

國立清華大學

碩士論文

Variable selection in Cox proportional
hazards model for prevalent survival data

在 Cox 比例風險模式之下對盛行倖存

資料進行模式選擇



所別：統計研究所 組別：生物統計組

指導教授：鄭又仁 Cheng, Yu-Jen 博士

學生：楊聲濤 Yang, Sheng-Tao

中華民國 104 年 6 月 29 日

Abstract

We study the variable selection problem in Cox proportional hazards model for prevalent survival data. In this study, we face some challenges. Firstly, from many potential predictors, we would like to select a small number of key risk factors, including continuous or discrete variables. Secondly, data were collected from a prevalent sampling which is exactly a biased sampling scheme. The proposed method not only can select and estimate variables simultaneously but also can correct the sampling bias. Further, the proposed method can allow for different penalty functions, including continuous or discrete variables. The results of simulation study show that the proposed procedure is stable and more accurate to select the true model. We also apply the proposed method to a real data.



Abstract

我們研究了在 Cox 比例風險模式 (Cox proportional hazards model) 分析盛行倖存資料 (Prevalent survival data) 時模式選擇議題，然而在此研究中我們遇到了一些挑戰。第一，在盛行倖存資料中我們蒐集到許多項共變數 (Covariates)，研究的目標是希望從這些共變數挑僅選出些許重要的共變數即可，包含連續型與離散型的共變數；第二，盛行倖存資料事實上是有偏誤之抽樣方法 (Biased sampling)。我們提供了一套方法，不僅可以同時進行模式選擇與參數估計，還可以矯正因抽樣方法造成的偏誤。更進一步地，我們的方法可以允許使用不同的懲戒函數 (Penalty function)，適用於連續型與離散型的共變數。模擬結果顯示我們提供的機制是穩定的，並且可以選出正確的模式。我最後，們也將此方法應用在一筆有關女性乳癌的真實資料上進行分析。



Acknowledgement

真的非常感謝鄭又仁老師兩年以來的教導與清華大學統計研究所兩年以來的訓練，從碩一下開始跟隨鄭又仁老師閱讀與倖存資料分析和模式選擇有關的文章，途中遇到很多讀不懂文章內容的挫折，鄭老師會耐心引導學生理解文章內容，老師不拘泥於數學式的推演與數學細節，而是更重視數學式的精神架構與直觀解釋意義，因為了解這些才能對於以後的議題推廣延伸，且閱讀資料的時候能對議題更有感覺，這樣的觀念對我來說是一個相當大衝擊，畢竟大學念數學系，數學系又特別從重視定理和證明，不過我是樂於接受這樣的觀念的，因為把數學與實際議題結合正是我轉念統計研究所的原因。這兩年真的收穫很多，也體會到作研究的樂趣，有老師的指導才能完成這篇論文，非常感謝鄭又仁老師的教導！除此之外老師也很關心學生們畢業後的狀況，給予鼓勵和協助，很令人感動。

謝謝我的家人，一路上都非常尊重且支持我念書和參加音樂活動，讓我從小就不用顧慮經濟方面的問題，可以專心的念書和利用剩餘的時間參加學校社團，雖然小時候的混的不好讓你們失望，還好長大後比較懂事會替自己的未來打算，也找到自己的興趣所在，未來我會努力爭取出國深造的機會，真的非常謝謝家人的栽培和鼓勵，給我這麼好的機會來完成學業。

在新竹整整六年，從懵懂的大一到碩士班畢業，雖然途中跌跌撞撞出了很多包吃了許多虧，但也遇到很多貴人和好朋友互相鼓勵和批評指教，特別感謝交通大學應用數學系、雄友會、星聲社、地板組、幼幼社、清華大學統計研究所、迴聲社、大仁家除的學長姐學弟妹和同學們，從你們身上學習到的東西，不管是好的壞的，都令我成長許多，所以我今天才會在這裡，和你們一起念書、一起玩音樂、一起練習活動、一起運動打電動、一起吃飯聊天都是我學生時代最珍貴的回憶，我將要離開新竹了，有緣會在相聚的！

