

利用技術指標與總體經濟變數預測台灣股市股價漲跌之研究

一以支持向量機建構

Forecasting TAIWAN Stock Markets with Technical Indicators and the

Variable of Macroeconomics -Based on Support Vector Machines

林綉月	陳榮昌	陳同孝	林昭明
Hsiu-yueh Lin	Rong-chang Chen	Tung-Shou Chen	Chao-Ming Lin
alin@ntit.edu.tw	rcchens@ntit.edu.tw	tschen@ntie.edu.tw	yebhee@yahoo.com.tw
國立台中技術學院 進修學院	國立台中技術學院 流通管理系	國立台中技術學院 資訊管理系	國立台中技術學院 資訊管理系

摘要

近十年來，SVM(Support Vector Machines)被廣泛運用於資料分類及迴歸處理，支持向量機由於其在許多領域良好的表現而受到矚目。支持向量機中包含一個學習演算法和一個輸入空間，輸入空間內含一個訓練集和一個測試集。藉由訓練集的輸入，學習演算法可以找出一個辨別器，並經由測試集我們可以知道此辨別器的辨別正確度。支持向量機的目的是找出正確度足夠的分類器，針對日後未知的輸入作辨別。基本分析與技術分析是股市投資者使用最廣的分析方法，藉由分析的結果可幫助投資者作出較正確的買賣決策，進而在股市中獲利。本研究即利用 SVM 之迴歸特性，試圖在基本分析與技術分析的理論基礎下，找出台灣股市未來可能之走勢。在 SVM 架構下，透過此兩種指標的預測模式分析，可發現台灣股市並非無法預測，在預測隔月股價漲跌方面，所預測的電子類個股，大多有不錯的準確率。整體而言，股本較小的個股使用技術分析訓練的模型有較佳的準確率，股本較大的個股，兩種方法皆有不錯的準確率(45.5%~69.4%)，但就股本較大的個股而言，基本分析模式的表現較佳，部分個股更高達 60% 以上。

關鍵字：SVM、迴歸分析、基本分析、技術分析、總體經濟變數

Abstract

Support vector machine (SVM) is used to classify data in the input space. SVMs are applied to many fields, and have good performance in a wide variety of applications. In this paper, we first introduce the support vector machines.

The basic analysis and technique analysis both are the methods that often be used by most of the investors, by analytical of the result can help the investors to make the righter decision, then making a profit in stock market. In recent years, SVM extensive be used in the data classification, and regression. This research used the regression characteristic of the SVM, trying to find out the possible trend in the Taiwan stock market.

Through the analytical result found that the Taiwan stock market is can to predict in the aspects of predicting the stock price next month, the electronics stocks chosen have good accurate rate mostly. By all accounts, there have a better accurate rate for the small capital stocks that use the technique analysis, for big capital stocks, two kinds of methods all have good accurate rate (45.5%~69.4%), but the technique analysis is better for big capital stocks, there are a few stocks being up to 60% above.

Keyword: SVM (Support Vector Machines) , Regression Analysis, Technical Analysis, Macro Economic

1.前言

股票市場是社會大眾重要的投資途徑，然而多數的投資者無法都擁有深厚的投資背景與財經知識，導致投資者在判斷與預測股市未來可能之走向時，顯得較無效率，投資風險無法有效規避。因此，如何建構一套簡易有效的分析機制，能讓投資者不再盲目的進出市場，而且在股票市場中有一定的獲利水準，即是本研究的動機與研究方向。現今學者的研究中，用來預測個股走趨的理論非常繁多，舉凡：基本分析（Fundamental analysis）學派[9]、技術分析（Technical analysis）學派[8]、隨機漫步（Random walk）學派[1]等，都有其理論基礎與推論依據，本研究利用基本分析中的總體經濟變數與技術分析中的基本指標，透過支持向量機(Support vector machine 以下簡稱 SVM)模型的建構來預測台灣電子個股未來隔月股價之漲跌，提供投資者決策之參考。

股市預測的方法很多，著名的有：1.時間序列法。時間趨勢有一定的軌跡，投資者可依此軌跡來預測未來股市發展；但時間序列法是採線性組合，對於股市非線性的特性，成效不彰。2.專家系統。此系統是廣徵許多專家的意見，再加以歸類、彙整；但執行的法則易受限制，維護系統比較費時費力。3. 倒傳遞類神經網路。在股市預測中，最常被運用的，其原理乃是運用模擬人類神經元的概念，找出資料中的特定規則，藉以預測股市未來發展，對於股市走趨的預測能力，有不錯的準確率，但有學習速度慢、模型建構階段風險較大的缺點。

有鑑於此，本研究嘗試利用 SVM[14]之特性與優點，來預測未來股市的走向，SVM 具有分類與迴歸的功能，其特性類似於類神經網路，但 SVM 的訓練方式較類神經網路要來得快速及簡單，且訓練時不須給予額外的知識幫助。SVM 的應用相當廣泛，例如人類臉部辨識、影像分類及搜尋、股市分析、文字辨視、語音辨視等，皆有不錯的成果[16][7]。而本研究，針對個電子類股透過基本分析與技術分析指標加以篩選，再利用 SVM 的迴歸功能，分別建構出各自的模型，並比較兩種模式的優劣，以提供給投資者在進行投資決策時，能夠作出較正確的判斷，降低決策風險，並增加投資的獲利率。實驗結果發現，股本較小的個股，使用技術分析訓練的模型，有較佳的準確率。股本較大的個股，使用兩種模型皆有一定的準確率(45.5%-69.4%)，但基本分析的模型準確率優於技術分析模型。此兩種分析法，對於某些特定條件的個股，有較佳準確率的特性，此結果的發現，可做為投資者在分析股本大小時選擇的依據。

本研究共分為五個部分，第一部份為前言。說明研究動機及目的、目前成

果簡介以及本實驗結果簡介。第二部份為 SVM 理論簡介。第三部份為文獻探討。針對國內外相關的文獻作一探討，並分析其所使用的方法。第四部份為本研究的研究方法。針對實驗的方法，作一詳細的介紹，包括資料來源、研究對象與範圍、研究限制、技術指標定義、實際作法及研究流程加以探討。第五部份為實證架構。第六部份為實證結果詳細的分析及探討。第七部份為結論。說明研究的貢獻、後續的研究及建議。

