尋找時間數列最佳預測模型

中央研究院統計科學研究所　銀慶剛

一、時間數列選模的簡介

舉凡股票指數，大眾運輸工具的客運量石油價格，失業率，降雨量及太陽黑子個數等，這些隨著時間呈現高低起伏變化的資料，不管是在自然科學或社會科學的研究裡都處處可見，我們稱這些資料爲時間數列。預測是收集(或觀測)時間數列資料最重要的目的之一，精確的預測常能幫助我們做出正確的決策。預測進行的方式如下：首先以一個數學模型來描述資料間的前後關聯，其次利用觀測到的資料，估計模型中的未知參數，最後，將所估計的參數帶入模型產生預測值。然而，如果有一個以上的候選模型可供選擇時，如何挑選出最佳的預測模型呢？這個問題十分有趣而重要，1970 年以來也引起學界廣泛的討論。

在諸多選擇模型的統計方法中，日本統計學家 Akaike[1] 提出的赤池訊息準則(Akaike's Information Criterion, AIC) 是最廣受重視學界的一個，這可從它被引用的次數一窺端倪。根據美國科學資訊研究所(Institute for Scientific Information, ISI) 的統計，上述文章迄今已被引用了 4546 次，衡諸近年來其他幾篇重要的統計(或相關領域) 文獻被引用的次數，如 Efron [2] (Bootstrap 法，2283 次)，Dempster et al. [3](EM 演算法，5783 次)，及 Geman and Geman [4] (Gibbs 抽樣法，3319 次) 等，[1]文顯然不遑多讓。事實上，AIC 是在測量眞實模型與候選模型間 Kullback-Leibler 距離(K-L distance)時，導入預測的概念而產生。令 $\mathbf{y}_n=(y_1,\ldots,y_n)$與 $\mathbf{x}_n=(x_1,\ldots,x_n)$爲同分佈(identically distributed)的資料，且 $\mathbf{y}_n$ 與 $\mathbf{x}_n$ 獨立。此地，$\mathbf{x}_n$ 是用來表示觀測到的資料而 $\mathbf{y}_n$ 則表示尚未得見的"未來"資料。此外，令 $f(\mathbf{y}_n)$ 表 $\mathbf{y}_n$ 眞正的機率密度函數(probability density function, pdf)，$g(\mathbf{y}_n, \theta)$表候選模型(或候選 pdf)，用以近似 $f(\cdot)$，其中 θ 爲未知參數。因爲 θ 未知，我們可利用 $\mathbf{x}_n$ 得到θ的最大概似估計 $\hat{\theta}_n=\hat{\theta}_n(\mathbf{x}_n)$，並進一步得到 $g(\mathbf{y}_n,\hat{\theta}_n(\mathbf{x}_n))$。當我們用觀察到的資料 $\mathbf{x}_n$ 估計參數後，應用到未來資料 $\mathbf{y}_n$ 的模型(或 pdf)時，"預測"這個概念就被悄然引進了。Akaike 發現，$f(\mathbf{y}_n)$與 $g(\mathbf{y}_n,\hat{\theta}_n(\mathbf{x}_n))$ 的 K-L distance，

$$E\left(\log\frac{f(\mathbf{y}_n)}{g(\mathbf{y}_n,\hat{\theta}_n(\mathbf{x}_n))}\right), \tag{1.1}$$

可適當地由，

$$\text{Constant} - \log g(\mathbf{x}_n,\ \hat{\theta}_n(\mathbf{x}_n)) + K, \tag{1.2}$$

來估計(注意到(1.1)式中的 $g(\mathbf{y}_n,\hat{\theta}_n(\mathbf{x}_n))$ 已被替換爲 $g(\mathbf{x}_n,\hat{\theta}_n(\mathbf{x}_n))$，其中 Constant 是一個與候選模型 $g(\cdot,\cdot)$ 無關的常數而 K 是模型中參數的個數。他因此定義

$$\text{AIC} = -2\log g(\mathbf{x}_n,\ \hat{\theta}_n(\mathbf{x}_n)) + 2K. \tag{1.3}$$