Contents

1	緒論	1
1.1	倖存資料	1
1.2	模式挑選	3
2	文獻回顧	5
2.1	符號定義與資料型態	5
2.2	部分概似函數	5
2.3	模式挑選與懲戒部分概似函數	6
2.4	群懲戒函數	10
2.5	正規解路徑	11
3	研究方法	14
3.1	群懲戒 log-部分概似函數	14
3.2	演算法	16
3.3	變異數估計	21
3.4	調整參數之選取	21
4	數值模擬	23
5	實例分析	26
5.1	女性乳癌資料	26
6	結論	30
7	附錄	31
7.1	附錄一	31
7.2	附錄二	40
7.3	附錄三	47

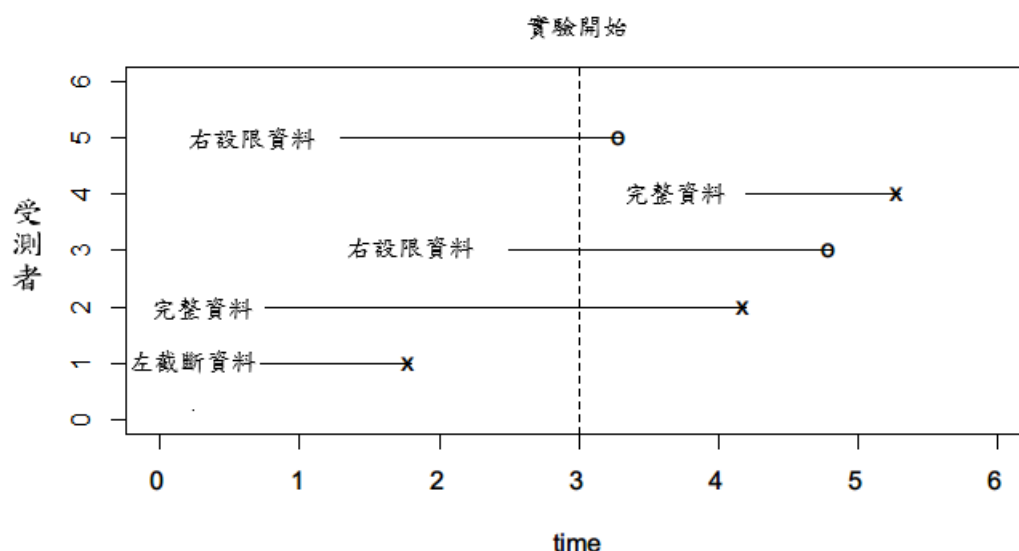
1 緒論

1.1 倖存資料

在倖存分析中，研究人員有興趣的議題是某個事件發生後的存活時間 (Survival time)，例如某種疾病從潛伏期至發病的時間、病人得病後存活的時間、電子產品的壽命、金融產品的壽命等眾多議題，繁族不及備載；然而在倖存分析中，常常無法完整地觀測事件的存活時間，因此造成分析資料的困難，例如資料右設限 (Right censored) 的問題 (Kalbfleisch and Prentice, 2002)。

資料會發生右設限的原因是因為受測者還未經歷死亡事件 (Failure event) 即離開實驗，使得研究人員無法關查到受測者的存活時間，其中受測者發生事件到離開實驗之間的時間我們稱為右設限時間，此類型資料在研究疾病時時常發生，例如：受測者的身體狀況太差需要接受治療而離開實驗、受測者還未經歷死亡事件但實驗已經中止，等等諸多因素，使得研究人員只能觀測到事件的右設限時間但不能觀測到存活時間，只能得知事件的存活時間至少大於右設限時間，若直接使用右設限時間代替存活時間進行參數估計，將會造成低估存活函數 (Survival function)。