2. SVM 理論

支持向量機(Support Vector Machines ; SVM)在近代的統計習理理論上是一個新的衍生學習系統，學者謝東宏[21]對理論架構有簡淺的整理。實際的應用工作中如：文字分類、手寫辨識、影像分類、生物基因序列分析等，都有不錯的表現。SVM 在 90 年代首先被介紹，立刻造成運用上的突破並加劇理論上分析，現在已成為機械學習(machine learning)與資料開採(data mining)標準工具之一[6]。

典型完美的分類器是建立在最小誤差估計(risk minimization)之上，而經驗的最小誤差估計(Empirical risk minimization)是最常用的技術，其目的是找尋一些參數設定使誤差降到最低：

$$R_{EMP}(\alpha) = \frac{2}{2l} \sum_{i=1}^l |Y_i - f(X_i, \alpha)| \dots\dots\dots(1)$$

經驗的最小誤差函式在(1)式中 α 是可調整的參數， y_i 是吾人所期待的輸出， x_i 是吾人給定的輸入。現在我們就舉一個例子來說明：圖 2.1 所示，是二類別分類的問題，區分平面(separating hyperplanes)將各類別區分開，圖中 C0、C1、C2 都可達到良好的區分類別的效果，而 C0 是最佳的，因為 C0 到邊界 H_1, H_2 的距離最大 [14]。這種型式的學習稱為結構的最小誤差計法 (Structural Risk Minimization, SRM)其目的是學習分類器在待估測的誤差 (expected risk)中如何找最小值。SVM 就是以 SRM 為基礎引申而來。

支持向量分類(Support Vector Classification)的目標是在高維度的特徵空間中找出一個區分平面(separating hyperplanes)。而此區分平面(separating hyperplanes)可以找出最佳的邊界。

如圖 2.1 的例子來說明，給定一些線性的訓練樣本 $S = ((X_1, y_1), \dots, (X_l, y_l))$ ，亦即訓練資料 $\{X_i, y_i\}, i = 1, 2, \dots, l, y_i \in \{-1, 1\}, X_i \in R^d$ ，假定我們可以找到一些區分平面(separating hyperplanes)C0，將對應到 $y_i = +1$ 的 X_i (正樣本)和對應到 $y_i = -1$ 的 X_i (負樣本)區分開來，若此區分平面滿足 $w \cdot X + b = 0$ ， w 是區分平面的法向量，則區分平面到原點的距離為 $\frac{|b|}{\|w\|}$ ，假設 d_+ (d_-)是正(負)樣本到區分平

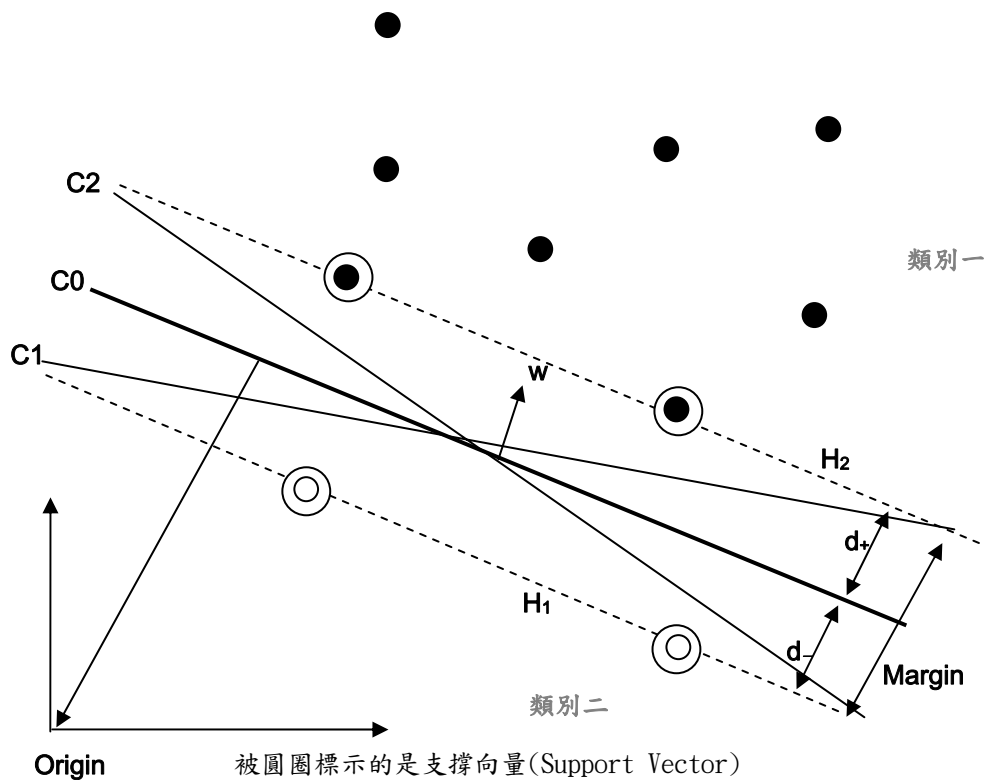


圖 2.1

面最短的距離，定義區分平面的“margin”等於 $d_+ + d_-$ ，對此線性分類的例子來說，支持向量的演算法簡單的說就是找出區分平面最大的“margin”。[2] 假設訓練樣本滿足以下的條件：

