

應用支援向量迴歸於國際旅遊需求之預測

Application of support vector regression in forecasting international tourism demand

陳寬裕 Kuan-yu Chen*

何嘉惠 Chia-Hui Ho

蕭宏誠 Horng-Cherng Shiau

(2004/10/20 收稿, 2005/1/4 接受刊出)

摘要

本研究將運用支援向量迴歸技術 (Support Vector Regression, SVR) 建構國際旅遊需求之預測模型。為了建構穩健、可信賴的預測模型, 需慎選建構模型時所必需的參數 γ 、C 和 ϵ 。因此先使用格子點法 (grid method) 配合 5-fold 交叉驗證技術於訓練資料集上, 以求得各種不同參數組合下的交叉驗證 MAPE (mean absolute percentage error), 再藉由本研究所提出的, 以支援向量的數量為主、評估交叉驗證 MAPE 為輔的參數選取機制, 篩選出最佳的參數組合, 最後再利用此參數組合建構實際的支援向量迴歸預測模型。此外研究中使用了敏感度分析技術 (sensitivity analysis), 以證明不當的選取參數將使模型容易陷於過度擬合 (over-fitting) 或不足擬合 (under-fitting) 的危機中。

關鍵字：支援向量機、支援向量迴歸、類神經網路、國際旅遊需求

ABSTRACT

This study applies support vector regression technique to construct the prediction model of international tourism demand. To build stable and reliable prediction model, parameters (γ , C and ϵ) are selected carefully. Since this study proposes that

* 長榮大學經營管理研究所博士班；聯絡地址：台南縣新市鄉中華路 49 號遠東技術學院企業管理系 (Email: t530903@ms7.hinet.net)

* 遠東技術學院企業管理系
遠東技術學院企業管理系

the number of support vector, as the main technique, associated with 5-fold cross-validation technique together to build parameters selection mechanism and finds out the optimal parameter set, and uses the parameter set to construct real support vector regression model. Moreover, this study implements sensitivity analysis technique, the analysis demonstrates that incorrectly selected parameters will lead the model's results in the risk of over-fitting, or under-fitting.

Keywords: Support Vector Machine、Support Vector Regression、Artificial Neural Network、International Tourism Demand

一、緒論

根據世界觀光組織 2000 年版(World Tourism Organization, 簡稱 WTO)的分析報告指出,在全球各國的外匯收入中約有超過百分之八來自觀光收益,高居各類收入的第一名,由此可知觀光產業已漸成為許多國家賺取外匯的首要來源。世界觀光旅遊委員會(World Travel and Tourism Council, 簡稱 WTTC)更進一步預估,至 2005 年,全球觀光人數將成長至十億人次,全球觀光收益亦將達到一兆美元,且至 2010 年,全球觀光就業人口數更可達到三億三千八百萬人次。因此觀光產業之於全球,乃至於單一國家的經濟發展,在可見的未來均將扮演著重要的角色。

基於此,旅遊需求儼然已成為各個國家社會與經濟發展的重要觸媒,然而觀光產業具備易消逝性(Perishability)的經濟特徵,也就是說,其產品具有無法儲存(如旅館空房、航空班機座位等),產出後即消逝之特性,因此其供給無法隨著市場的需求或變化而任意加以調整,而必須在供需方面加以配合管理,由此可知旅遊業之需求預測將顯的格外重要。所以若能對旅遊需求預測得當,則將可預先得知未來旅遊需求的相關統計資料,這些預測資料將會對計畫旅遊者或旅遊政策制訂者,於其旅遊決策的形成過程中帶來重要的影響。

過去三十幾年中,對於旅遊需求預測的研究方法中,以計量經濟模型為最多(Uysal and Crompton,1984、Sheldon and Var,1985、Martin and Witt,1987、Smeral,Witt and Witt,1992、Manuel and Croes,2000),此外運用時間序列方法進行旅遊需求預測的研究也不少(Sheldon and Var,1985、Chan,1993、Pilar and Molar,1995、Pattie and Snyder,1996、Kulendran and King,1997、Chu,1998),至於其他的量化方法被使用的頻率則較少。利用計量經濟模型進行旅遊需求預測的好處,在於可掌握旅遊需求的重要影響因素,但相對的所花費的時間與成本也較高,而且影響因素之資料亦較難蒐集。而時間數列方法雖無法瞭解變數間的關係,但其所需的實證資料較少,因此較適合旅遊需求資料闕如的研究課題。

近幾年,類神經網路(artificial neural networks)已經被成功的應用在預測問題的建模上(Cheng et al,1996、Sharda and Patil,1992、Van and Robert,1997),它是一種泛用型的函數逼近器(universal function approximators),縱使對於資料的性質沒有任何事前的認知下,它也能映射任何的非線性函數(Haykin,1999),不同於傳統的統計模型,類神經網路是種資料驅動、無母數弱模型且它能让「資料能自己說話」。相較於一般參數模型,類神經網路較不容易掉入需確認模型種類問題中,且類神經網路也較具有對雜訊的容忍力,在資料具有不完整或錯誤資料的環境下,也較有能力去學習此種較複雜的系統。此外類神經網路還甚具彈性,在不

斷的使用新資料的情況下，它能透過不斷的重複學習而具有學習動態系統的能力，因此，和傳統的統計模型比較起來，在描述動態的時間序列方面，類神經網路更具威力(Kaastra and Milton, 1995、Chiang et al,1996)。然而類神經網路也具有一些弱點：比如它需要大量的控制參數、可能只得到區域最小解和可能冒著過度擬合的風險。所謂過度擬合之意義為：所建立的模型在訓練資料中雖能得到良好的擬合結果，但是在測試資料中卻只得到非常差的預測效果。過度擬合是個重要的問題，它常導致較差的預測結果，這是由於所建立的類神經網路模型對訓練資料具有較強的配適能力(fitness capacity)而導致它不只捕捉了訓練資料中的有用資訊而且也包含了無用的資訊。結果它將只是去記憶並擬合訓練資料中的資料點，但是對於測試資料的預測結果卻是相當的差。這個有關預測結果方面的問題早已被很多研究學者所關注。最典型的方法為使用交叉驗證技術來避免過度擬合現象，但這個方法需花費大量的計算時間。

最近，Vapnik(1995)等人提出一種叫做支援向量機(Support Vector Machine, SVM)的類神經網路演算法，它不像大部分傳統的類神經網路模型去遵循經驗風險最小化法則(empirical risk minimization principle)，它是一種以統計學習理論為基礎之VC維理論(Vapnik – Chervonenkis dimension)和遵循結構風險最小化原則(structural risk minimization principle)的新型學習機器。這法則試圖最小化預測誤差的上界(結構化風險)，而不是最小化訓練誤差(經驗風險)。由此，支援向量機能藉著經驗誤差及VC維之信賴區間的權衡(trade-off)而完成最佳化的網路結構。而其最後結果將比其他的類神經網路模型具有更好的預測結果。另一個支援向量機的優點為：訓練支援向量機的過程等價於解一個線性且具限制式的二次規劃問題(quadratic programming)，其所隱含的意義為支援向量機的解是唯一的、最佳化的且跳脫了所得到的解為區域最小解的可能性，而不像其他的網路訓練，在解非線性問題之最佳化過程中仍跳脫不了得到區域最小解的風險。支援向量機的發展最先被運用在模式識別領域(pattern recognition) (Schmidt,1996、Joachimes,1996)，然而藉著Vapnik(1997)的 ϵ -不敏感損失函數(ϵ -insensitive loss function)的導入，支援向量機已經被擴展到解決非線性迴歸估計的問題上了，這類新崛起的技術被稱為是支援向量迴歸(Support Vector Regression, SVR)，且這技術已被證明具有相當優秀的操作績效。

