

最小絕對壓縮挑選機制(LASSO)於天氣迴歸預報的應用

王政忠 陳雲蘭
中央氣象局氣象預報中心

摘要

在建立天氣預報迴歸方程的過程中，預報因子(解釋變數)往往具有很強的共線性且具有龐大的數量。在共線性的影響下，若採用一般常用的最小平方方法(Ordinary least squares)來建立模式，迴歸係數估計顯得很不穩定；傳統變數選取方法-逐步迴歸方法(stepwise)雖可對待選預報因子做挑選，但離散型的選取方式易受到資料些微變動而影響結果，而且一旦預報因子過多將影響其運算效率。為改善迴歸係數估計的穩定性及逐步迴歸方法的缺點，本研究將介紹壓縮係數的概念，引進同時具有壓縮係數與挑選變數的最小絕對壓縮挑選機制(LASSO)，並以台北站日均溫資料為例，分別使用LASSO方法、複迴歸法、逐步迴歸法建立預報方程，再經十年交叉驗證資料分析比較，以驗證壓縮係數迴歸方法在天氣預報可能的效益。

關鍵字：變數選取、迴歸預報、LASSO

一、前言

迴歸分析目的主要就是要去探索反應變數(Y)與解釋變數(X_1, \dots, X_d)之間的關係，並希望利用這個關係經由觀察(X_1, \dots, X_d)去對反應變數(Y)做預測。如果分析目的只是為了做預測，我們可以採納所有解釋變數去做預測而不遺漏任何可用資訊。但引入變數過多，將使得預測偏誤(bias)較小，而預測變異(variance)較大；相反的引入變數太少，將使得預測偏誤(bias)較大，而預測變異(variance)較小，這之間存在著權衡關係。因此在建立迴歸模式的過程中，最重要的就是如何決定哪些變數可納入於模型及在可用的變數中去建立一個好的預測。傳統的變數選取方法有子集挑選(subset selection)、逐步迴歸(stepwise regression)；其中子集挑選即在所有迴歸組合中，利用某些準則去選定其中一子集使得準則達到最佳，當解釋變數的量太大，將因計算量龐大而受限制；逐步迴歸則利用序列假設檢定去決定合適的子集，雖然計算量較小，就某些準則而言，它並不保證可以達到最佳。在另一方面，壓縮係數方法則是放棄傳統最小平方估計和子集挑選，利用所有解釋變數配適模型，控制偏誤的估計大小來做變數選取。其中如脊迴歸(ridge regression)(Hoerl and Kennard, 1970)、LASSO(Tibshirani, 1996)等，已經成為各領域在處理多維度資料的重要工具。

在天氣資料中建立迴歸預報模式時，面臨解釋變數間具有高共線性且大量解釋變數個數的挑戰。在共

線性的影響下，使得傳統最小平方方法對迴歸係數估計不穩定；雖然最小平方估計為最佳線性不偏估計，但是龐大的解釋變數個數，將造成迴歸係數估計的變異很大，而迴歸模式建立後對於未來預報的精確度也因而降低。利用傳統變數選取方法可以減少變數個數，而且適當的加入相關係數判斷可以排除共線性變數，但離散的變數選取過程，資料些微變動易使挑選結果出現極大差異而降低預測準確度。那麼是否有方法可以改善傳統變數選取缺點，而且在建立迴歸預報模式後可以提供穩定的預測，並可兼顧建立預報模式的時效性？

為了改善傳統變數選取的缺點，本研究引進壓縮係數法於建立天氣預報模式時做變數選取，改以偏誤的估計去降低迴歸係數估計的變異，進而提高預測準確度。根據壓縮係數法可有效提供迴歸模式穩定度的特性，本研究將以天氣實例資料建立迴歸模式進行預報測試，進一步驗證此法在天氣預報的可能效益。

二、方法

2.1 壓縮的概念

壓縮(shrinkage)的概念源自於 James 與 Stein(1961)所提。假設觀察 d 維度 Z 向量且 $d > 2$ ，即 $Z = (Z_1, Z_2, \dots, Z_d)$ ，並假設 Z 向量服從 d 維度多元常態分布，其期望值為 θ ，其共變異數矩陣為 $\sigma^2 I_d$ ，即 $Z \sim N_d(\theta, \sigma^2 I_d)$ ，其中