$$X_i \cdot w + b \geq +1, \text{ for } \dots y_i = +1 \dots \dots \dots (2)$$

$$X_i \cdot w + b \leq -1, \text{ for } \dots y_i = -1 \dots \dots \dots (3)$$

第(2)(3)式可以結合成

$$y_i(X_i \cdot w + b) - 1 \geq 0, \quad \forall i \dots \dots \dots (4)$$

現在我們考慮滿足(2)式的樣本，在邊界 $H_1: X_i \cdot w + b = 1$ ，其法向量為 w ，到原點的距離為 $\frac{1-b}{\|w\|}$ ；相同在滿足(3)式的樣本，其樣界 $H_2: X_i \cdot w + b = -1$ ，法向量為 w ，到原點的距離為 $\frac{|-1-b|}{\|w\|}$ 。因此 $d_+ = d_- = \frac{1}{\|w\|}$ ，亦即其“margin” = $\frac{2}{\|w\|}$ ，所以我們尋找區分平面的最大“margin”，就如同找最小的 $\|w\|^2$ ，但需符合(4)式的條件。所以吾人改以 Lagrangian 最佳化問題的解法來處理這個問題，有二個理由可以來說明：第一，條件((4)式)可被 Lagrange multipliers 取代，更容易處理；第二，經過整理後的公式中訓練資料將以向量內積的形式來表示。Lagrangian function 如下所示：

$$L_p \equiv L(w, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^L \alpha_i y_i (\langle w, x_i \rangle + b) + \sum_{i=1}^L \alpha \dots \dots \dots (5)$$

在(5)中 $\alpha_i, i=1,2,\dots,l$ 為 Lagrange multipliers

經過事先的觀察，吾人知僅有少數的訓練樣本會對區分平面具有影響，也就是這些樣本在(5)式中 $\alpha_i \neq 0$ ，而這樣的樣本就稱為支撐向量(Support Vectors)。

倘若訓練資料為非線性樣品，又該如何處理呢或是如何針對一般的狀況來找區分平面最大的“margin”？從上述的例子中，吾人觀察到一個重點，即樣本資料是以內積的型式($X_i \cdot X_j$)在做訓練。現假設訓練資料以一種映射方式對映到其它的歐幾里得空間(Euclidean space)H，我們以 Φ 來表示，即 $\Phi: \mathcal{R}^d \rightarrow H$ 。當然資料訓練的演算法在 H 上仍依靠內積方式處理，也就是 $\Phi(X_i) \cdot \Phi(X_j)$ ，我們將在訓練演算法時使用 K，但吾人並不需要知道 Φ 是什麼。 [2]

在非線性支向機(Support Vector Machine, SVM)處理樣式辨識(pattern recognition)的問題上，其核心函式的研究有下述三類：

$K(x,y)=(x \cdot y+1)^p$多項式(P 為次數)

$K(x,y)=e^{-\|x-y\|^2 / 2\sigma}$ 放射型

$K(x,y)=\tanh(kx \cdot y-\delta)$S 型

最後，分類器以下面式子來表示：

$$f(x) = \sum_{i=1}^L \alpha_i y_i K(x, x_i) + b \dots \dots \dots (6)$$

在(6)中 L 是支撐向量的個數，樣本的類別就過函式 f 來決定。此為二類別的分類器。多類別的分類器(Multi-classifiers),仍以二類別的分類器為基礎，以兩兩(one-against-one)的方法處理，以三個類別的分類器為例：需做 $\frac{3 \times (3-1)}{2} = 3$ 個二類別的分類器，再以投票的方式決定採用那一個二類別分類器，再送出結果。 [15]

3.文獻探討

3.1 國外部分

一、Kimoto、Asakawa、Yoda、Takeoka (1990) [5]

利用模組化的類神經網路，預測東京股市(TOPIX)的最佳買賣時機，其使用的指標包括，向量曲線、利率、營業額、國外交易比、美國紐約道瓊平均等等。資料的擷取為每週平均，並以 Moving Simulation 的方式，來訓練網路。模組化乃指多個神經網路分別學習單一變數，在經處理整合，以使類神經網路產生可被解釋的輸出，了解各輸入變數間的關係。實驗結果顯示，與買入持有相較，有更佳的獲利。

3.2 國內部分

一、陳榮昌、陳同孝、劉育辰（2003）[17]

篩選基本面分析中的總體經濟指標，投入 SVM 中作未來隔月股價漲跌預測。實驗結果顯示，對於電子資訊類股的漲跌預測，有不錯的預估準確率，尤其在針對傳統電子類股方面更有 6 成以上的平均準確率。

二、林萍珍、陳稼興（1998）[18]

透過人工智慧之遺傳演算法及台灣股市做實證研究，並以公司基本面分析及使用者為導向發展出一套投資組合選擇模型。研究顯示，投資組合年報酬率高於國內較佳共同基金績效相當，甚至超越部分基金報酬近 30%，而最差投資組合年報酬率也高於國內較差的共同基金績效及台灣加權指數年報酬率。

三、王春笙、李瑞庭（1996）[19]

運用倒傳遞及複迴歸，預測 6 日、12 日、18 日後之漲跌，輸入變數為十個技術指標。實驗結果顯示，倒傳遞優於複迴歸，漲跌預測方面，為 18 日較為準確，大約 5 至 6 成。

4. 研究方法

4.1 資料來源

總體經濟指標方面是由行政院主計處第三局的整體統計資料庫、經濟建設委員會及台灣證券交易所取得，蒐集從 1983 年 1 月起的各項所需資料，包括證券代碼、公司名稱、每月平均收盤價、各季及各月的總體經濟指標，總計 23 種研究所需資料。技術指標方面則是由台灣經濟新報資料庫中取得，其中包含台灣電子類上市公司之每月基本資料。

4.2 研究對象與範圍

研究對象為電子類第一類股，股票代碼為 2301~2345 之間之股票，第一階段選股乃針對國內基金組合中最常被投資及被認為最易帶動此類股波動的股票(此即所謂的龍頭股)中，挑選出十支個股，第二階段選股乃由剩下的個股中隨機挑選出五支股票，並針對兩階段所挑選出的個股(共十五支股票)，作隔月股價漲跌之預測。代碼為 2301~2345 之間之股票，所能收集到的資料較為完整；由於台灣股票的市值較小，容易受到人為的操作，而電子股的平均股本相對於其他類股大，可避免上述的情況。

研究範圍方面，基本分析[11]所使用到的資料介於 1988 年 1 月至 2002 年 12 月，其中 1988 年 1 月至 1999 年 12 月為訓練資料；技術分析[12]方面，