Law and Au(1999、2000)曾將類神經網路應用在旅遊觀光之需求預測中，其預測結果較傳統的簡算法(Naive)、移動平均法(Moving average)、指數平滑法(Exponential smoothing)和複迴歸法(Multiple regression)優秀。然而，類神經網路的本質上具有一些前述的弱點，因此本文乃欲使用嶄新的支援向量迴歸預測技術，建構國際旅遊之需求預測模型，並將其預測結果和 Law and Au 的類神經預測模型比較。同時由於尚無結構化、系統化的方法輔助選取建構支援向量迴歸模型時所需使用的參數，因此本文也提出了一種以支援向量的數量為主、評估交叉驗證 MAPE 為輔的參數選取機制，試圖為解決模型的選取問題上提供一個可行的方向。最後再透過敏感度分析技術，解析參數之設定對支援向量迴歸模型之預測能力的影響程度並探討支援向量迴歸模型的特性。

本研究將包含七個小節，第二節將對支援向量迴歸的相關理論作明確的介紹，第三節將說明資料來源、實驗設計方法，第四節將建構支援向量迴歸模型，第五節將討論、分析實證的結果，第六節將對支援向量迴歸的各參數進行敏感度分析，以參數對支援向量迴歸模型之預測能力的影響程度並探討支援向量迴歸模型的特性，最後，將勾勒出本文的結論與未來待努力的方向。

二、支援向量迴歸的理論基礎

當支援向量機被運用來估計迴歸方程時有三個比較獨特的特性，第一：支援向量機估計迴歸時，使用了定義在高維空間中之線性方程式的集合。第二：支援向量機藉著結構風險最小化的原則實現了迴歸估計的功能，且此結構風險將使用 Vapnik 的 一不敏感損失函數來加以估計。第三：支援向量機使用了由經驗誤差和結構風險最小化原則所導出的懲罰項 (regularization term) 所組合而成的風險函數 (Tay and Cao, 2001b)。

2.1 支援向量迴歸

給定一個資料點的集合 $G = \{(x_i, d_i)\}_i^n$ (x_i 是輸入向量、 d_i 是目標值、 n 是樣本資料量)，支援向量迴歸的基本概念是透過一個非線性映射 Φ ，將資料 x 映射到高維特徵空間，並在這個空間進行線性迴歸。即

$$f(x) = \mathbf{w}\Phi(x) + b \quad (1)$$

$$\Phi: R^n \rightarrow F, \mathbf{w} \in F$$

其中 \mathbf{w} 為權重向量 (weight vector)，它代表 $f(x)$ 在高維空間中的平坦程度 (flatness)，其歐幾里德長度 (Euclidean norm) $\|\mathbf{w}\|^2$ 可用來衡量 $f(x)$ 的平坦度，因此，也可認為 $\|\mathbf{w}\|^2$ 即代表著模型的複雜度。 b 是門限值， Φ 是所謂的高維特徵空間，它是輸入空間 x 的非線性映射。這樣，在高維特徵空間的線性迴歸便可對應於低維輸入空間的非線性迴歸，免去了在高維空間中 \mathbf{w} 和 $\Phi(x)$ 之內積計算。由於 Φ 是固定不變的，為使 $f(x)$ 能更平坦，因此必須尋找最小化的 $\|\mathbf{w}\|^2$ 。且在遵循結構風險最小化原則之下，藉著參數 C 的導入，為使支援向量迴歸的誤差 ($R_{reg}(C)$) 能權衡於模型複雜度與經驗誤差之間，因而定義式(2)的懲罰風險函數 (regularized risk function)，同時藉著最小化式(2)的過程而估計出 \mathbf{w} 和 b 的係數。

$$R_{reg}(C) = R_{emp} + \frac{1}{2} \|\mathbf{w}\|^2 = C \times \frac{1}{n} \sum_{i=1}^n L_e(d_i, y_i) + \frac{1}{2} \|\mathbf{w}\|^2, \quad (2)$$

$$L_e(d, y) = \begin{cases} |d - y| - e & |d - y| \geq e \\ 0 & otherwise \end{cases} \quad (3)$$

式 (2) 中 $R_{reg}(C)$ 即為迴歸誤差 (regression risk)，等號右邊第一項的 $C \times \frac{1}{n} \sum_{i=1}^n L_e(d_i, y_i)$ 就是經驗誤差 (風險)，它將藉由式(3)中的 一不敏感損失函數來加以估計。導入這個損失函數所將產生的好處是只要使用較少的資料點就能提供給式(1)的決策函數足夠的樣本。第二項 $\frac{1}{2} \|\mathbf{w}\|^2$ 就是所謂的懲罰項。 C 是懲罰常數 (regularized constant)，用於控制產生誤差時的懲罰程度，它將在經驗誤差和懲罰項 (代表迴歸的預測能力) 之間進行權衡。增加 C 值時，將導致經驗誤差相對於懲罰項的重要性也增加。其次，只有當擬合誤差大於 e 時才接受懲罰，使用 一不敏感損失函數的好處是提高了估計的穩健性，也就是說它具有一定的抗雜訊能力。因此 如同一個管狀空間 (Tube size)，其意義等價於在訓練資料上的預估精確度，如圖 1。未來實證時， C 和 e 都是研究者需自己選取的參數。

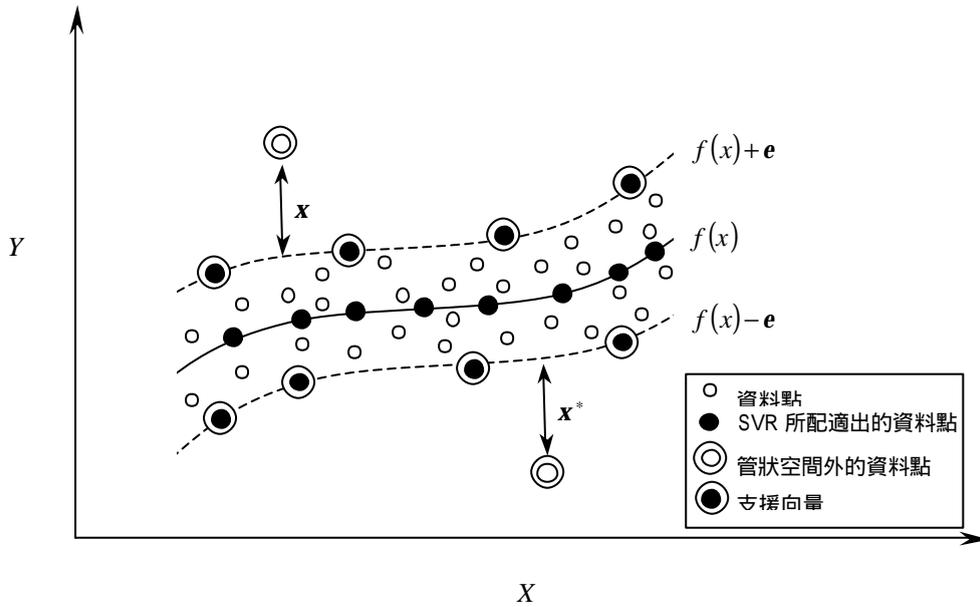


圖 1 導入 一不敏感損失函數的支援向量迴歸

為了能估計出 w 和 b ，在此將藉由式(4)中所導入之正的鬆弛變數 (slack variable) 而將式(2)轉換成新的目標函數，如下：

$$\begin{aligned} \text{Minimize} \quad & R_{reg}(w, x^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (x_i + x_i^*) \quad (4) \\ \text{Subjected to} \quad & d_i - wf(x_i) - b \leq e + x_i \quad x_i \geq 0 \\ & wf(x_i) + b - d_i \leq e + x_i^* \quad x_i^* \geq 0 \end{aligned}$$