$E(Z_i) = \theta_i, i = 1, \dots, d$ 。顯然地， θ 的最大似估計量

為 Z 。現在考慮計算 $\|Z\|^2$ 的期望值，即

$$E(\|Z\|^2) = \sum_{i=1}^d E(Z_i^2) = \sum_{i=1}^d (\sigma^2 + \theta_i^2) = d\sigma^2 + \|\theta\|^2$$

可發現 $\|Z\|^2$ 的期望值大於 $\|\theta\|^2$ ，所以此估計量 Z 以此觀點

來看並不讓人滿意。因此以 $\|Z\|^2$ 的期望值來看， Z 估計量可能估的太大，隱含了將 Z 壓縮(shrinkage)的可能性。

現在考慮另一估計 $\tilde{\theta}$ 為 $cZ, 0 < c < 1$ 。雖然 $\tilde{\theta}$ 是偏的(biased)估計量，但是可利用微分去找到合適的 c 使得 $\tilde{\theta}$ 具有最小的均方誤差(MSE)。即考慮：

$$f(c) = E(\|\tilde{\theta} - \theta\|^2) = c^2 d\sigma^2 + (1-c)^2 \|\theta\|^2$$

經微分可得到理想的估計 $\hat{\theta}^c$ 為：

$$\hat{\theta}^c = cZ = \left(1 - \frac{d\sigma^2}{d\sigma^2 + \|\theta\|^2}\right) Z$$

由於 $\|\theta\|^2$ 未知，可利用上述推導所得 $\|Z\|^2$ 為 $d\sigma^2 + \|\theta\|^2$

的不偏估計量，將 $\tilde{\theta}$ 改為為：

$$\tilde{\theta} = cZ = \left(1 - \frac{d\sigma^2}{\|Z\|^2}\right) Z$$

James 與 Stein(1961)證得 $(1 - b/\|Z\|^2)Z$ 此種形式的估計，

其中當 $b = (d-2)\sigma^2$ 且 $(d > 2)$ 時，具有最小的MSE。

即 James-Stein 估計量(JS)

$$\hat{\theta}^{JS} = \left(1 - \frac{(d-2)\sigma^2}{\|Z\|^2}\right) Z$$

由此可見，根據估計偏誤與估計變異權衡關係可知，若放棄估計不偏性，適當的對估計做壓縮，犧牲一些估計偏誤而降低估計變異，因而得均方誤差(MSE)更小的估計。相同的概念引伸到迴歸模型之迴歸係數估計，高斯-馬可夫定理中闡述

在所有線性不偏估計中，最小平方估計具有最小的MSE，但如果在具有偏誤的估計量中是否可以找到使得 MSE 更小的估計量？由上述的概念可以得知答案是確定的。來看一個簡單的例子，考慮兩個迴歸係數估計，一個為傳統最小平方方法估計，即 $\hat{\beta}_k^{OLS}$ ，另一個將傳統最小平方方法估計乘上一個壓縮常數，稱此估計為壓縮係數估計，即：

$$\hat{\beta}_k^{shrink} = \frac{1}{1+\lambda} \hat{\beta}_k^{OLS}$$

分別計算其迴歸係數估計的MSE總合，可得到MSE總合如下：

$$\begin{aligned} \sum_{k=1}^d E[(\hat{\beta}_k^{OLS} - \beta_k)^2] &= d\sigma^2 \\ \sum_{k=1}^d E[(\hat{\beta}_k^{shrink} - \beta_k)^2] &= d\sigma^2 \left(\frac{1}{1+\lambda}\right)^2 + \left(\frac{\lambda}{1+\lambda}\right)^2 \sum_{k=1}^d \beta_k^2 \end{aligned}$$

由於最小平方估計為不偏估計，其MSE總合為一個常數；而壓縮係數估計其MSE總合成變異項與偏誤平方項，隨著不同的 λ 值而改變，而MSE變化可由(圖1)展現出適當的挑選 λ 值，偏誤估計可以讓MSE降低。