對於右設限資料的處理，Kaplan and Meier (1958) 提出可以藉由估計風險函數 (Hazard function) 來估計修正存活函數的估計，此方法之後又繼續被推廣為乘積極限估計量 (Product limit estimator)，此為沒有加入模式解釋的無母數估計方法。定義有興趣事件之存活時間為隨機變數 T ，對應死亡事件時間點 t 的倖存函數記作 $S(t) = P(T > t)$ ，其意義為有興趣事件的存活時間超過 t 的機率。另外定義風險函數 $\lambda(t) = \lim_{h \rightarrow 0} \frac{P(t < T < t+h | t < T)}{h}$ ，其意義為事件在給定事件的存活過時間 t 後瞬間死亡的機率 (Kalbfleisch and Prentice, 2002)。而後 Cox (1972) 在分析存活函數時加入了迴歸模式的解釋，對風險函數作了一個半母數



形式的比例風險的假設，並建立部分概似函數 (Partial likelihood function)(Cox, 1975)，藉由最大化部分概似函數得到參數的漸近不偏的估計量。

本文主要探討如何分析盛行倖存資料 (Wang, 1989; Wang et al., 1993)，為了節省金錢與時間成本，此抽樣方式已經常用於流行病學、癌症研究等需長時間追蹤的研究之中；傳統的世代研究 (Cohort study) 蒐集資料的方式為蒐集實驗開始之後發生疾病的受測者，並持續追蹤至觀察到受測者的存活時間或者右設限時間；然而存活時間較長的疾病，研究時間需要夠長才能觀察到足夠的死亡事件，使得世代研究在研究存活時間較長的疾病蒐集資料上較不具效率，雖然此種實驗方法可以同時探討多種疾病，卻相當花費時間且不適用於稀有疾病；相較之下，盛行倖存資料針對某個特定的疾病，研究其存活時間，在實驗開始時蒐集當下已經發生該疾病並且還未發生死亡事件的病患的資料，如此可以比世代研究蒐集到更多資料；然而盛行倖存資料實際上是偏誤樣本而非隨機樣本，除了考慮資料右設限的問題，需要考慮資料左截斷 (Left truncation data) 的問題 (Tsai et al., 1987)。

左截斷資料發生原因是因為存活時間越短的受測者越不容易進入實驗，資料忽略了存活時間較短的樣本而傾向於蒐集存活時間較長的樣本，導致蒐集到的

樣本為偏誤樣本 (Biased sample)，無法忠實呈現疾病母體的訊息；定義左截斷時間為病患發病之後到實驗開始之間的時間，假設每個病患皆有存活時間與左截斷時間，若某病患的左截斷時間小於存活時間，則該病患的資料無法進入研究人員的數據之中，若研究人員未考慮此偏誤，將得到偏誤的訊息分析結果。Tsai et al. (1987) 同時考慮左截斷與右設限資料的問題，提出了無母數的乘積極限估計量用來估計存活時間，此為沒有加入模式解釋的無母數估計方法；Wang et al. (1993) 考慮左截斷資料加入模式解釋的問題，在 Cox 比例風險模式之下推論出漸近不偏估計量。

1.2 模式挑選

模式挑選在統計學科中一向是重要的課題，以分析數據的經驗而言，重要的變數其實並不多，而不重要的變數不需要放入模式之中，以免過度解釋模式造成參數估計效率較差，這顯示挑選出重要的變數是重要的。令一方面，由於科技的進步，人類已經進入大數據時代，不但蒐集資料相對以往來說非常簡單且快速，更可以記錄更多樣本與變數，例如關於基因議題的資料可能動輒幾萬個基因位點，又或者生醫影像的像素 (pixel) 皆數已經數以百萬計，然而對於有興趣事件重要之變數卻僅有數個，這也凸顯了辨別重要變數的議題的重要性。

在迴歸模式中，傳統的模式選擇選方法如所有子模式選擇 (All subset selection)，其方法為跑遍所有模式組合並估計參數，訂定某個挑選標準，選擇符合該挑選標準的模式作為模式選擇與參數估計的結果；然而此方法非常不具效率；假設資料有 p 個變數，則模式的組合數為 $2^p - 1$ 呈現指數成長，如此一來變數個數增會使得模式組合數急劇增加，跑遍所有模式組合相當不具效率性。