介於 1991 年 12 月至 2002 年 12 月，其中 1991 年 12 月至 1999 年 12 月為訓練資料，兩種分析法的訓練範圍，依各股上市時間不同及可用資料而有所調整，兩分析法的測試資料皆為 2000 年 1 月至 2002 年 12 月。

在輸入變數的選擇上，基本分析乃將所收集到的總體經濟變數，以主成分分析篩選出對股價漲跌最有影響力的總體經濟變數組合[13]，其中包括海關出口值（電子類）、海關出口值（資訊類）、受雇員工每人每月平均工時、躉售物價指數、證券交易統計表—C 股票、貨幣總計數（M1B）、產出分類指數、領先指標綜合指數、同時指標綜合指數等九種；而技術分析方面則以陳榮昌、陳同孝、徐士翔[16]篩選後使用的技術指標，作為此建構模型的輸入變數，其中包括 MA6、MA12、MA24、RSI6、RSI12、K9、D9、W%R12、PSY10、BIAS6、DMI10、DIF、MACD 等 13 種。在附錄 A，針對此兩種分析法所用到的指標作簡單的介紹。

4.3 研究限制

本研究主要是運用基本分析與技術分析的指標，透過 SVM 架構，對未來股價漲跌進行預測，然而影響股價的因素很多，僅利用基本分析與技術分析可能無法完全反應出影響股價變動的全部因素。尤其消息面對台灣股市影響甚鉅，且兩岸政治關係微妙，要進行準確的因素歸納實屬不易。另一方面，影響電子類股漲跌的因素也是每年翻新。不同經濟體系影響股價的因素，也會隨著產業的轉型或其他原因而有所改變。在訓練資料與預測資料方面，因為資料的收集無法達到最理想的完整情況，尤其是公司內部密而不宣的機密，這些都是本研究無法克服的部分。

5 實證分析架構

5.1 基本分析—總體經濟指標

首先，利用[10]的方法在許多總體經濟變數的組合中，運用主成分分析篩選出與未來股價較有關聯之總體經濟指標，並將篩選出來的指標(海關出口值（電子類）、海關出口值（資訊類）、受雇員工每人每月平均工時、躉售物價指數、證券交易統計表—C 股票、貨幣總計數（M1B）、產出分類指數、領先指標綜合指數、同時指標綜合指數)作為 SVM 預測模型之輸入，並運用 SVM 之非線性迴歸功能，訓練出一預測模型，並將要預測的資料集合，投入此預測模型中，以產生預測結果。

5.2 技術分析—基礎指標

首先，將所收集的個股基本資料，利用試算表軟體（Microsoft EXCEL）的函數功能，計算出個股每月之技術指標。然而，若投資者僅使用一種技術分析法，並不能非常準確的預測股價[3]。因此，將此 13 種技術指標（MA6、MA12、MA24、RSI6、RSI12、K9、D9、W%R12、PSY10、BIAS6、DMI10、DIF、MACD），共同作為 SVM 預測模型之輸入，並運用 SVM 之非線性迴歸功能，訓練出一預測模型，並將要預測的資料集合，投入此預測模型中，以產生預測結果。

在準確率評估上，此兩種分析法都是將實際股價隔月正負值（漲、跌）與預測資料作一比較，當實際情況與預測情況相同時，表示預測正確；當實際情況與預測情況相異時，表示預測錯誤。

5.3. SVM 架構

SVM 是一種 Machine learning 最常被使用在分類問題，尤其是用在預測上，不但能夠處理一般迴歸預測的問題，在處理當 Y 是類別資料的時候，通常也都能夠有很好的效果[11]。其資料形式，如表 5.1。

表 5.1 資料形式

X_1	Y_1
:	:
:	:
X_n	Y_n
X_{n+1}	Y_{n+1}
:	:
X_{n+m}	Y_{n+m}

我們把一組資料分成兩部份(如表 5.1)， $X_1 \dots X_n$ 部份當做訓練資料(Training data)通常解釋變數(X)視為輸入(input)，而經過 SVM 的分類之後會得到反應變數(Y)的分類視為輸出(output)。為了檢測這個 Machine 好不好，都是在反應變數已知的情況下，求得訓練誤差(Training error)和測試誤差(Testing error)，至少訓練誤差不能太大且兩個誤差不能相差太遠，這個 Machine 才值得採用。我們利用這個 Machine 來預測我們的新進資料，要測試的新進資料反應變數是不知道或是不清楚的。

舉例來說：我們想預測學生修習統計學的成果，這一屆預測某學生修習統計學的成果(1：及格，0：不及格)，我們可以由其他的一些行為模式(解釋變數)來預測他能不能及格，例如相關科目的成績，看書的時間等等，當我們預測他能及格時但實際上卻不及格，或是預測不及格時但實際上卻及格就稱之為誤差，而為測試這個 **Machine**，我們就把某學生這個班級的其他學生一部份當做訓練資料一部份當做測試資料來得到訓練誤差和測試誤差，假如這個 **Machine** 是可接受的，那麼我們就可以在統計學學習成果未知的情況之下，使用這個 **Machine** 來預測下一屆學生的情況。

本研究嘗試利用 **SVM** [14]之特性與優點，來預測未來股市的走向，**SVM** 具有分類與迴歸的功能，其特性類似於類神經網路，但 **SVM** 的訓練方式較類神經網路要來得快速及簡單，且訓練時不須給予額外的知識幫助。針對個電子類股進行基本分析與技術分析加以篩選，再利用 **SVM** 的迴歸功能，分別建構出各自的模型，並比較兩種模式的優劣，以提供給投資者在進行投資決策時，能夠作出較正確的判斷，降低決策風險，並增加投資的獲利率。

本研究之流程如圖 5.1 所示:

6.實證結果

6.1 準確率與資料多寡

整體而言，本研究所挑選出之個股在進行隔週股價漲跌的預測上，都有不錯的整體正確率，在基本分析方面分別如表 6.1：

表 6.1 基本分析個股正確率

個股名稱	代碼	準確率	個股名稱	代碼	準確率
光寶	2301	62.5%	矽品	2325	66.7%
聯電	2303	65.2%	台積電	2330	58.3%
日月光	2311	60.8%	精英	2331	50.0%
金寶	2312	29.2%	致伸	2336	54.2%
台揚	2314	37.5%	旺宏	2337	45.5%
鴻海	2317	69.5%	英群	2341	37.5%
佳錄	2318	54.2%	華邦	2344	63.6%
中環	2323	58.3%			