最後，藉著拉格郎奇乘算子 (Lagrange multipliers) 的導入和最佳化的限制式，式(1)的決策函數將變成如下的簡單型式，其推導過程請詳見附錄 A：

$$f(x, a_i, a_i^*) = \sum_{i=1}^n (a_i - a_i^*) K(x, x_i) + b \quad (5)$$

2.2 拉格郎奇乘算子

在式(5)中， a_i 和 a_i^* 就是所謂的拉格郎奇乘算子，它們滿足 $a_i \times a_i^* = 0$ 、 $a_i \geq 0$ 及 $a_i^* \geq 0$ ， $i = 1, 2, \dots, n$ 。由於拉格郎奇乘算子導入式(4)中，而得到其最大化的對偶方程式，如下：

$$\begin{aligned} \text{Max} \quad & R(a_i, a_i^*) = \sum_{i=1}^n d_i (a_i - a_i^*) - e \sum_{i=1}^n (a_i + a_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) \quad (6) \\ \text{Subjected to} \quad & \sum_{i=1}^n (a_i - a_i^*) = 0, \quad 0 \leq a_i \leq C \quad i = 1, 2, \dots, n \\ & \quad \quad \quad \quad \quad \quad \quad \quad 0 \leq a_i^* \leq C \quad i = 1, 2, \dots, n \end{aligned}$$

以 Karush-Kuhn-Tucker (KKT) 解二次規劃問題的條件為基礎，式(6)中的 $(a_i - a_i^*)$ ，只有一些會被假設為非0值。這些係數為非0值的資料點之估計誤差將大於等於 e 且將被稱為是支援向量 (support vector)。也就是說，這些資料點將落在決策方程的 $-bound$ 之上或外側。參考式(6)，明顯的，所謂的支援向量只是指那些被用在決定決策方程的過程中，除去 $(a_i - a_i^*)$ 係數為0的其他資料點，它們對最佳解將具有決定性的影響力。相對的，在管狀空間內的點則對解並沒有直接

的貢獻。一般而言 愈大、支援向量就愈少，對於解而言，每個支援向量相對的重要性就越高，亦即解就越具有稀疏性(sparseness)，因此只要使用訓練資料集的一個小的子集合就可用來表示解，這在計算上具有相當大的優勢。然而，較大的卻也降低了訓練資料於估計時的精確度。所以，在這樣的概念下，將權衡於解的稀疏性和對資料點逼近能力之間。

2.3 核函數

式(6)中的 $K(x_i, x_j)$ 被定義為核函數(kernel function)，核函數的值為特徵空間 $f(x_i)$ 和 $f(x_j)$ 中兩向量 x_i 和 x_j 的內積，也就是說 $K(x_i, x_j) = f(x_i) * f(x_j)$ 。使用核函數的好處是能處理任意維度的特徵空間而不必須實際的去計算 $f(x)$ 。任何的函數如果能滿足 Mercer 條件的話 (Vapnik, 1995) 都能被拿來當做核函數。最典型的核函數範例為多項式核函數 (polynomial kernel function) : $K(x, y) = (x * y + 1)^d$ 和高斯核函數 (Gaussian kernel function) : $K(x, y) = \exp(-1/s^2(x - y)^2)$ 。在此 d 為多項式核函數的次方項；而 s^2 則為高斯核函數的頻寬 (bandwidth)。將來建立模型時，核函數的參數都必須很謹慎的選取，因為它定義了高維特徵空間的結構且也控制了最後解的複雜度。

三、實驗設計

3.1 資料集

為了和 Law and Au (1999) 所建構的類神經網路預測模型比較，本研究中所使用的資料集為 1967 年至 1996 年間，各年度日本人赴香港旅遊的人口數 (Number of visitors)、日本對香港的貨幣匯率 (Foreign exchange)、日本人口數 (Population)、日本國民支出毛額 (Gross domestic expenditure) 與香港旅館平均房價 (Average hotel rate) 等五項時間序列資料，這些資料都是正式官方統計文件所提供的資料，和 Law and Au 所蒐集的資料類似，Law and Au 的模型中運用了 6 個輸入變數 (Foreign exchange、Population、Gross domestic expenditure、Average hotel rate、Service Price、Marketing expenses) 架構類神經網路預測模型以預測日本人赴香港旅遊的人口數。該類神經網路屬前饋式類神經網路模型 (feed-forward neural network model)，該模型具有三個層次，輸入層有 6 個輸入神經元 (foreign exchange、population、gross domestic expenditure、average hotel rate、service price、marketing expenses)、隱藏層有 10 個神經元、輸出層有一個輸出神經元 (number of visitors)。其預測結果顯示，類神經網路的預測績效比簡算法、移動平均法、指數平滑法和複迴歸法等傳統預測技術優秀。

由於類神經網路具有前述的一些本質上的缺點，因此本研究將運用支援向量迴歸建構預測模型，以便能將預測結果和 Law and Au 的類神經網路預測模型進行比較。實證時，資料點將隨機的被分成兩部分，一部分為訓練資料集 (20 筆資料)、另一部分為測試資料集 (10 筆資料)。訓練資料集將用來訓練支援向量迴歸模型，然後再利利用所得到的模型對測試資料集中的資料點進行預測，並評估其預測效果。

3.2 實驗設計

利用支援向量迴歸模型進行預測的前提是認為某一因變數與某些自變數之間存在著某種函數關係，就本研究而言，所考量的自變數有旅館平均房價 (AHR)、匯率 (FER)、人口 (Pop) 與國民支出毛額 (GDE) 等，而因變數則為日本人赴香港旅

遊的人口數 (Y)，因此函數關係可描述如下：

$$Y = f(AHR, FER, Pop, GDE) \quad (7)$$

在此， $f(x)$ 的角色是不明確的，因此欲利用支援向量迴歸來擬合這種函數關係 $f(x)$ 。也就是說將利用各自變數所組成的每一組輸入資料 $(AHR_i, FER_i, Pop_i, GDE_i)$ ，透過支援向量迴歸的訓練、預測機制來推導出因變數 Y 的值。

3.3 模型評估

對於支援向量迴歸模型之預測績效，將以 MAPE(mean absolute percentage error)、NMSE(normalized mean squared error)、Z、R 等統計量來評估，這些統計量的定義如表 1 所示。MAPE 和 NMSE 可用來衡量真實值和預測值之間的差異程度，當 MAPE 和 NMSE 越小時，代表所預測的時間序列值會相當接近真實值。Z 值代表預測值之 MAPE 0.15 的百分比，R 值可評估真實值和預測值的逼近水準與相關程度。