另外的方法如逐步迴歸 (stepwise regression)，其方法為配飾模式之後，依據參數的顯著程度決定是否將與該參數對應之變數從迴歸模式中剔除，或者多加

入某個未被選入模式之中的變數，依據與該變數對應之參數的顯著程度決定是否將該變數加入迴歸模式之中，反覆執行直至沒有變數需要被加入或者剔除。然而此方法已經由 Breiman (1996) 討論為不穩定的模式選擇方法，也就是即使兩筆資料皆從同一母體抽樣，藉由逐步迴歸的方式也很可能得到不同的模式選擇解果，也因此使得參數估計的性質難以推論。另外分開執行模式選擇與參數估計的過程，也是此方法的缺點。

自從 Tibshirani (1996) 提出了 Least Absolute Shrinkage and Selection Operator (LASSO) 機制，得到具有稀疏性的估計量，得以同時進行模式選擇與參數估計，接續著有一系列關於模式選擇議題的方法被提出與討論，此部分將會在第二章文獻回顧中更加詳細地介紹，包含如何挑選調整參數 (Tuning parameter) (van Houwelingen et al., 2006)、該使用何種懲戒函數以描繪正規解路徑 (Regularized solution path) (Fan and Li, 2002; Zou, 2006; Zhang, 2010)、如何有效率的描繪正規解路徑 (Efron et al., 2004; Friedman et al., 2010; Simon et al., 2011; Breheny and Huang, 2015)、模式選擇與參數估計的性質 (Fan and Li, 2002)、離散型隨機變數的處理方式 (Yuan and Lin, 2006)、如何推廣至廣義線性模式甚至 Cox 比例風險模式 (Tibshirani et al., 1997; Zou, 2008) 等等。第三章研究方法的部分，我們針對左截斷與右設限型態的盛行倖存資料，在 Cox 比例風險模式之下的模式選擇與參數估計之議題進行討論，並提出一套機制得以穩定且有效率同時進行模式選擇與參數估計。第四章呈現有無考慮左截斷時間的模式選擇與參數估計之數值模擬結果，並且比較有無考慮左截斷時間的模擬結果之差異。最後我們導入一筆真實盛行倖存資料進行統計分析。

2 文獻回顧

2.1 符號定義與資料型態

令 T^* , X^* , W^* , $W^* + C^*$ 分別表示為研究人員有興趣事件之存活時間、共變數、左截斷時間與右設限時間，承緒論所述，研究人員所蒐集到的資料並非完整資料。若存活時間小於截斷時間，研究人員將無法蒐集到資料，即研究人員觀察到的資料僅有集合 $S = \{i | T_i^* \geq W_i^*\}$ ；其次，受測對象的存活時間不總是觀察得到，因為受測者可以自由選擇何時離開實驗，又或者受測者還處於存活狀態但實驗已經結束等諸多因素，因此研究人員只能觀測到受測者的存活時間或者右設限時間其中之一；定義 $Y^* = \min(T^*, W^* + C^*)$ ， $\Delta^* = I(T^* \leq W^* + C^*)$ 代表觀測到之時間與其狀態， $\Delta^* = 1$ 或 0 分別代表 $Y^* = T^*$ 或 $Y^* = W^* + C^*$ 。綜合以上兩種造成資料不完整的因素，研究人員只能中觀察到不完整資料 $(Y, X, W, \Delta) = \{(y_i, \mathbf{x}_i, w_i, \delta_i) | i \in S\}$

2.2 部分概似函數

令資料 $(y_1, \mathbf{x}_1, w_1, \delta_1), \dots, (y_n, \mathbf{x}_n, w_n, \delta_n)$ 為盛行倖存資料之隨機樣本， y_i 為觀察到之時間、 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ 為受測者的 p 個共變數、 w_i 為左截斷時間、 δ_i 為觀察到之時間的狀態， 1 代表存活時間， 0 代表右設限時間。Cox 的比例風險模式對事件存活時間的風險函數做了一個半母數型式之假設 (Cox, 1972)

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t)e^{\mathbf{x}_i\beta} \quad (1)$$

其中 $\lambda(t|\mathbf{x}_i) = \lim_{h \rightarrow 0} \frac{P(t \leq T^* \leq t+h | t \leq T^*, \mathbf{X}^* = \mathbf{x}_i)}{h}$ 、 $\lambda_0(t)$ 為無母數假設的風險函數、 β 為長度 p 之行向量，是我們有興趣的參數。