2.2 OLS、Ridge 與 LASSO

傳統最小平方方法(ordinary least square)的估計量，在一般迴歸模式的假設下為不偏估計量，但其預報精確度(prediction accuracy)差，且不具有變數選取功能，使得解釋模型困難，並得依賴其他的統計準則，如配適度(goodness of fit)、預報誤差(prediction error)、梅洛 Cp (Mallow's Cp)準則等，才可達到變數選取目的。近年來壓縮係數法發展，將傳統最小平方估計加入合適的限制式，由於迴歸係數限制於某一個合理範圍下，所得的迴歸係數估計就具有壓縮的效果，並達到改善預報精確度目的。

脊迴歸(ridge regression)(Hoerl and Kennard, 1970)其方法是將傳統最小平方估計加入一個 2-norm 的限制式，即

$$\min_{\beta} \|Y - X\beta\|^2 \text{ subject to } \|\beta\|^2 = \sum_{j=1}^d \beta_j^2 \leq t$$

利用拉格朗日(Lagrange)方法可以去解出

$\hat{\beta}^{Ridge} = (X'X + \lambda I)^{-1} X'Y$ ，其中 λ 由限制式的門檻值 t 所決定。在限制式下，解決了原本反矩陣因共線性造成迴歸係數估計變異膨脹的問題，而改善預報精確度，但是它無法將任何一個變數壓縮到零，保留了所有解釋變數，無法得到

簡化模型的目的。

最小絕對壓縮挑選機制(Least absolute shrinkage and selection operator)由 Tibshirani 在 1996 年所提，簡稱 LASSO，其方法是將傳統最小平方估計加入一個 1-norm 的限制式，

$$\text{即：} \min_{\beta} \|Y - X\beta\|^2 \text{ subject to } \|\beta\| = \sum_{j=1}^d |\beta_j| \leq t$$

並考慮改善脊迴歸(ridge regression)無去簡化模型取及子集挑選(subset selection)易受到資料變動影響選取結果的缺點。其計算在此限制式下，除了係數會被壓縮，限制區或會出現奇異點(singular point)，當極值發生在奇異點位置，某些 β 中的元素將被壓縮到零，所以 LASSO 在估計的同時，也經由壓縮係數而做了變數選取。因此改善了脊迴歸無去做變數選取，並建立了一個不同於離散子集挑選的連續型變數選取過程。以 2 維為例，如圖 2 所示，左邊為 LASSO 示意圖，其限制區域為一個旋轉後的正方形(灰色區域)；右邊為脊迴歸示意圖，其限制區域為一個圓形(灰色區域)。若 $\hat{\beta}^{LASSO}$ 發生於奇異點(如圖所示)，即挑選到了 x_2 變數，而 $\hat{\beta}^{Ridge}$ 無去做變數

挑選，但是兩者皆將原本的 $\hat{\beta}^{OLS}$ 壓縮至限制式的範圍。由於 LASSO 估計的計算為數學上的 QP(Quadratic Programming)問題，可以利用數值分析演算法來求算。

三、資料與分析結果

為了解 LASSO 方法在氣象綜合預報的可應用性，本研究以台北站 12 月及 6 月的逐日氣溫為預報目標，選定東亞區域共 7 個重要層場(MSLP、H700、H500、T850、V850、U200、RH850)的大氣環流格點資料做為候選預報因子，分別使用複迴歸法、傳統變數選取、逐步迴歸法(stepwise)及壓縮係數變數選取法-LASSO 方法建立迴歸預報方程，以 RMSE、絕對誤差度數天數、模式計算效率去比較三個模式十年交叉驗證(cross-validation)預報差異。

首先將台北站每日均溫(Y；反應變數)與 7 個層場的環流格點資料進行相關分析(圖 3)，根據結果，選取具有相當相關程度的格點資料列入待選預報因子(X；解釋變數)。而為了比較測試，另外又依不同顯著相關標準設定多組資料集。以測試 12 月均溫所使用的資料為例，共建立 6 組資料集，分別為：Multi-zero、Multi-I、Multi-II、Multi-III、Multi-EA2 及 Multi-EA。其中 Multi-zero 為各層場則站上空單一格點，因此解釋變數個數僅有 7 個，其他資料集所對應的相關程度限制收(表 1)，表