表 1 評估模型的統計量

統計量	定義
MAPE	$MAPE = \frac{\sum_{i=1}^n a_i - p_i / a_i}{n} \times 100\%$
NMSE	$NMSE = 1 / (d^2 n) * \sum_{i=1}^n (a_i - p_i)^2$ $d^2 = 1 / (n - 1) * \sum_{i=1}^n (a_i - \bar{a})^2$
Z	$Z = \frac{\sum_{i=1}^n j}{n} \times 100\% \quad \text{for} \quad \begin{cases} j=1 & \text{if } \frac{a_i - p_i}{a_i} \leq 0.15 \\ j=0 & \text{otherwise} \end{cases}$
R	$R = \frac{\sum_{i=1}^n (a_i \times p_i)}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n p_i^2}}$

註：* a_i 和 p_i 分別代表真實值與預測值

四、建構支援向量迴歸模型

在本研究中，高斯函數(Gaussian function)將被用來當做支援向量迴歸的核函數，因為如果對資料的性質沒有任何事先的認知時，使用高斯函數將可得到較好的預測結果(Smola, 1998)，這概念也曾被以實驗的方式比較高斯核函數和多項式核函數所得的結果，也證明了使用高斯核函數可以得到較好的預測結果，而且若使用多項式核函數時，除了得到較差的預測結果外，其所花費在訓練支援向量迴歸模型的時間也比較長。

在建構支援向量迴歸模型的過程中，尚有三個參數需要由研究者自行決定，這三個參數分別是： C 、 γ 和 ϵ 。 C 為懲罰常數、 γ 為高斯核函數的頻寬、 ϵ 為不敏感區域的寬度。任意的選取這些參數，雖然也可以很容易的經由計算而得到訓練資料的訓練誤差，但是它隱含著具有過度配適的風險。為了避免過度配適的現象發生，可使用交叉驗證技術，但是交叉驗證的過程則是相當耗時的。Duan, Keerthi, and Poo(2001)曾運用了 k-fold 交叉驗證技術於分類辨識的領域，得到了不錯的效果，且該技術能於計算的時間成本和參數估計的可信度間取得適當的平衡。Duan 同時也建議 k 取 5，因為在他的實驗裡使用 5-fold 交叉驗證技術來訓練樣本資料時，能很適當的估計出支援向量迴歸模型的預測誤差。在此，所謂 5-fold 交叉驗證技術，即是將訓練資料集分成 5 等份，每等份具有相同個數的輸入資料，且無任何交集，然後利用其中的 4 等份以某一參數組合 $\{C, \gamma, \epsilon\}$ 來訓練支援向量迴歸模型，訓練完成後所得到的模型再用來測試剩餘的另一等份，並評估其預測誤差，重複此類實驗共 5 次，最後求得這 5 次預測的平均誤差，並以這平均誤差當作在該特定參數組合下所建構之模型的交叉驗證誤差，依參數組合之不同進行多次實驗，並分別紀錄各不同參數組合下所得到的交叉驗證誤差，取具有最小交叉驗證誤差時的參數組合，做為未來建構支援向量迴歸模型的最佳參數組合。

根據 Tay and Cao(2001)， γ 的取值範圍在 1 到 100 之間，而 C 的取值範圍在 10 到 100 之間時，所建構的支援向量迴歸模型將得到較佳的預測結果。為能更廣泛的尋找最佳狀態時的參數，在此將使用格子點法搜尋最佳參數組合，最佳參數組合的搜尋範圍如表 2。實驗將利用各格子點上三個參數之組合，以 5-fold 交叉驗證技術訓練支援向量迴歸模型，最後求取各格子點上的交叉驗證 MAPE。

表 2 參數的取值範圍

參數	取值範圍	增量
γ	0.1 $0.1 * 2^{19}$	$2 * 2$
C	1 100	1
	2^{-16} 2^0	$* 2$

運用格子點法配合 5-fold 交叉驗證技術，總共對訓練資料進行了 $20 * 100 * 17 * 5$ 次訓練(即建構了 $20 * 100 * 17 * 5$ 個支援向量迴歸模型)，於參數的選取決策上將由 $\gamma = 0.1$ 開始，圖 2 是各 γ 值下平均支援向量個數佔訓練樣本個數之比例，觀察圖 2 可發現，平均支援向量個數隨 γ 值而遞減，這個結果和 Vapnik(1995)所提出的支援向量個數是 γ 值的遞減函數之結論是一致的。另外根據 Tay and Cao(2001)的實驗，當支援向量個數佔訓練樣本個數之比例太大時，容易產生過度配適(over-fitting)或不足配適(under-fitting)的問題。因此本研究將先篩選出比例小於 90% 的 γ 值(0.003906 \sim 1)。圖 3 是各 γ 值下交叉驗證 MAPE 小於等於 15 的模型數。圖 3 所示，在 $\gamma = 0.007813$ 時，MAPE 小於等於 15 的模型數最多，代表以 $\gamma = 0.007813$ 訓練支援向量迴歸模型時，將可得到較穩定的預測結果。綜合觀察圖 2、圖 3 所獲得的結論，本研究中未來將選取 $\gamma = 0.007813$ 建立支援向量迴歸模型。

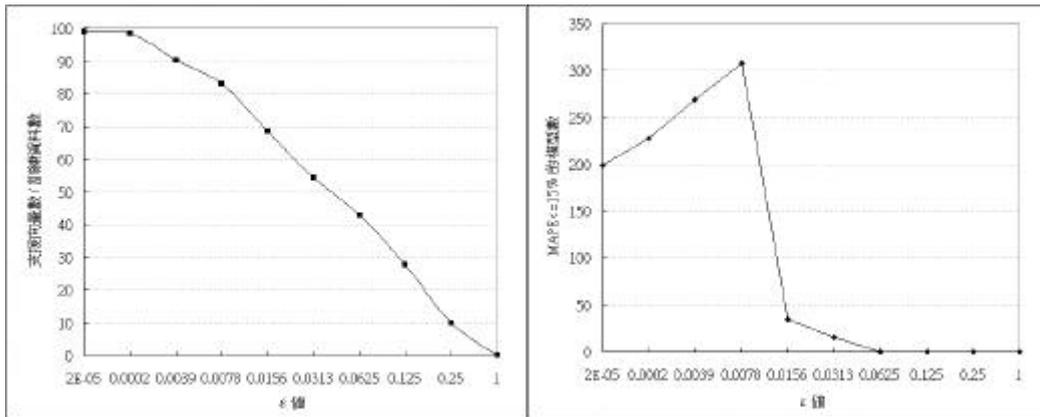


圖 2 各 λ 值下平均支援向量個數佔訓練樣本個數之比例

圖 3 各 λ 值下交叉驗證 MAPE 小於等於 15 的模型數

值確定之後，接下來將選取 λ^2 值和 C 值。圖 4 是當 $\lambda = 0.007813$ 時，各種不同的 λ^2 、C 值組合下交叉驗證的 MAPE，觀察得知當 $\lambda^2 = 1.6$ 、C=34 時交叉驗證的 MAPE 最小。因此未來可選取 $\lambda^2 = 1.6$ 、C=34 以建立支援向量迴歸模型。