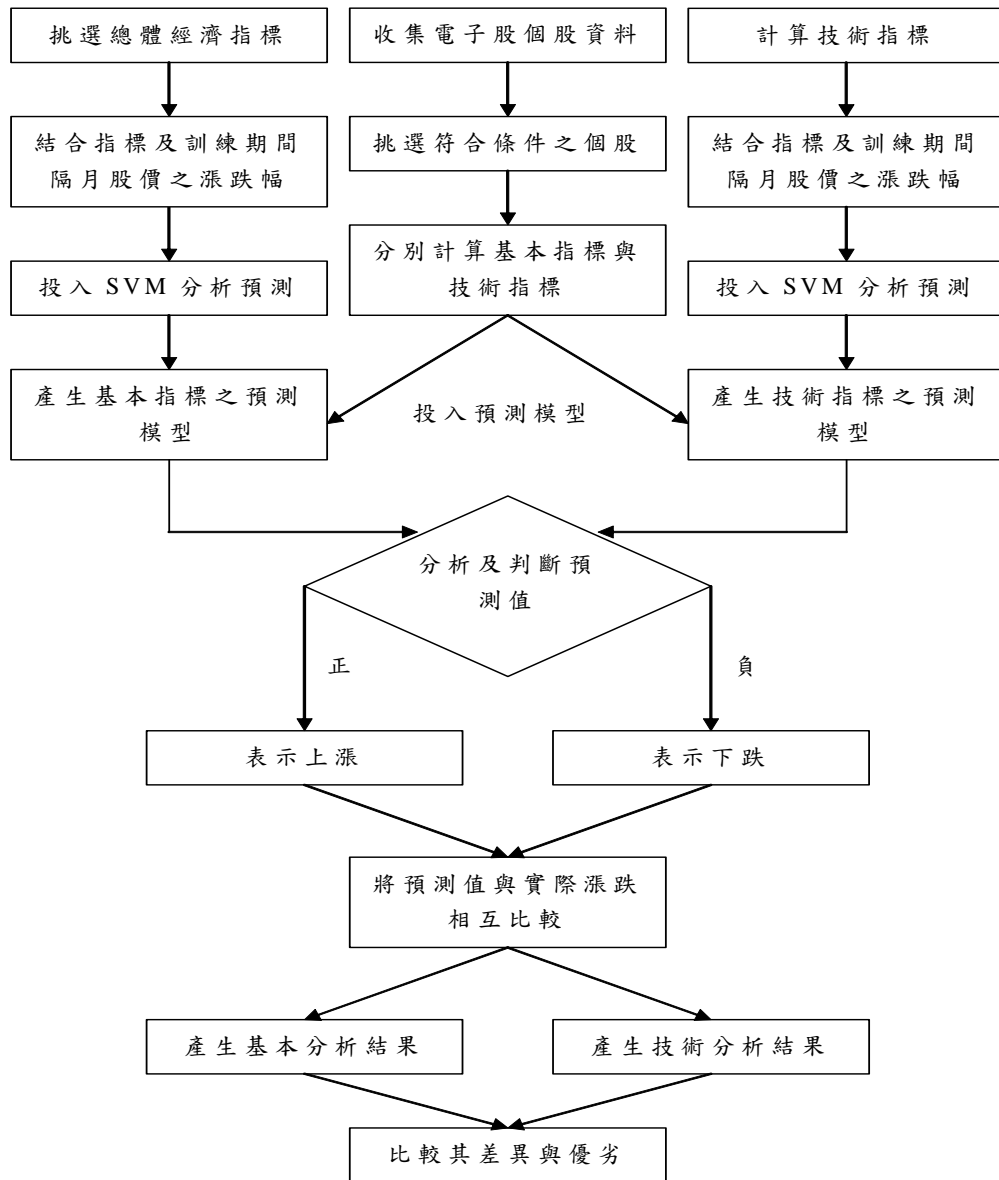


圖 5.1 研究流程

其中有六支個股之準確率高達 60% 以上，有著相當令人滿意的準確率，而在準確率與訓練資料多寡間可發現資料多寡因實際可得之資料而有所不同，基本分析對於訓練資料的多寡與準確率上，並無明顯的關係，越早期的資料，並不一定能對近期的股市漲跌造成影響，相對的也有可能對準確率造成負面的影響，這也可能是造成某些個股準確率低落的原因。而在技術分析方面，對相同的個股進行預測，個股的準確率分別如表 6.2：

表 6.2 技術分析個股正確率

個股名稱	代碼	準確率	個股名稱	代碼	準確率
光寶	2301	61.1%	矽品	2325	52.7%

聯電	2303	69.4%	台積電	2330	55.5%
日月光	2311	55.5%	精英	2331	61.1%
金寶	2312	52.7%	致伸	2336	50.0%
台揚	2314	47.2%	旺宏	2337	61.1%
鴻海	2317	47.2%	英群	2341	52.7%
佳錄	2318	55.5%	華邦	2344	52.7%
中環	2323	52.7%			

其中有四支個股之準確率高於 60%，在準確率與訓練資料多寡間可發現(資料多寡因實際可得之資料而有所不同)，技術分析面對越多的訓練資料，預測的結果則有較為準確的趨勢，這也正好符合了技術分析的基本特性，即相信股市是一種循環，如果能夠掌握充足且詳細的歷史資料，對於未來股價漲跌的預測準確率就越能夠期待。