中可見隨著相關程度限制放寬，解釋變數的個數也將隨之增加。在其他資料處理方面，每月的取樣除了當月以外，還加上前後十五日，以預報台北站 12 月逐日均溫為例，本研究使用的資料為自 1998 年至 2007 年共十年，每年 11 月 15 日至隔年 1 月 15 日的資料。在不同迴歸方法的比較方面，將以複迴歸法為評比的基準模式，此法是以各資料集中各層場所有選取格點平均值作為解釋變數，再對此 7 個解釋變數建立複迴歸，其中並不再做變數挑選。我們期望經由變數挑選方法所建立的迴歸模式可比此基準模式佳。以下分別說明 12 月及 6 月份資料的測試結果。

12 月逐日均溫預報結果，Multi-zero 資料集由於解釋變數僅有 7 個，RMSE 在 1 度至 1.8 度之間，三種方法表現一致；Multi-I 資料集開始，stepwise 與 LASSO 逐漸拉開與複迴歸法的差異，RMSE 降到 0.8 至 1.4 度之間；Multi-II 資料集，LASSO 與 stepwise 差異變大，LASSO 的 RMSE 始終低於複迴歸法及 stepwise，stepwise 的 RMSE 在 2004 年與 2005 年還出現比複迴歸法還高的狀況；Multi-III 資料集中，stepwise 的 RMSE 每年皆高於複迴歸法，而 LASSO 的 RMSE 每年依舊低於複迴歸法；逐漸增加解釋變數個數，Multi-EA2 資料集增加到 534 個接近 9 年樣本個數 540 個，此時 stepwise 其 RMSE 上升到 4.5 度至 7.5 度間，遠高於 LASSO 與複迴歸法的 0.8 度至 1.5 度；當解釋變數個數增加而超過樣本個數，如 Multi-EA 資料集中，stepwise 則因演算法限制而無去求算，LASSO 的 RMSE 保持在 0.8 度至 1.4 度間。觀察 LASSO 法對各資料集建立台北站 12 月均溫預報模式，經十年交叉驗證所得 RMSE 值(見圖 4)，可發現無論資料集中的解釋變數如何變化，每年預報的 RMSE 變異不大，反映出連續型的變數選取方法 LASSO 不因資料的少許變化而改變預報結果；若改以 stepwise 測試(見圖 5)，可以明顯看到 stepwise 隨著變數個數增加，RMSE 逐漸增大，預報結果易受到資料集改變而變化。進一步以絕對誤差度數為 0 之天數為標準，比較不同迴歸模式在各資料集預報 2007 年 12 月的預報成效(見圖 6)，可發現 stepwise 從 Multi-II 資料集開始，絕對誤差度數為 0 之天數持續下降至 Multi-EA2 資料集僅剩下 2 天，而 LASSO 在 Multi-II 與 Multi-III 資料集中都有 20 天以上，且在 Multi-EA2 與 Multi-EA 資料集亦還有 13 天以上，並可發現其天數皆大於複迴歸法基準模式。

6 月逐日均溫預報結果，亦可發現隨著資料集樣本數增加，LASSO 的 RMSE 趨於穩定，各資料集交叉驗證的 RMSE 均可低於 1.5 度，而 stepwise 從 Multi-II 資料集開始 RMSE 開始上升，直到 Multi-III 資料集 RMSE 均高於 1.5 度，甚至有超過 2 度的狀況發生。從六月份的 10 年交叉驗證表現(見圖 7、

圖 8)，可見與 12 月具有相同狀況，stepwise 容易受到資料集變化而改變預報結果。由於夏季預報準確率較低且各資料集 2007 年 6 月均溫絕對誤差度數為 0 度之天數差異不大，進一步去比較不同迴歸模式在各資料集預報 2007 年 6 月均溫絕對誤差度數小於或等於 1 度之天數(見圖 9)，可發現 LASSO 略高於 stepwise。關於於模式計算效率，我們去比較預報 12 月各資料集執行十年交叉驗證程式所需時間，由(表 2)可明顯看到 LASSO 在處理 Multi-I、Multi-II、Multi-III 資料集都可以迅速完成，僅需要約 3 分鐘以內，而 stepwise 在 Multi-III(解釋變數個數為 323 個)就要耗費約 870 分鐘，甚至遠高於 LASSO 完成 Multi-EA 資料集(解釋變數個數為 589 個)僅需要約 53 分鐘，stepwise 顯得相當耗時且無效率。