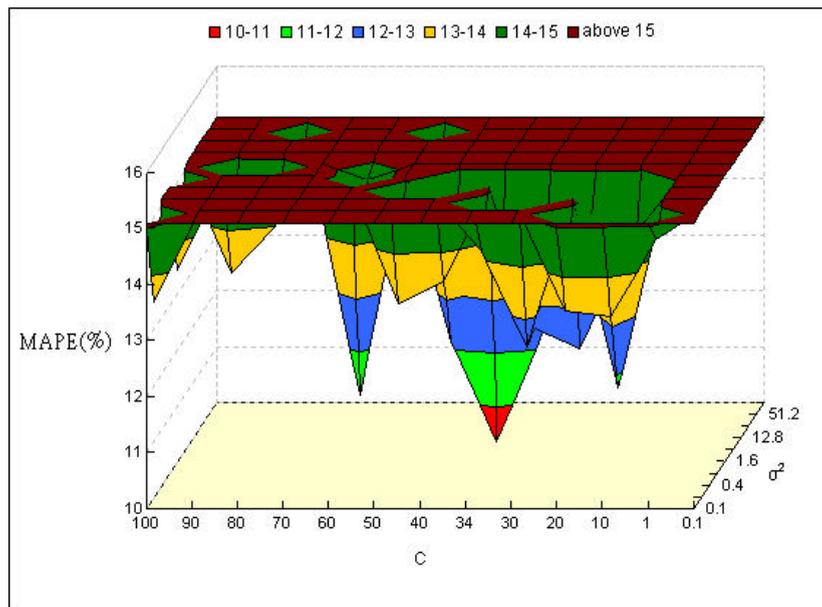


圖 4 當 $\lambda = 0.007813$ 時，各個 λ^2 、C 值組合下交叉驗證 MAPE

五、實證結果

經對訓練資料進行 5-fold 交叉驗證後，將選定 C=34、 $\lambda^2 = 1.6$ 、 $\lambda = 0.007813$ 以建構支援向量迴歸之預測模型。這個支援向量迴歸預測模型對測試資料集進行模擬預測後，所得到的結果如表 3、表 4 與圖 5。

表 3 日本人赴香港之旅遊需求的預測結果

年	真實值	類神經網路模型 (Law)		支援向量迴歸模型	
		預測值	MAPE(%)	預測值	MAPE(%)
1970	168473	139388	17.26	170201	1.025684
1971	237950	208801	12.25	263008	10.53078
1972	349212	354247	1.44	317684	9.028327
1973	486677	421785	13.33	383706	21.15798
1975	382740	428112	11.86	428664	11.99875
1980	472182	509941	8.00	458473	2.90333
1987	1033525	1066340	3.18	1067517	3.288938
1988	1240470	1122840	9.48	1186749	4.330697
1993	1280905	1414250	10.41	1409003	10.00059
1994	1440632	1710490	18.73	1652343	14.6957
MAPE			10.59		8.896077

表 4 不同預測模式的精確度比較

預測模型	MAPE	Z	r
類神經網路模型(Law)	10.59	70%	0.9851
支援向量迴歸模型	8.896077	90%	0.9964

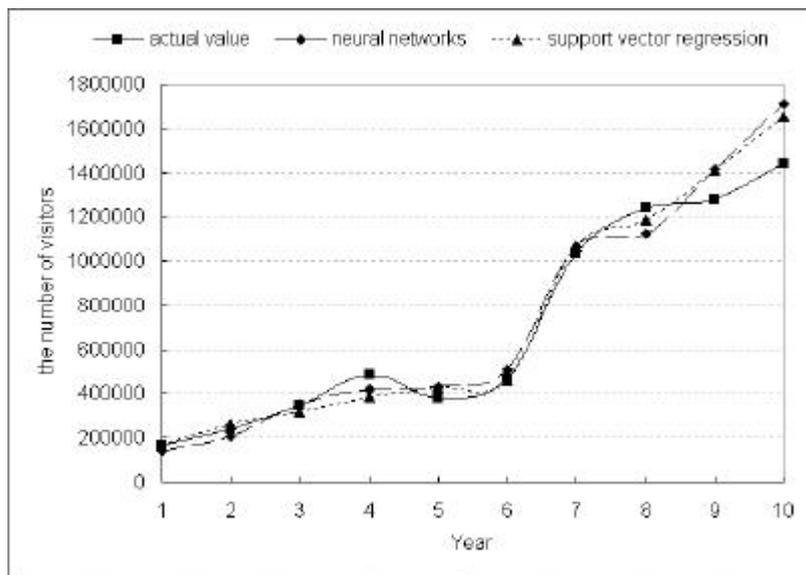


圖 5 不同預測模式的預測圖

由實驗結果的列表可知，支援向量迴歸模型所得的預測結果相當的接近真實值，換句話講，支援向量迴歸模型所得的預測結果是相當精確的，而較小的 MAPE 值 (平均 MAPE=8.896077 %) 也說明了預測結果和真實值之間的離差是相當小的，依據 Lewis(1982)的等級評定，可評定支援向量迴歸模型的預測結果是具有高度準確性的，甚至比類神經網路模型更準確。而由 Z 值來看，支援向量迴歸模型的預測值中，MAPE 值小於等於 15 的預測值有 90%，然類神經網路模型卻只有 70% 而已，顯示支援向量迴歸模型具有優越且穩定的預測能力。此外，以正規化的相關係數來看，支援向量迴歸模型和類神經網路模型都相當高，且都相當接近於 1，但

還是以支援向量迴歸模型之預測結果的相關係數較高。這意味著支援向量迴歸模型之預測結果和真實值之間具有很高的相關性。

六、支援向量迴歸模型的敏感度分析

在上面的實證過程中，運用了 5-fold 交叉驗證技術而得到核函數之參數 γ 、 C 和 σ^2 的值，在這節中，將探討模型的 NMSE 和這 3 個參數之間的關係，以便能更深入了解這些參數對支援向量迴歸模型的重要性並獲知支援向量迴歸模型的一些特性。

圖 6(a)中先將 C 和 σ^2 固定為 34 和 0.007813，而觀察在不同的 γ 值下 NMSE 的變化情形，圖中顯示出訓練資料集的 NMSE 隨著 γ 的增加而增加，另一方面，在測試資料集中，隨著 γ 的增加，NMSE 先是緩慢遞減，然後微幅遞增再遞減，最後再急速的增加。這暗示著愈小的 γ 值 (< 0.4)，在預測時會產生過度擬合的現象，而太大的 γ 值 (> 102.4)則會產生不足擬合的現象。因此 γ 的預估值大約在 0.8 到 12.8 之間，因為這段期間訓練資料集和測試資料集的 NMSE 較為穩定，較不易有過度擬合或不足擬合的現象發生。在這樣的觀點之下，我們可說在支援向量迴歸的預測效果中， γ 將扮演著重要的角色，因為不當的選取 γ 將產生過度擬合或不足擬合的現象，而這將降低預測模式的可信度。而圖 6(b)則顯示了當 γ 漸增時，支援向量的數量先是遞增，然後遞減，最後再遞增。特別值得注意的是，在不足擬合時約有 90% 的訓練資料點都收斂變成了支援向量。

圖 7(a)是當 γ 仍固定為 0.007813，而 σ^2 的值設為 1.6 時，不同的 C 值下 NMSE 的變化情形，從圖中可觀查出，在訓練資料集中，隨著 C 值的遞增 NMSE 會遞減，但相對的，在測試資料集中，當 C 從 0.1 增加到 1 時，NMSE 會遞增，然後在 C 從 1 變化到 100 的過程中 NMSE 遞減，而從 100 變化到 10000 的過程中 NMSE 會遞增。其原因可能為較小的 C 值 (< 1) 將不足擬合訓練資料集，因為放在訓練資料集上的權重太小，因此導致在訓練資料集會有較大的 NMSE 出現，相反的，當 C 值太大 (> 100)，支援向量迴歸將過度擬合訓練資料集，導致較差的擬合結果，在此情形下， C 值大約可以取為 1 到 100 之間，因為這段期間訓練資料集和測試資料集的 NMSE 同時穩定的緩慢遞減，較不易有過度擬合或不足擬合現象的發生。此外由圖 7(b)中亦可發現，當不足擬合、過度擬合現象發生時，分別有高比例 (95%、80%) 的訓練資料點都收斂變成了支援向量。