以上的討論顯示，AIC 強調的不僅只是候選模型 $g(\cdot,\cdot)$ 與眞實模形 $f(\cdot)$之間的差異，它更透過區別現有資料 $\mathbf{x}_n$ 與未來資料 $\mathbf{y}_n$，強調了模型的"預測"能力。由(1.3)式得知，AIC 以候選模型擬合度(goodness of fit)與複雜度(complexity)之總和做爲預測能力的指標。擬合度以 $-2\log g(\mathbf{x}_n,\hat{\theta}_n(\mathbf{x}_n))$ 衡量，此值越小表候選模型與觀測值之間的擬合程度越好，因此有利於預測；複雜度，2K，則與模型中未知參數的個數成正比，未知參數越多意味著模型要負擔較多的估計誤差，因此預測品質較差。因爲高擬合度與低複雜度常常不能得兼，選取 AIC 值最小的候選模型，可幫助我們在此兩指標上獲得一個好的平衡，進而達到好的預測效果。但，這是否保證 AIC 眞能選到時間數列最佳的預測模型？過去二十餘年，這個問題雖屢屢被提及，卻始終未能有令人滿意的答案。在第二、三節裡，我們將討論進一步的細節。

二、早期理論

1980 年以前，統計學者在選模上的研究大多偏重在證(說)明 AIC 與交互驗證法(Cross Validation，CV)或 Mallow's Cp 的(漸進)等價性，以及證明貝氏訊息準則(Bayes Information Criterion, BIC)在選取時間數列模型及線性迴歸模型時的一致性(指選到眞實模型的機率，隨樣本數趨近於無窮大而趨近於 1)。然而，由於討論一致性需假設眞實模型包含在候選模型中，爲了避免這一個特殊的假設，1980 年以後考慮眞正模型不必然包含在候選模型中的研究漸漸受到重視。Shibata[5]開啓了這個方向研究的先河：他主要的想法是用複雜(高維度)的模型逼近眞模(極複雜需用到無窮多個參數)，並容許逼近模型的複雜度隨著資料量增加而增加以降低逼近誤差。在此架構下，他證明了 AIC 及 Cp 最終可以選到預測能力最好的逼近模型，但 BIC 傾向選擇過分簡潔的模型而不具漸進有效性，這個結果同時也對無母數迴歸中帶寬(bandwidth)選取的研究產生深遠的影響。雖然 Shibata 的文章很快地就被學界認可，但對時間數列的分析者而言，文章中卻隱含一個無法忽略的瑕疵，那就是，爲了配合 Akaike[1] 的預測概念，Shibata 是對觀測到的時間數列$\{xn\}$進行估計與選模，但將選到的模型用來預測另一組 i.i.d 時間數$\{\mathbf{y}_n\}$的下一步，$\mathbf{y}_n$+1。不幸的是，這樣的作法似乎從未被實際採用，幾乎所有時間數列選模的目的都是爲了預測觀察值本身的未來，$\mathbf{x}_n$+1。但是，當我們將$\{\mathbf{x}_n\}$與$\{\mathbf{y}_n\}$獨立這個條件改成 $\mathbf{x}_n = \mathbf{y}_n$ 時，立刻面臨困難的數學問題，使得獲選模型的預測誤差變得難以計算，這也解釋了，爲何預測 $\mathbf{x}_n$+1 的最佳選模理論，始終未能建立的原因。

另外一方面，奠基在解碼理論的研究上，Rissanen[6]年提出了另一類廣被討論的選模原則,稱爲累積預測誤差(Accumulated Prediction Error, APE)，它的特色是完全根據模型的線上(on-line)預測表現來選模。相關文獻自 1988 至 1995 年間大量湧現，證明 APE 的一致性與它與 BIC 間的漸近等價性，其中 Wei[7]可做爲這個方向研究的代表。然而作爲一個以預測爲目的所設計出來的選模方法，APE 應有一些超越一致性的特質，這些特質卻始終未能被挖掘出來。

三、我的研究方向

茲擇要敘述本人對時間數列選模的研究如后：

1. 證明 AIC 在單一樣本預測時的漸近有效性

單一樣本預測(same-realization prediction)指的是對觀察值本身$\{\mathbf{x}_n\}$，的未來，$\mathbf{x}_{n+h}, h \geq 1$，作預測。爲了方便討論，我將 h 固定爲 1，並假設$\{\mathbf{x}_n\}$爲一無窮階(infinite-order)的自迴歸過程(AR (∞) process)。此時選模的目的，是根據$\{\mathbf{x}_n\}$來選定一個有限階(finite-order)的自迴過程來逼近眞實模型，希望達到理想的單一樣本預測。如果 $\hat{k}_n$ 是被某一選模準則(model selection criterion)所選出的階數，並滿足：

$$\lim_{n\to\infty} \frac{E\left(x_{n+1} - \hat{x}_{n+1}(\hat{k}_n)\right)^2 - \sigma^2}{\min_{1\le k\le K_n} E\left(x_{n+1} - \hat{x}_{n+1}(k)\right)^2 - \sigma^2} = 1, \quad (3.1)$$