四、結論

LASSO 可同時做到壓縮係數及變數選取，壓縮係數功能發生些許不偏性而降低迴歸係數估計值的變異，穩定了變數間共線性對模式的影响，並提高了模式的預報準確度；變數選取功能，解決了建立迴歸模式時面臨大量解釋變數的問題，而且 LASSO 為連續型的變數選取方式，可以提供更穩定的預報方程，不易受資料變動而影響結果。另外，在發展程式階段，我們發現引用 LARS(Efron, Hastie, Johnstone, and Tibshirani, 2004)演算法還能更進一步改良 LASSO 的演算，使其更有效率。因此，為了因應天氣迴歸預報模式建立的挑戰，具有準確性、穩定性、效率性的 LASSO 是符合我們所期待的工具。

五、參考文獻

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression (with discussion). *The Annals of Statistics* 32 407 - 451.

Hocking, R. R. (1976) The analysis and selection of variables in linear regression. *Biometrics*, 32, 1-49

Hoerl, A. E. and Kennard, R. W. (1970) Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, 12,55-67

James, W. & C. Stein (1961), "Estimation with quadratic loss", Proc. Fourth Berkeley Symp. Math. Statist. Prob., vol. 1, at 311-319

Richards, John A. (1999) An introduction James-Stein estimation
url : <http://sag.mit.edu/group/alumni/johnrich/docs/js.ps.gz>

Stein, C. (1956), "Inadmissibility of the usual estimator for the mean of a multivariate distribution", Proc. Third Berkeley Symp. Math. Statist. Prob., vol. 1, at 197-206

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B*, 58, 267 - 288.