在未考慮左截斷時間之下為了推論 β ，我們可以藉由最大化部分概似函數 (Cox, 1975)

$$L(\beta) = \prod_{i=1}^n \left[\frac{e^{\mathbf{x}_i \beta}}{\sum_{j \in R_i} e^{\mathbf{x}_j \beta}} \right]^{\delta_i}, \quad R_i = \{j \mid y_i < y_j\}$$

得到 β 的漸近不偏估計量，稱最大部分概似估計量，記作 $\hat{\beta}^{\text{cox}}$ 。由於 \log 函數為單調遞增函數，因此最大化部分概似函數 $L(\beta)$ 等價於最小化負 \log -部分概似函數

$$\ell(\beta) = -\log L(\beta) = -\sum_{i=1}^n \delta_i \left[\mathbf{x}_i \beta - \log \left(\sum_{j \in R_i} e^{\mathbf{x}_j \beta} \right) \right] \quad (2)$$

根據 (Wang et al., 1993)，若資料需考慮左截斷時間，需要修正部分概似函數為

$$\ell^*(\beta) = -\sum_{i=1}^n \delta_i \left[\mathbf{x}_i \beta - \log \left(\sum_{j \in R_i^*} e^{\mathbf{x}_j \beta} \right) \right] \quad R_i^* = \{j \mid w_j \leq y_i < y_j\} \quad (3)$$

最大化方程式 (3)，即可得到 β 的漸近不偏估計。

2.3 模式挑選與懲戒部分概似函數

Tibshirani (1996) 提出了利用 LASSO 機制，得到能在線性模式之下同時進行模式選擇與參數估計的結果，並在 (Tibshirani et al., 1997) 推廣至 Cox 比例風險模式 (方程式 (1))，得到在 Cox 比例風險模式之下可以同時進行模式選擇與參數估計的結果；由於當樣本數大的時，參數估計值會收斂至參數真值，但是參數真實為零的參數估計值不可能剛好等於零，LASSO 機制的概念是將原本就對於推論參數有幫助的函數加上一個懲罰項，可以使原本的參數估計值朝壓縮些許，使沒有效應變數的參數估計被壓縮至 0，也就是希望參數估計結果具有稀疏性 (Sparsity)，意指參數估計值中有些為 0 有些則否，藉此達到可以同時進行模式選擇與參數估計的效果；然而參數估計值的解釋意義實際上與變數原本的單位有關，因此必須先進行去單位的動作，也就是要先將變數單範化；令 $\sum_{i=1}^n x_{ij} = 0$, $\sum_{i=1}^n x_{ij}^2 = n$, 即 x_{ij} 為已經標準化之下考慮 Cox 比例風險模式，給定調整參數 t ，並以負 \log -部分概似函數 (方程式 (2)) 作為目標函數

(objective function) , LASSO 機制則為解

$$\hat{\beta} = \operatorname{argmin}_{\beta} \{\ell(\beta)\} \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq t \quad (4)$$

做為參數估計值；根據 KKT(Karush-Kuhn-Tucker) 條件；在命題

$$x^* = \operatorname{argmin}_x f(x) \quad \text{subject to} \quad \begin{cases} h_i(x) = 0 & i = 1, \dots, m \\ g_i(x) \leq 0 & i = 1, \dots, n \end{cases}$$

之下， x^* 須滿足

$$\begin{aligned} \text{(i)} \quad & \nabla_x f(x^*) + \sum_{i=1}^m \lambda_i \nabla_x h_i(x^*) + \sum_{i=1}^n \mu_i \nabla_x g_i(x^*) = 0 \\ \text{(ii)} \quad & \nabla_{\lambda} f(x^*) + \sum_{i=1}^m \lambda_i \nabla_{\lambda} h_i(x^*) + \sum_{i=1}^n \mu_i \nabla_{\lambda} g_i(x^*) = 0 \\ \text{(iii)} \quad & \mu_i g_i(x^*) = 0, \mu_i \geq 0, i = 1, \dots, n \end{aligned} \quad (5)$$