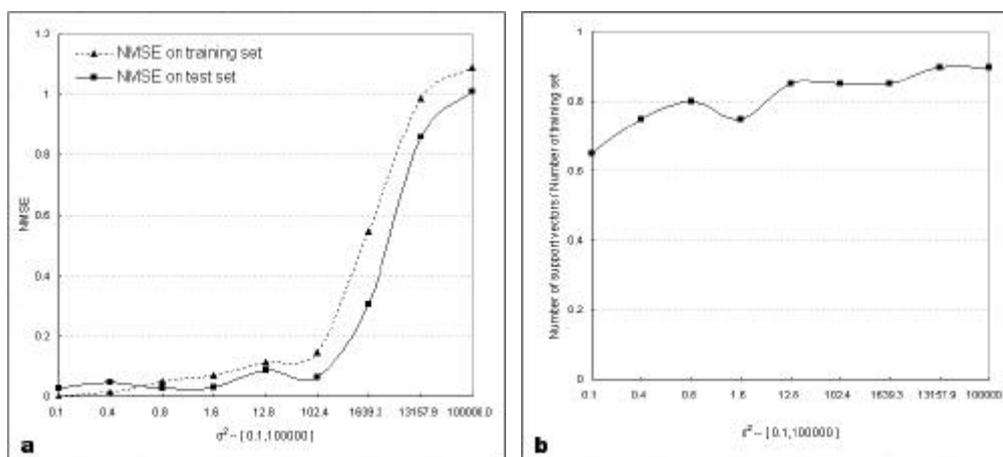


圖 6(a) 固定 C 和 σ^2 ，變動 γ 值時，NMSE 的變化

圖 6(b) 固定 C 和 σ^2 ，變動 γ 值時，支援向量數量的變化

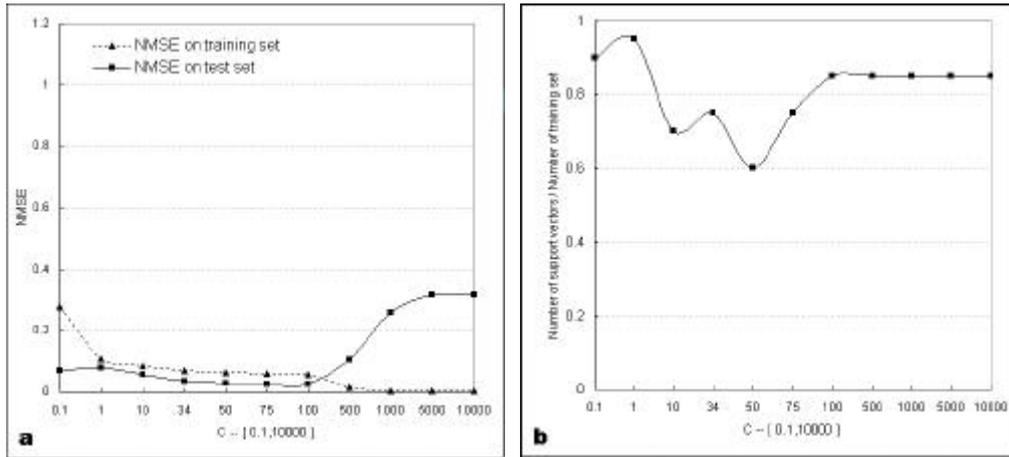


圖 7 (a) 固定 λ^2 和 ϵ ，變動 C 值時，NMSE 的變化

圖 7 (b) 固定 λ^2 和 ϵ ，變動 C 值時，支援向量數量的變化

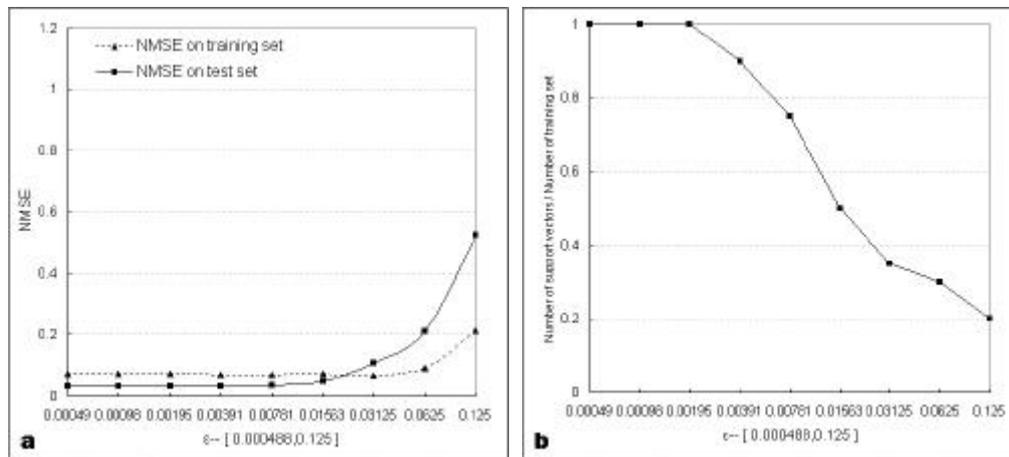


圖 8(a) 固定 λ^2 和 C，變動 ϵ 值時，NMSE 的變化

圖 8(b) 固定 λ^2 和 C，變動 ϵ 值時，支援向量數量的變化

圖 8(a)是當 λ^2 值固定為 1.6 和 C 值固定為 34 時，在不同的 ϵ 值下 NMSE 的變化情形，圖 8(a)顯示了，當 ϵ 值小於 0.007813 時不管在訓練資料集或測試資料集中 NMSE 都相當的穩定，這意味著在這段期間內 NMSE 是不受 ϵ 值所影響的。但是當 ϵ 值超過 0.03125 時，訓練和測試資料集的 NMSE 卻都急速的增加，這可能是因為訓練資料集中的資料點本來就少，又因當 ϵ 值增大時，可用以擬合估計的支援向量個數就更少，因此最佳解的代表性不足，導致產生較大的 NMSE。這點可從圖 8(b) 得到印證，當 ϵ 值超過 0.03125 時支援向量的個數是相當少的，這個結果和 Vapnik(1995)所提出的支援向量個數是 ϵ 值的遞減函數之結論是一致的。

七、結論

本研究使用了支援向量迴歸技術於國際旅遊需求的預測領域中，為了建構穩健、可信賴的預測模型，需慎選建構模型時所必需的參數 λ^2 、C 和 ϵ 。由於 5-fold 交叉驗證的訓練誤差可用於估計模型的預測誤差，因此先使用格子點法配合 5-fold 交叉驗證技術於訓練資料集上，以求得各種不同參數組合下的交叉驗證 MAPE，再藉由本研究所提出的以支援向量的數量為主、評估交叉驗證 MAPE 為輔的參數選取機制，篩選出最佳的參數組合，最後再利用此參數組合建構實際的支援向量迴歸預測模型。其結論為在國際旅遊需求之預測領域中，支援向量迴歸確實是個

可信任的預測工具，其預測能力甚至比類神經網路更精確。此外為了驗證參數對支援向量迴歸模型的重要性並獲知支援向量迴歸模型的一些特性，研究中使用了敏感度分析技術，分析中也證實了，不當的選取參數將使模型容易陷於過度擬合或不足擬合的危機中。這些結論是由以下的原因所架構起來的：

