1-141.
16. Griffin, W. L., Fisher, N. I., Friedman, J. H., & Ryan, C. G. (1997). Statistical techniques for the Classification of Chromites in Diamond Exploration Samples. *Journal of Geochemical Exploration*, 59, 233-249.
17. Ince, H. and T. B. Trafalis, (2006), "A hybrid model for exchange rate prediction," *Decision Support Systems*, 42, pp.1054–1062.
18. Karras, D. A. and B. G. Mertzios, (2004), Time series modeling of endocardial border motion in ultrasonic images comparing support vector machines, multilayer perceptrons and linear estimation technique, *Measurement*, 36, pp. 331-345.
19. Kim, K. J., (2003), Financial time series forecasting using support vector machines, *Neurocomputing*, 55, pp. 307-319.
20. Kiran, N. J., & Ravi, V. (2008). Software reliability prediction by soft computing techniques. *The Journal of Systems and Software*, Vol. 81, 576-583.
21. Ko, M., & Osei-Bryson, K.M. (2004). Using regression splines to assess the impact of information technology investments on productivity in the health care industry. *Information Systems Journal*, Vol. 14, 43-63(21).
22. Koike A. and T. Takagi, (2004), "Prediction of protein-protein interaction sites using support vector machines," *Protein Engineering Design & Selection*, 17, pp. 165-173.
23. Leigh, W., Hightower, R., and Modani, N., (2005), Forecasting the New York stock exchange composite index with past price and interest rate on condition of volume spike, *Expert Systems with Applications*, 28, pp. 1-8.
24. Meissner, G. & N. Kawano, (2001), Capturing the volatility smile of options on high-tech stocks-a combined GARCH-neural network approach, *Journal of Econ Finance*, 25, pp. 276-293
25. Mohandes, M. A., T. O. Halawani, S. Rehman and A. A. Hussain, (2004), Support vector machines for wind speed prediction, *Renewable Energy*, 29, pp.939-947.
26. Nag, A.K. and A. Mitra, (2002), "Forecasting daily foreign exchange rates using genetically optimized neural networks," *Journal of Forecasting*, 21, pp. 501– 511.
27. Nguyen-Cong, V., Van, D. G., & Rode, B. M. (1996). Using multivariate adaptive regression splines to QSAR studies of dihydroartemisinin derivatives. *European Journal of Medicinal Chemistry*, 31, 797-803.

28. Pai, P. F. and C. S. Lin, (2005), "Using support vector machines in forecasting production values of machinery industry in Taiwan," *International Journal of Advanced Manufacturing Technology*, 27, pp.205-210.
29. Roh, T. H. (2007). Forecasting the volatility of stock price index. *Expert Systems with Applications*, 33, 916-922.
30. Schittenkopf, C., G. Dorffner & E. J. Dockner, (2000), Forecasting time-dependent conditional densities: a semi-nonparametric neural network approach, *Journal of Forecast*, 19, pp. 355-374.
31. Steinberg, D., Bernard, B., Phillip, C., & Kerry, M. (1999). *MARS user guide*. San Diego, CA: Salford Systems Inc.
32. Tay, F. E. H. and L. Cao, (2001), Application of support vector machines in financial time series forecasting, *Omega*, 29, pp. 309-317.
33. Vapnik, V.N., S. Golowich and A. Smola, (1997), "Support vector method for function approximation, regression estimation and signal processing." In: Mozer, M., Jordan, M., and Petsche, T. (Eds.), *Advance in Neural information processing system 9*. Cambridge, MA: MIT Press, pp.281-287.
34. Vapnik, V., (1999), *The nature of statistical learning theory*, 2nd ed, Berlin : Springer.
35. Zareipour, H., Bhattacharya, K., & Cañizares, C.A. (2006). Forecasting the Hourly Ontario Energy Price by Multivariate Adaptive Regression Splines. *Power Engineering Society General Meeting*, 1-7.
36. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural network : The state of the art. *International Journal of Forecasting*, 14, 35-62.
37. Zhou, Y. & Leung, H. (2007). Predicting objected-oriented software maintainability using multivariate adaptive regression splines. *The Journal of Systems and Software*, Vol. 80, 1349-1361.