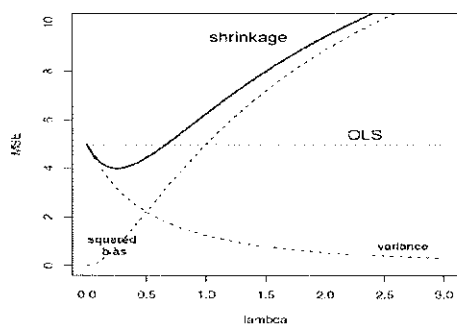


圖 1、壓縮概念示意圖

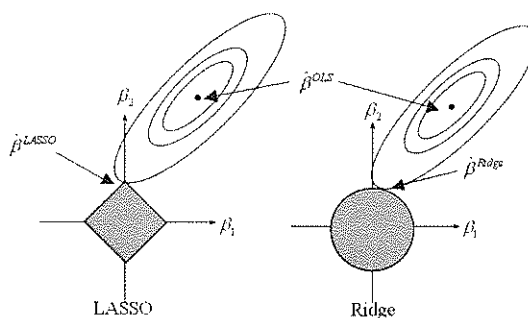


圖 2、LASSO 與 Ridge 方法 2 維示意圖

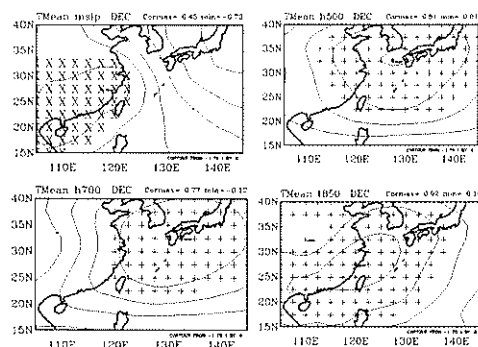


圖 3、資料預報因子選取示意圖

表 1、12 月資料集

12 月	Multi-rec		Multi-I		Multi-II		Multi-III	
	Cor	Pa Number	Cor	Pa Number	Cor	Pa Number	Cor	Pa Number
M6p		1	Cor=0.7	6	Cor=0.6	31	Cor=0.5	55
H700		1	Cor=0.7	21	Cor=0.6	53	Cor=0.5	89
H500		1	Cor=0.7	30	Cor=0.7	30	Cor=0.6	63
T850		1	Cor=0.5	26	Cor=0.7	51	Cor=0.6	87
V850		1	Cor=0.6	1	Cor=0.5	17	Cor=0.5	17
U200		1	Cor=0.5	9	Cor=0.5	9	Cor=0.5	9
RH850		1	Cor=0.5	3	Cor=0.5	3	Cor=0.5	3
Total		7		99		194		323
			Multi-EA2		Multi-EA			
	Cor	Pa Number	Cor	Pa Number	Cor	Pa Number		
M6p	Cor=0.4	75	Cor=0.4	88				
H700	Cor=0.4	95	Cor=0.4	146				
H500	Cor=0.3	111	Cor=0.4	122				
T850	Cor=0.4	147	Cor=0.4	165				
V850	Cor=0.4	11	Cor=0.4	25				
U200	Cor=0.4	41	Cor=0.4	41				
RH850	Cor=0.4	18	Cor=0.4	18				
Total		534		589				

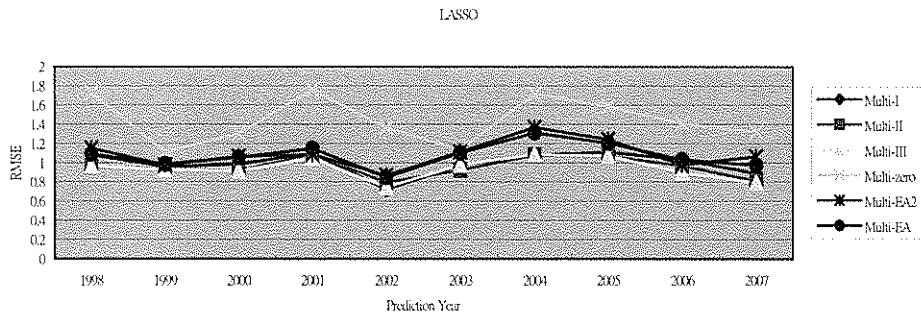


圖4、以LASSO法對各資料集建立台北站12月均溫預報模式，經十年交叉驗證所得RMSE值

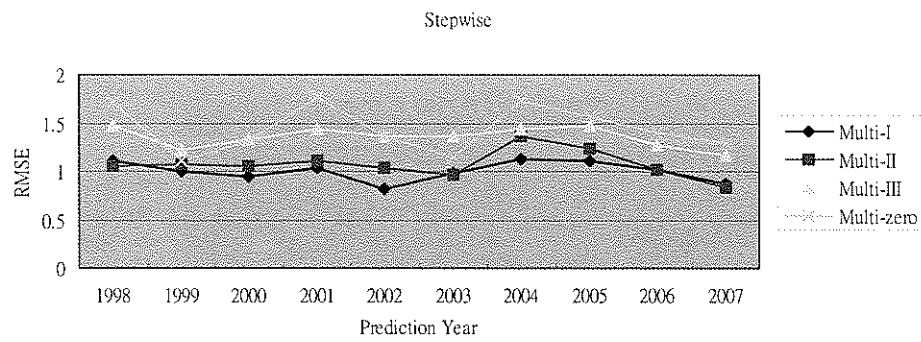


圖5、同圖4，改用逐步迴歸法，省略Multi-EA2、Multi-EA資料集

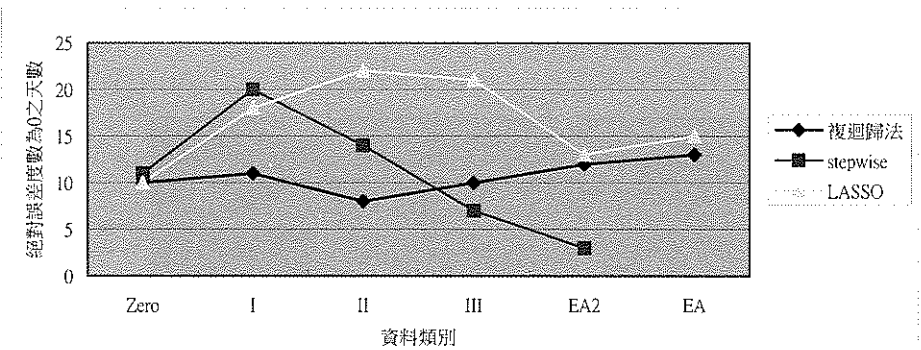


圖6、不同迴歸模式在各資料集預報2007年12月均溫絕對誤差度數為0度之天數

表2、預報12月各資料集執行十年交叉驗證程式所需時間

	Multi-I	Multi-II	Multi-III	Multi-EA2	Multi-EA
解釋變數個數(個)	99	194	323	534	589
LASSO(分鐘)	0.3	0.9	3.6	50.5	52.3
Stepwise(分鐘)	12.0	235.2	869.2	1944.6	none