- 一、無論在類神經網路或支援向量迴歸中，都可使用交叉驗證技術來訓練模型，以建構較可信賴的預測模型，然這需要花費相當多的時間與經驗且須帶點運氣。在支援向量迴歸中，由於使用者需自定義的參數較少，因此可簡單的使用敏感度分析探討參數選取的技巧，以避免過度擬合或不足擬合等現象的發生，然而使用類神經網路進行預測時，需有相當多的參數待決定，如隱藏層的數目、隱藏層的神經元數、學習率、動量項、代數、轉換函數、初始加權方法等。每一個參數都必須根據經驗而確實的被決定出來，因此為了要得到一個較好的預測結果，需努力的先去得到一個最佳化的參數組合，而這真是一個艱巨的任務呀！而且也無法保證預測過程不會有過度擬合或不足擬合的情形存在，而這通常會被認為是類神經網路的一個弱點。
- 二、支援向量迴歸實作了結構風險最小化原則，它最小化了預測誤差的上界而不是最小化了訓練誤差。這個重要的本質導致了支援向量迴歸的預測誤差較以實作經驗風險最小化的類神經網路的結果要來得好。
- 三、類神經網路將可能無法收斂於全域解。然而在支援向量迴歸中，訓練支援向量迴歸的過程等同於求解一連串的線性且具限制式的二次規劃過程，且支援向量迴歸的解是獨立的、最佳的、全域的。

在支援向量迴歸之參數的敏感度分析研究中，證明了 γ 、 C 在預測的績效上扮演著重要的角色，不當的選取這兩個參數將導致訓練資料過度擬合或不足擬合現象的發生，雖然支援向量迴歸之預測績效對於 γ 來說較不敏感，但是當使用較大的 γ 時，支援向量的數量將大幅度的下降，因此最佳解將較不具有代表性，導致得到較差的預測效果。以上述所得到的結果為基礎，因此當考慮到這些參數對預測效果之顯著影響時，發展一個選取參數的結構化方法，將變成一件很重要的工作。和類神經網路比較起來，由於支援向量迴歸能提供這麼多的優點，將來的研究更應致力於探索更精進的支援向量迴歸技術以追求更具效率且更好的預測效果。

參考文獻

1. Chang, C.C. & Lin, C.J., LIBSVM : a library for support vector machines, Technical Report, Department of Computer Science and Information Engineering, Nation Taiwan University, Available at <http://www.csie.edu.tw/~cjlin/papers/libsvm.pdf>, 2001.
2. Chan, Y.M., “Forecasting tourism: A sine wave time series regression approach”, Journal of Travel Research, 32(2), pp.58-60, 1993.
3. Cheng, W., Wanger, L., & Lin, CH., “Forecasting the 30-year US treasury bond with a system of neural networks”, Journal of Computational Intelligence in Finance, 1(4), pp.10-15, 1996.
4. Chiang, WC., Urban, TL., & Baildridge, G., “A neural network approach to mutual fund net asset value forecasting”, Omega, 24(2), pp.205-219, 1996.
5. Chu, Fong-Lin, “Forecasting Tourism: a Combined Approach”, Tourism Management, 19(6), pp.515-520, 1998.
6. Duan, K., Keerthi, S. & Poo, A., Evaluation of simple performance measures for tuning SVM hyperparameters. Technical Report, Department of Mechanical

- Engineering, National University of Singapore, 2001.
7. Haykin, S. , Neural networks: a comprehensive foundation. Englewood Cliffs, NJ: Prentice Hall, 1999.
 8. Joachimes, T. , Text categorization with support vector machines. Technical Report , [ftp://ftp-ai.informatik.unidortmund.de/pub/Reports/report23.ps.z](http://ftp-ai.informatik.unidortmund.de/pub/Reports/report23.ps.z), 1996.
 9. Kaastra, I. & Milton, SB., “Forecasting futures trading volume using neural networks”, The Journal of Futures Markets, 15(8), pp.853-970,1995.
 10. Kulendran, N. & King, Maxwell., “Forecasting International Quarterly Tourist Flows Using Error-Correction and Time-Series Model”, International Journal of Forecasting, 31, pp.319-327, 1997.
 11. Law, Rob . & Au, Norman., “A Neural Network Model to Forecast Japanese Demand for Travel to Hong Kong”, Tourism Management, 20, pp.89–97, 1999.
 12. Law, R., “Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting”, Tourism Management, 21, pp.331–340, 2000.
 13. Lewis, C.D., International and Business Forecasting Methods, London : Butterworths, 1982.
 14. Manuel, V. and Croes, R. , “Evaluation of Demand US Tourists to Aruba”, Annals of Tourism Research , 27(4), pp.946–963, 2000.
 15. Martin, Christine A., & Stephen, F. Witt., “Forecasting Tourism Demand : A Comparison of the Accuracy several Quantitative Methods”, International Journal of Forecasting, 5, pp.7-19, 1989
 16. Pattie, Douglas C. & Snyder, John., “Using a neural network to forecast visitor behavior”, Annals of Tourism Research, 23(1), pp.151-164, 1996.
 17. Pilar, Gonzalez. & Moral, Paz., “An Analysis of the International Tourism Demand in Spain”, International Journal of Forecasting, 11, pp.233-251, 1995.
 18. Schmidt, M. , “Identifying speaker with support vector networks”, Interface ‘96 Proceedings, Sydney, 1996.
 19. Sheldon, Pauline J. & Var, Turgut. , “Tourism Forecasting: A Review of Empirical Research”, Journal of Forecasting, 4(2), pp.183-195, 1985.
 20. Sharda, R. & Patil, R. B. , “Connectionist Approach to Time Series Prediction : An Empirical Test”, Journal of Intelligent Manufacturing, 3, pp.317-323, 1992.
 21. Smeral, E., Witt, S. F. & Witt, C. A., “Econometric Forecasts: Tourism Trends to 2000”, Annals of Tourism Research, 19(3), pp.450-466, 1992.
 22. Smola AJ., Learning with Kernels. PhD Thesis, GMD, Birlinghoven, Germany, 1998.
 23. Tay, Francis. E. H. & Cao, Lijuan., “Financial Forecasting Using Support Vector Machines”, Neural Computing & Application, 10, pp.184–192, 2001a.
 24. Tay, Francis. E. H. & Cao, Lijuan., “Application of support vector machines in financial time series forecasting”, Omega, 29(4), pp.309–317, 2001b.
 25. Uysal, M. & Crompton, J.L., “Determinants of Demand for International Tourist Flow to turkey”, Tourism Management, 15(4), pp.288-297, 1984.
 26. Vapnik, V., The nature of statistical learning theory, New York: Springer, 1995.
 27. Vapnik, V., Golowich, S., & Smola, A., Support vector method for function approximation , regression estimation and signal processing, Advance in Neural information Processing System 9. (pp. 281-287).Cambridge:MIT Press, 1997.
 28. Van, E. & Robert, J., The application of neural networks in the forecasting of share price, Haymarket, VA, USA: Finance & Technology Publishing, 1997.
 29. Witt, S. F. & Witt, C. A., Modeling and Forecasting Demand in Tourism, London: Academic Press, 1992.