6.2 準確率與股本大小

實驗過程中，利用兩種分析法分別建構之模型，預測所挑選出的個股，其中基本分析所預測的十五支個股中有六支個股的準確率高於 60%，而利用技術分析預測相同的十五支個股中，只有四支個股高於 60%，但準確率介於 50%與 60%的個股卻高達九支，遠高於基本分析的四支。再進一步觀察股本對於準確率的影響可發現，利用基本分析預測未來股價漲跌，而準確率低於 50%的五支個股中，有四支個股的股本小於 200 億(見表 5.1)。可見基本分析模型對於股本小的個股，預測的準確度較低。此可能因為股本小的個股，比較容易受到外來因素的影響，股價操控較易，進而造成股價無法忠實反應基本面指標所欲傳達的訊息。另一個技術分析模型，相對的較看不出準確率與股本間有明顯的關係；此可能因為技術指標是一種較為敏感的股市分析指標，股市中的變化，經由可由分析的法則，整理出一些比較的原則，故技術分析模型的實驗結果，比較不容易受到外來因素影響的原因。實驗所得之詳細結果，請參照表 6.3。

表 6.3 準確率與股本的關係

	基本分析	技術分析
整體準確率 $\geq 60\%$ 之個股數目	6	4
60% $>$ 整體準確率 $>50\%$ 之個股數目	4	9
整體準確率 $\leq 50\%$ 之個股數目	5	2
整體準確率 $\leq 50\%$ 之個股數目	4	1

(股本<200 億)		
整體準確率> 50%之個股數目 (股本<200 億)	4	6
整體準確率> 50%之個股數目 (股本>200 億)	6	7
整體準確率≤ 50%之個股數目 (股本>200 億)	1	1

6.3 個股與分析法

本研究利用基本分析與技術分析的理論，藉由 SVM 所訓練的模型預測未來股價的漲跌。經由實驗結果顯示，利用基本分析與技術分析兩種模型預測第一階段所挑選出的十支個股(光寶、聯電、日月光、鴻海、佳錄、中環、台積電、精英、旺宏、華邦)中，準確率高於 55%的個股皆有七支(見表 5.4)，而基本分析的平均準確率為 59.7%，技術分析的平均準確率為 57.2%，此兩種分析模型對於電子產業中較具代表性的個股皆有不錯的準確率。進一步觀察可發現利用基本分析，準確率高於 55%的七支個股中，有六支個股的股本大於 200 億，即股本較大的個股，其股價與基本指標較有正向的連動；而利用技術指標則較看不出股本與股價間的關聯(見表 5.3)。接著觀察第二階段所挑選出的五支個股(金寶、台揚、矽品、致伸、英群)，利用基本分析方面，此五支個股的股本皆小於 200 億(見表 5.5)，其中準確率高於 50%的只有一支個股(見表 5.3)，這更證明了股價與基本指標有一定的正向關聯；而技術指標方面，準確率與股本間的關係則顯得較為薄弱，然而股本越小之個股越易受到外來因素影響，股價波動較為劇烈，因而造成準確率較無法達到令人滿意的水準，但基本上此五支個股的準確率大多有 50%以上水準(見表 5.5)，且優於基本分析之準確率。

此兩種分析法，投資者並不需具備深厚的投資經驗或財經背景，即可依照本研究的實驗步驟對股票進行未來股價漲跌之預測，並透過個股的特性，選擇較適合的分析法。經由實驗的結果發現，只需透過簡單的指標即可創造出不錯的準確率，對於投資者在進行買賣決策時，較能夠作出正確的判斷。實驗之詳細結果，請參照表 6.4、6.5。

表 6.4 第一階段選股

公司名稱	模型指標	準確率	股本(億)	訓練期間	預測期間
光寶	基本分析	62.5%	191	1995/11~1999/12	2000/01~2002/12

(2301)	技術分析	61.1%		1993/11~1999/12	
聯電	基本分析	65.2%	1,540	1988/01~1999/12	2000/01~2002/12
(2303)	技術分析	69.4%		1993/11~1999/12	
日月光	基本分析	60.8%	325	1989/07~1999/12	2000/01~2002/12
(2311)	技術分析	55.5%		1993/11~1999/12	
鴻海	基本分析	69.5%	206	1991/06~1999/12	2000/01~2002/12
(2317)	技術分析	47.2%		1993/11~1999/12	
佳錄	基本分析	54.2%	16	1991/06~1999/12	2000/01~2002/12
(2318)	技術分析	55.5%		1993/11~1999/12	
中環	基本分析	58.3%	260	1992/02~1999/12	2000/01~2002/12
(2323)	技術分析	52.7%		1994/12~1999/12	
台積電	基本分析	58.3%	1,862	1994/09~1999/12	2000/01~2002/12
(2330)	技術分析	55.5%		1997/07~1999/12	
精英	基本分析	50.0%	43	1994/09~1999/12	2000/01~2002/12
(2331)	技術分析	61.1%		1997/07~1999/12	
旺宏	基本分析	45.5%	373	1995/03~1999/12	2000/01~2002/12
(2337)	技術分析	61.1%		1998/01~1999/12	
英群	基本分析	37.5%	44	1995/08~1999/12	2000/01~2002/12
(2341)	技術分析	52.7%		1998/06~1999/12	

表 6.5 第二階段選股

公司名稱	模型指標	準確率	股本(億)	訓練期間	預測期間
金寶	基本分析	29.2%	108	1989/11~1999/12	2000/01~2002/12
(2312)	技術分析	52.7%		1993/11~1999/12	
台揚	基本分析	37.5%	43	1990/08~1999/12	2000/01~2002/12
(2314)	技術分析	47.2%		1993/11~1999/12	
矽品	基本分析	66.7%	189	1993/04~1999/12	2000/01~2002/12
(2325)	技術分析	52.7%		1996/02~1999/12	
致伸	基本分析	54.2%	41	1995/01~1999/12	2000/01~2002/12
(2336)	技術分析	50.0%		1997/11~1999/12	

英群 (2341)	基本分析	37.5%	44	1995/08~1999/12	2000/01~2002/12
	技術分析	52.7%		1998/06~1999/12	

7.結論

本研究主要以基本與技術指標為基礎，運用 SVM 的迴歸功能，來預測股市未來一週的漲跌。在預測工具方面，SVM 與現今被廣泛運用於股市預測的類神經網路相比，其訓練步驟比較簡單(並不用建構訓練網路)，只需將訓練資料投入到 SVM 即可建構模型，對於一般投資者來說簡單易學、容易上手。藉由實驗的解果證明，SVM 在預測未來股價漲跌方面有不少的準確率，對於投資者在進行股市交易時，有助於投資決策上的正確判斷，進而在股市中獲得期望的報酬。