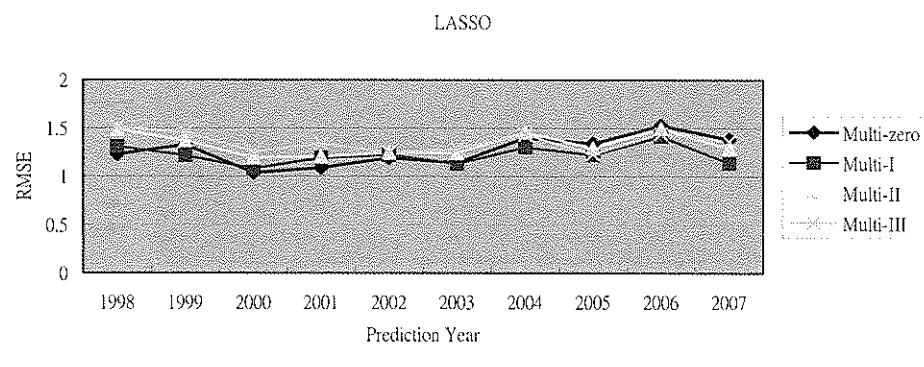


圖7、以LASSO法對各資料集建立台北站6月均溫預報模式，經十年交叉驗證所得RMSE值

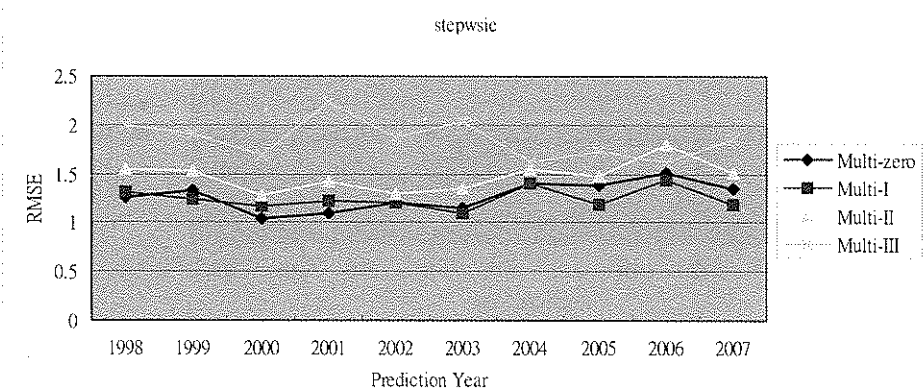


圖8、同圖7，改用逐步迴歸法

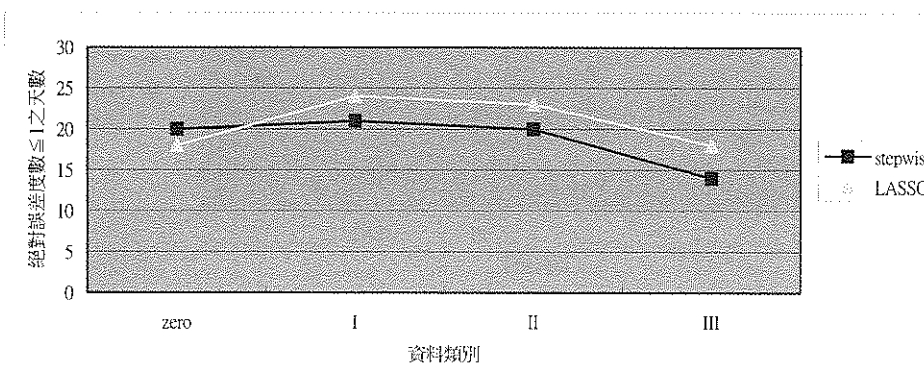


圖9、不同迴歸模式在各資料集預報2007年6月均溫絕對誤差度數≤1度之天數