然而 $L_{1-\text{norm}}$ 在零點不可微分的性質，造成牛頓-拉佛森法無法應用，使得解方程式 (4) 並不容易，(Tibshirani et al., 1997) 的做法為將負 log-部分概似函數在 $\beta = \check{\beta}$ 泰勒展開，得到二次逼近式

$$\begin{aligned} \ell(\beta) &\approx \ell(\check{\beta}) + (\mathbf{X}\beta - \check{\eta})^T \ell'(\check{\eta}) + (\mathbf{X}\beta - \check{\eta})^T \ell''(\check{\eta})(\mathbf{X}\beta - \check{\eta})/2 \\ &= \frac{1}{2}(z(\check{\eta}) - X\beta)^T \ell''(\check{\eta})(z(\check{\eta}) - X\beta) + C(\check{\eta}, \check{\beta}) \end{aligned} \quad (6)$$

其中 $\eta = \mathbf{X}\beta$ 、 $\ell'(\eta)$ 、 $\ell''(\eta)$ 表示 $\ell(\beta)$ 對於 η 的梯度向量 (Gradient) 與海森矩陣 (Hessian)、 $z(\check{\eta}) = \check{\eta} - \ell''(\check{\eta})^{-1} \ell'(\check{\eta})$ 、 $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ ，且 $C(\check{\eta}, \check{\beta})$ 為與 β 無關之函數，因此方程式 (4) 的解得以以二次規劃之命題逼近

$$\hat{\beta} = \operatorname{argmin}_{\beta} \frac{1}{2}(z(\check{\eta}) - X\beta)^T \ell''(\check{\eta})(z(\check{\eta}) - X\beta) \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq t \quad (7)$$

根據 Lagrange Multiplier，此命題等價於 $L_{1-\text{norm}}$ 懲戒 log-部分概似函數

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left[\frac{1}{2n} \sum_{i=1}^n v(\eta)_i (z(\eta)_i - \mathbf{x}_i \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right] \quad (8)$$

其中 $v(\eta)_i = \ell''(\eta)_{ii}$ 、 λ 為調整參數且與 t 有對應關係。如上所述， $\hat{\beta}$ 必須滿足 KKT(Karush-Kuhn-Tucker) 條件；然而絕對值函數 $|\beta|$ 在 $\beta = 0$ 無法定義微分，

因此在 $x = 0$ 時須考慮 subgradient

$$\frac{d|\beta|}{d\beta} = \begin{cases} \text{sign}(\beta) & \beta \neq 0 \\ \{u : |u| \leq 1\} & \beta = 0 \end{cases} \quad (9)$$

則簡化方程式 (5) 為

$$\begin{cases} n^{-1} \mathbf{X}_j^T \mathbf{V}(\hat{\eta})(z(\hat{\eta}) - \mathbf{X}\hat{\beta}) = \lambda \text{sign}(\hat{\beta}_j) & \hat{\beta}_j \neq 0 \\ n^{-1} |\mathbf{X}_j^T \mathbf{V}(\hat{\eta})(z(\hat{\eta}) - \mathbf{X}\hat{\beta})| \leq \lambda & \hat{\beta}_j = 0 \end{cases} \quad (10)$$

其中 $\mathbf{V}(\eta) = \text{diag}\{v(\eta)_i\}_{i=1}^n$ ，因為 $\hat{\beta}_j = 0$ 是一個區間，所以 $\hat{\beta}$ 具有稀疏性，而達到了同時進行模式選擇與參數估計的效果。

Fan and Li (2002) 則考慮不同的懲戒函數；不同於 $L_{1-\text{norm}}$ 懲戒函數，他們提出了 Smoothly Clipped Absolute Deviation (SCAD) 懲戒函數

$$p_{\lambda,\gamma}^S(|\theta|) = \begin{cases} \lambda\theta & \text{if } |\theta| \leq \lambda \\ \frac{\gamma\lambda\theta - 0.5(\theta^2 + \lambda^2)}{\gamma - 1} & \text{if } \lambda < |\theta| \leq \gamma\lambda \\ \frac{\lambda^2(\gamma^2 - 1)}{2(\gamma - 1)} & \text{if } |\theta| > \gamma\lambda \end{cases} \quad (11)$$