而在分析方法方面，基本分析與技術分析對於某些特定條件的個股皆有其一定的準確率，然而，對於某些個股而言其成效仍然不夠顯著，這或許是台灣股市存在著資訊不對稱，容易受到非經濟因素的影響，單靠基本分析中的總體經濟指標與技術分析中的基礎指標，並不足以充分反映台灣股市的所有訊息。因此，可能尚需配合其它可能影響股市波動的資訊，以及觀察股價波動與時間區段間的關係，才能有效預測台灣股市未來之走趨。但以本研究所得到的結果可以證明，在理性環境中，台灣股市是可以預測的，運用基本指標於股本較大的個股，較易獲得令人滿意的準確率，而運用技術指標於未來股價漲跌上，較不易受到外來因素影響而造成準確率大幅波動的情況，後續的研究可朝結合其他相關資訊或預測其它類股方面，繼續探討。

參考文獻及著作

- [1] Cootner, P. ed. The Random Character of Stock Market Prices. MIT, Cambridge, MA, 1964.
- [2] Christopher J.C.B wrges, "A Tutorial on Support Vector Machines for Pottern Recognition", Bell Laforatories, Lucent Techuologies.
- [3] Granville, Joseph E. Granville's New Strategy of Daily Stock Market Timing for Maximum Profit. Englewood Cliffs: Prentice-hall, Inc, 1976, pp.10.
- [4] John Maynard Keynes, "The General Theory of Employment, Interest, and Money", New York: Harcourt. Brace and Company, 1936.
- [5] Kimoto, Takashi. "Stock Market Prediction System with Modular Networks." IEEE International Joint Conference on Neural Networks, Vol.1, 1990.

- [6] Nello Cristianini and John Shawe-Taylor, "An introduction to Support Vector Machines and other Kernel-based learning methods", CAMBRIDGE UNIVERSITY PRESS 2000.
- [7] Olivier Chapelle, Patrick Haffner, Vladimir N. Vapnik, "Support Vector Machines for Histogram-Based Image Classification," IEEE Transaction on Neural Networks, Vol. 10, No. 5, Sep.1999, pp. 1055-1064.
- [8] Rebert, A. Levy. "Conceptual Foundation of Technical Analysis." Financial Analysis Journal, July-August, 1966.
- [9] Reilly, Frank. K. Investments. New York, CBS College Publishing, 1982.
- [10] Rhea, Robert: Dow Theory, Barrons, New York, 1932.
- [11] Scholkopf, B. and Smola, A. (2001). Learning with Kernels-support Vector Machines, Regularization, Optimization and Beyond.
- [12] Steve Gunn, "ISIS Technical Report-Support Vector Machines for Classification and Regression", May 14th, 1998
- [13] Stefan Rüping, mySVM, Version 1.4.7, March 8th, 2001
- [14] V. Vapnik, The Nature of Statistical Learning Theory. Springer-Verlag, New York. 1995.
- [15] W.N Venaffes, D.M. Smith and R Development Core team, "An Introduction to R", Version 1.4, 2001
- [16] Zhang Lei, Fuzong Lin, Bo Zhang, "Support Vector Machine Learning for Image Retrieval," IEEE Image Processing, Vol. 2, Oct. 2001, pp. 721-724.
- [17] 陳榮昌、陳同孝、劉育辰著，以總體經濟變數預測台灣電子股漲跌之實際研究—運用 SVM 技術，國立台中技術學院技術報告，民國 92 年。
- [18] 林萍珍、陳稼興著，遺傳演算法在使用者導向的投資組合選擇之應用，中央大學資訊管理研究所碩士論文，87 年 6 月。
- [19] 王春笙著，以技術指標預測台灣股市股價漲跌之實證研究-以類神經網路與複回歸模式建構，台大資管所碩士論文，民國 85 年。
- [20] 陳榮昌、陳同孝、徐士翔著，植基於 SVM 的電子股價漲跌預測，國立台中技術學院技術報告，民國 92 年。
- [21] 謝東宏，運用支向機於華語單音節混淆音組辨認之初步研究，長庚大學電

機工程研究所碩士論文，91 年 8 月。

附錄 A

A 指標的定義

A.1 基本分析—總體經濟變數[4]

一、海關出口值

此項統計按國家海關的出口總值及商品類別分別給予列計，本研究分別以全部各類的海關出口值（加總）、電子類及資訊類的出口值來作為研究中的輸入變數。

二、受雇員工每人每月平均工時

本研究所使用的指數為製造業的每位受僱員工每個月所工作的平均時數。

三、躉售物價指數 WPI (Wholesale Price Index)

又稱批發物價指數，就是衡量物品批發價格的物價指數。其意義為廠商獲利的好壞。其中的物品通常不包括各項勞務。

四、證券交易統計表—C 股票

以西元 1966 年為基期，計算股市證券交易綜合指數。

五、貨幣總計數 M1B

銀行活期存款、活期儲蓄存款、支票存款再加上流通在社會上的現金，其對股市資金增減有最直接的影響。

六、產出分類指數

衡量工業部門產品生產量在某時間與基期間之相對指標。

七、領先指標綜合指數

為經建會所編製的景氣動向指標之一，是根據能夠提前反映景氣變動情況且敏感度較高的 7 項指標編製而成，在季節調整、標準化因子調整及加權平均等統計處理後，合併成一個綜合指數時間數列，可用來預測未來景氣變動的方向與幅度。

八、同時指標綜合指數

目前經建會所編製的同時指標綜合指數，係由能夠即時反映經濟景氣的 6 項指標所組成，經過季節調整、標準化因子調整及加權平均等統計處理後，合併成一個綜合指數時間數列，可反映出當時的景氣狀況，用以衡量景氣之變動。

A.2 技術分析—基礎指標[10]

一、移動平均線 MA

MA 是一種最基本，而又最有效的趨勢判別方法。其主要特性是將過去某段時間內的收盤價相加，計算其平均數。例如週平均線，即是將一段時間內之每週平均算出，連接起來，便行成一條移動平均線。