則我們稱此選模準則具漸進有效性(asymptotic efficiency, AE)，其中 $\hat{x}_{n+1}(k)$ 是根據 AR(k)模型得到的最小平方預測，σ^2 爲白噪音(white noise)的變異數，K_n 表最大候選模型的階數並可隨樣本數 n 趨近無窮大。(3.1)說明了，AE 選模準則的(單一樣本)預測均方差與最佳預測模型的預測均方差漸進等價。因爲實際上無法知悉最佳預測模型爲何，我們可以 AE 選模準則選到的模型代替最佳模型，達到漸進最佳預測。但是，由於時間數列資料的前後關聯，造成了 $\hat{k}_n$ 與 x_{n+1} 間極爲複雜的相關結構，使得 $E(x_{n+1} - \hat{x}_{n+1}(\hat{k}_n)^2$ 的計算變得窒礙難行。爲了解決這個困難，我們在[8]中發展了一些十分有用的數學工具，藉由這些工具，我們證明了：AIC 能以很高的機率選到與最佳預測模型接近的模型。藉此，我們更進一步證明：AIC 滿足(3.1)式。

2. 證明修正後的 APE 在多步預測時的漸近有效性

在預測各式各樣科學數據及社會或經濟指標長期趨勢的需求與日俱增的情況下，時間數列的多期預測近年來已成爲熱門的研究主題。由於一個模型之下常有數個不同的多期預測方法，有別於傳統的選模問題，多期預測最大的挑戰在於如何找到模型與預測方法間的最佳組合。在[9]中，我假設$\{\mathbf{x}_n\}$產生自一有限階的自迴歸過程，

並假設眞實模型含括在候選模型的單純狀況成立。我舉例說明了，當一種以上的多期預測方法被納入考慮時，即便在上述單純的狀況下，最佳多期預測都無法透過一致性的選模產生。這個結果，挑戰了所有的一致性選模準則(包含 BIC 及傳統的 APE)，應用於多期預測時的正當性。爲了克服這個難題，我在[9]中修正了 APE，使它成爲一個同時能選模型與選預測方法的工具。[9]中也發展了一些新的數學技巧，它們幫助我在 martingale structure 被多期預測破壞的困難情況下，證明了(於適當條件下)修正後的 APE 選中模型與方法最佳組合的機率，會隨著樣本數增加而趨近於 1。

3. 時間數列選模的統一定理

時間數列選模裡另一個主要的難題是：當眞實模型極爲複雜時(如 AR(∞)模型)，AIC 是 AE 但 BIC 不是；而當眞實模型較簡單時(如有限階的 AR 模型)，BIC 具有一致性但 AIC 不具有。不幸的是，由於眞實模型是否複雜(或簡單)，永遠無法得知，故 AIC 與 BIC 的優劣之辯多年來始終未能偃旗息鼓。爲了解決這個問題，我在[10]中將 AIC 與 BIC 適當組合後，發展了一個兩階段的選模法。我並證明了這個兩階段的選模法，在眞模複雜的情況下與 AIC 漸進等價，而在眞模簡單的情況下與 BIC 漸進等價。

參考文獻

[1] H. Akaike, IEEE Transactions on Automatic Control, **19**, 716 (1974).

[2] B. Efron, Annals of Statistics, **8**, 1 (1979).

[3] A. P. Dempster, N. M. Laird and D. B. Rubin, Journal of the Royal Statistical Society, Series B, **39**, 1 (1977).

[4] S. Geman and D. Geman, IEEE Transactions on Pattern Analysis and Machine Intelligence, **6**, 721 (1984).

[5] R. Shibata, Annals of Statistics, **8**, 147 (1980).

[6] J. Rissanen, Essays in Time Series and Allied Processes (J. Gani and M. P. Priestley, eds.), Journal of Applied Probability, **23A**, 55 (1986).

[7] C. Z. Wei, Annals of Statistics, **20**, 1 (1992).

[8] C. K. Ing and C. Z. Wei, Annals of Statistics, **33**, 2423 (2005).

[9] C. K. Ing, Annals of Statistics, **32**, 693 (2004).

[10] C. K. Ing, Manuscript (2006).