附錄 A

給定一個資料點的集合 $G = \{(x_i, d_i)\}_i^n$ (x_i 是輸入向量、 d_i 是目標值、 n 是樣本資料量)，支援向量迴歸的基本概念是透過一個非線性映射 Φ ，將資料 x 映射到高維特徵空間，並在這個空間進行線性迴歸。即

$$f(x) = \mathbf{w}\Phi(x) + b \quad (1)$$

$$\Phi: R^n \rightarrow F, \mathbf{w} \in F$$

在遵循結構風險最小化原則之下，藉著參數 C 的導入，為使支援向量迴歸的誤差能權衡於模型複雜度與經驗誤差之間，因而定義式(2)的懲罰風險函數 (regularized risk function)，同時藉著最小化式(2)的過程而估計出 \mathbf{w} 和 b 的係數。

$$R_{reg}(C) = R_{emp} + \frac{1}{2} \|\mathbf{w}\|^2 = C \times \frac{1}{n} \sum_{i=1}^n L_e(d_i, y_i) + \frac{1}{2} \|\mathbf{w}\|^2, \quad (2)$$

$$L_e(d, y) = \begin{cases} |d - y| - e & |d - y| \geq e \\ 0 & otherwise \end{cases} \quad (3)$$

為了能估計出 \mathbf{w} 和 b ，在此將藉由導入正的鬆弛變數 (slack variable) 而將式(2)轉換成新的目標函數，如下：

$$\text{Minimize} \quad R_{reg}(\mathbf{w}, \mathbf{x}^{(*)}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\mathbf{x}_i + \mathbf{x}_i^*) \quad (4)$$

$$\text{Subjected to} \quad \begin{cases} d_i - \mathbf{w}\mathbf{f}(x_i) - b \leq e + \mathbf{x}_i \\ \mathbf{w}\mathbf{f}(x_i) + b - d_i \leq e + \mathbf{x}_i^* \end{cases} \quad \mathbf{x}^{(*)} \geq 0$$

藉著拉格郎奇乘算子 (Lagrange multipliers) 的導入，而建構拉格郎奇函數。

$$\begin{aligned} L(\mathbf{w}, b, \mathbf{e}^{(*)}) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\mathbf{x}_i + \mathbf{x}_i^*) \\ &\quad - \sum_{i=1}^n \mathbf{a}_i (\mathbf{e} + \mathbf{x}_i - d_i + \mathbf{w}\mathbf{f}(x_i) + b) \\ &\quad - \sum_{i=1}^n \mathbf{a}_i^* (\mathbf{e} + \mathbf{x}_i^* + d_i - \mathbf{w}\mathbf{f}(x_i) - b) \\ &\quad - \sum_{i=1}^n (\mathbf{m}_i \mathbf{x}_i + \mathbf{m}_i^* \mathbf{x}_i^*) \end{aligned} \quad (5)$$

式(5)中， $\mathbf{a}_i, \mathbf{a}_i^*, \mathbf{m}_i, \mathbf{m}_i^* \geq 0$ ，它們都是拉格郎奇乘算子，在鞍點 (saddle point) 上， L 對於 $\mathbf{w}, b, \mathbf{x}_i, \mathbf{x}_i^*$ 的微分都為 0，因此

$$\begin{aligned}
 \frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^n (\mathbf{a}_i - \mathbf{a}_i^*) \mathbf{f}(x_i) = 0 \\
 \frac{\partial L}{\partial b} &= \sum_{i=1}^n \mathbf{a}_i - \sum_{i=1}^n \mathbf{a}_i^* = 0 \\
 \frac{\partial L}{\partial \mathbf{x}_i} &= \mathbf{C} - \mathbf{a}_i - \mathbf{m}_i = 0 \\
 \frac{\partial L}{\partial \mathbf{x}_i^*} &= \mathbf{C} - \mathbf{a}_i^* - \mathbf{m}_i^* = 0
 \end{aligned} \tag{6}$$

重寫式(7)的方程式可得到：

$$\begin{aligned}
 \mathbf{w} &= \sum_{i=1}^n (\mathbf{a}_i - \mathbf{a}_i^*) \mathbf{f}(x_i) \\
 \sum_{i=1}^n \mathbf{a}_i &= \sum_{i=1}^n \mathbf{a}_i^* \\
 \mathbf{a}_i^{(*)} &\in [0, C]
 \end{aligned} \tag{7}$$

將式(7)代入到式(5)中，並應用對偶理論，可得到如下的對偶問題

$$\text{Max } R(\mathbf{a}_i, \mathbf{a}_i^*) = \sum_{i=1}^n d_i (\mathbf{a}_i - \mathbf{a}_i^*) - \mathbf{e} \sum_{i=1}^n (\mathbf{a}_i + \mathbf{a}_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{a}_i - \mathbf{a}_i^*) (\mathbf{a}_j - \mathbf{a}_j^*) K(x_i, x_j) \tag{8}$$

$$\text{Subjected to } \sum_{i=1}^n (\mathbf{a}_i - \mathbf{a}_i^*) = 0, \quad \begin{matrix} 0 \leq \mathbf{a}_i \leq C & i = 1, 2, \dots, n \\ 0 \leq \mathbf{a}_i^* \leq C & i = 1, 2, \dots, n \end{matrix}$$

在此， $K(x_i, x_j) = \mathbf{f}(x_i) \mathbf{f}(x_j)$ ，解此二次規劃問題，由式(7)可得最佳化時的 \mathbf{w}

$$\mathbf{w}^* = \sum_{i=1}^n (\mathbf{a}_i - \mathbf{a}_i^*) \mathbf{f}(x_i) \tag{9}$$

因而將式(9)代入式(1)可得

$$f(x, \mathbf{a}_i, \mathbf{a}_i^*) = \sum_{i=1}^n (\mathbf{a}_i - \mathbf{a}_i^*) K(x, x_i) + b \tag{10}$$

接著，可由互補的鬆弛條件 (complementary slackness conditions) 解最佳的 b 值，如下：

$$\begin{aligned}
 \mathbf{a}_i (\mathbf{e} + \mathbf{x}_i - d_i + \mathbf{w} \mathbf{f}(x_i) + b) &= 0 \\
 \mathbf{a}_i^* (\mathbf{e} + \mathbf{x}_i^* + d_i - \mathbf{w} \mathbf{f}(x_i) - b) &= 0 \\
 (\mathbf{C} - \mathbf{a}_i) \mathbf{x}_i &= 0 \\
 (\mathbf{C} - \mathbf{a}_i^*) \mathbf{x}_i^* &= 0
 \end{aligned} \tag{11}$$

從式(8)中可得到一些有用的結論：

- 1、 $\mathbf{a}_i^* = C$ 時，即表示資料點 (x_i, d_i) 將落在 一不敏感損失函數的管狀空間之外。

2. $a_i a_i^* = 0$ 時，則任何成對的對偶變數 a_i 、 a_i^* 將不可同時為 0，亦即 $a_i = 0$ 時， a_i^* 不能為 0，反之亦然，此外在上、下兩個方向上鬆弛變數都不能為 0。
- 3 如果 $a_i^{(*)} \in (0, C)$ 而 $x_i^{(*)} = 0$ ，即表示資料點 (x_i, d_i) 將落在 邊界 $(f(x) + e$ 或 $f(x) - e)$ 上。因此最佳的 b 值可藉由計算得到：

$$\begin{array}{ll} b^* = d_i - \mathbf{w}^* \mathbf{f}(x) - e & a_i \in (0, C) \\ b^* = d_i - \mathbf{w}^* \mathbf{f}(x) + e & a_i^* \in (0, C) \end{array}$$