計算公式:

$$MA_n = (W_1 + W_2 + W_3 + \dots + W_n) / n$$

二、相對強弱指標 RSI

RSI 目前已被市場普遍使用，為最基本的技術指標之一，在實質上，RSI 是一種動量指標，其主要意思是計算某一段時間內買賣雙方力量，以作為超買、超賣的參考依據與 K 線圖及其他技術指標配合使用，以免過早賣出及買進，造成少賺多賠的損失。

計算公式:

$$UP = \text{過去 } n \text{ 日內上漲點數總和} / n$$

$$DN = \text{過去 } n \text{ 日內下跌點數收總和} / n$$

$$RS = UP / DN$$

$$n \text{ 日 RSI} = 100 - 100 / (1 + RS)$$

三、隨機指標 KD

股票市場常使用的一套技術分析工具。其適用範圍以中短期投資的技術分析為最佳。KD 指標的理論認為：當股市處於多頭市場時，收盤價往往接近當日最高價；反之在空頭市場時，收盤價比較接近當日最低價，該指數的目的即在反映出近期收盤價在該段日子中價格區間的相對位置。

計算公式:

它是由 %K(快速平均值)、%D(慢速平均值)兩條線所組成，假設從 n 天週期計算出隨機指標時，首先須找出最近 n 天當中曾經出現過的最高價、最低價與第 n 天的收盤價，然後利用這三個數字來計算第 n 天的未成熟隨機值(RSV)。

$$RSV = (\text{第 } n \text{ 天收盤價} - \text{最近 } n \text{ 天內最低價}) / (\text{最近 } n \text{ 天內最高價} - \text{最近 } n \text{ 天內最低價}) \times 100$$

計算出 RSV 之後，再來計算 K 值與 D 值。

$$\text{當日 K 值}(\%K) = 2(\text{前一日 K 值}) + 1(RSV) / 3$$

$$\text{當日 D 值}(\%D) = 2(\text{前一日 D 值}) + 1(\text{當日 K 值}) / 3$$

若無前一日的 K 值與 D 值，可以分別用 50 來代入計算，經過長期的平滑的結果，起算基期雖然不同，但會趨於一致，差異很小。

四、威廉指標 W%R

W%R 是運用股市的擺動點來量度其超買超賣現象，可以量度循環期內的高點或低點，提出有效率的投資訊號，其為一有效的短期指標。

計算公式:

$$W\%R = 100 - (H_n - C) / (H_n - L_n) \times 100$$

H_n = n 日內最高價， L_n = n 日內最低價，C = 當日收盤價

五、心理線 PSY

PSY 是研究某段期間內投資人趨向於買方或賣方的心理，來做為買賣股票的依據，事實上，它就是一種人氣指標。

計算公式:

$$n \text{ 日 PSY} = (n \text{ 日內的上漲天數} / n) \times 100\%$$

六、乖離率 BIAS

BIAS 乃是用以測定當日股價與移動平均線偏離程度的一項技術指標：當股價偏離平均線時，它最終應該都會回歸平均線的平衡狀態。

計算公式:

$$BIAS \ n = \text{當日股價} - \text{最近 } n \text{ 日平均股價} / \text{最近 } n \text{ 日平均股價}$$

七、趨向指標 DMI

DMI 利用計量分析方法，以較客觀性的態度，研判股價漲跌的趨勢。在研判時，未摻雜個人主觀性的判斷，且能考慮股價每日的最高價、最低價及收盤價三者間的波動情形，可對股價的波動情形做完整性分析。

計算公式:

1. 先求出趨向變動值(DM)--趨向變動值為本日股價變動幅度大於昨日股價變動幅度的『最大值』。

$$『+DM』 = \text{本日最高價} - \text{昨日最低價}$$

$$『-DM』 = \text{本日最低價} - \text{昨日最低價}$$

DM 能表達出股價波動增減的幅度。『+DM』及『-DM』計算出來後，再分別求出其 N 日移動平均值(一般以 10 日、12 日、14 日為計算日期)。

2. 找出真實的波幅(真實的波動價位值，簡稱 TR)--TR 為本日行情與昨日行情比較後的最大變動值。該變動值需比較下列三種差價的『絕對值』後，取

其中最大者為本日之 TR。

本日最高價 - 本日最低價

本日最高價 - 昨日收盤價

本日最低價 - 昨日收盤價

TR 求出後，再計算其 N 日之移動平均值。

3. 求出方向線(DI)--為探測股價上漲或下跌方向的指標，以 +DI 表示上升方向指標，為最近 N 日內實際上漲的動量百分比；以 -DI 表示下跌方向指標，為最近 N 日內實際下跌的動量百分比。

+DI = + DI N 日平均 / TR N 日平均

-DI = - DI N 日平均 / TR N 日平均

4. 最後求出平均方向的移動平均值(ADX)--

方向平均值(DX) = | (+DI) - (-DI) | (絕對值) / (+DI) + (-DI)

再計算其 N 日移動平均值 ADX。

八、MACD

MACD 根據移動平均線(MA)較易掌握趨勢變動的優點所延伸出來的一種分析方式。其基本原理是運用兩條不同速度的指數平滑移動平均線來計算兩者之間的差離狀態(DIF)，然後再對 DIF 進行平滑移動平均即為 MACD 線。簡而言之，MACD 就是對長期與短期的移動平均線收斂或發散的徵兆，加以雙重平滑處理，用以研判買賣股票的時機與信號。

計算公式：

先行計算出快速線(n 日 EMA)與慢速線(m 日 EMA, n<m)。以此兩個數值，再計算兩者間的「差離值」(DIF),再計算 DIF 之 x 日 EMA,此即 MACD。

$nEMA = (\text{前一日 } nEMA * (n-1) + \text{今日收盤價} * 2) / (n+1)$

$mEMA = (\text{前一日 } mEMA * (m-1) + \text{今日收盤價} * 2) / (m+1)$

$DIF = nEMA - mEMA$

$xMACD = (\text{前一日 } xMACD * (x-1) + DIF * 2) / (x+1)$