如同方程式 (8)，可以寫出 SCAD 懲戒 log-部分概似函數

$$\hat{\beta} = \text{argmin}_{\beta} \left[\frac{1}{n} \ell(\beta) + \sum_{j=1}^p p_{\lambda,\gamma}^S(|\beta_j|) \right] \quad (12)$$

此篇文章也提出可以用區域二次逼近式 (Local quadratic approximation, LQA)

逼近懲戒函數

$$p_{\lambda}(|\beta|) \approx p_{\lambda}(|\beta_j^{(0)}|) + \frac{p'_{\lambda}(|\beta_j^{(0)}|)}{2|\beta_j^{(0)}|} (|\beta_j|^2 - |\beta_j^{(0)}|^2) \quad (13)$$

如此一來可以避開懲戒函數在零不可微分的問題，所以可以應用牛頓-拉佛森法，也因此可以估計 $\hat{\beta}(\lambda)$ 的變異 (關於估計量變異數估計的部分會在第三章研究方法的變異數估計部分更仔細的描述)；藉由估計量變異數估計，推論出最小化方程式 (12) 得到的估計量擁有 Oracle 性質，也就是估計量與 Oracle 估計量

的漸近性質一致，其中 Oracle 估計量意指已知有效應的變數為何並且模式只配適有效應的變數得到的最大部分概似估計量，因為沒有過度解釋模式，Oracle 估計量比配飾全部變數得到的最大部分概似估計量更具效率。

另外 Zhang (2010) 提出 Minimum concave penalty(MCP) 懲戒函數如方程式 (14)，也得到具有 Oracle 性質的參數估計量，然而不同於 $L_{1-\text{norm}}$ 是凸函數 (convex function)，SCAD 懲戒函數與 MCP 懲戒函數屬於非凹函數 (non-concave function)，使這兩個懲戒函數在給定 λ 之下較難推論唯一性

$$p_{\lambda, \gamma}^M(|\theta|) = \begin{cases} \lambda\theta - \frac{\theta^2}{2\gamma} & \text{if } \theta \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 & \text{if } \theta > \gamma\lambda \end{cases} \quad (14)$$

Zou (2006) 提出 adaptive LASSO(ALASSO)，令 $\lambda_j = \lambda/|\hat{\beta}^{\text{cox}}|$ ，如此一來本來是 p 維度的調整參數 $(\lambda_1, \lambda_2, \dots, \lambda_p)$ 可以以最大部分概似估計量為反比降至一維度；概念就是希望參數估計值大的參數應該朝零壓縮少一點，參數估計值小的參數應該朝零壓縮多一點，此文章也證明 ALASSO 估計量擁有 Oracle 性質；Zhang and Lu (2007) 將此概念推廣到 Cox 比例風險模式 (方程式 (15))，得到在 Cox 比例風險模式之下擁有 Oracle 性質的估計量

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left[\frac{1}{n} \ell(\beta) + \sum_{j=1}^p \lambda_j |\beta_j| \right] \quad (15)$$

不論是 SCAD、MCP 或者 ALASSO，最小化懲戒 log-部分概似函數皆須滿足 KKT 條件 (如方程式 (10))，而在計算懲戒函數的一階微分需考慮連鎖率，綜合以上四種不同的懲戒函數，KKT 條件可以寫成一般式

$$\begin{cases} n^{-1} \mathbf{X}_j^T \mathbf{V}(\hat{\eta})(z(\hat{\eta}) - \mathbf{X}\hat{\beta}) = p'_{\lambda_j}(|\hat{\beta}_j|) \operatorname{sign}(\hat{\beta}_j) & \hat{\beta}_j \neq 0 \\ n^{-1} |\mathbf{X}_j^T \mathbf{V}(\hat{\eta})(z(\hat{\eta}) - \mathbf{X}\hat{\beta})| \leq \lambda & \hat{\beta}_j = 0 \end{cases} \quad (16)$$

$$\begin{aligned}
\text{LASSO} & : p'_{\lambda_j}(|\theta|) = \lambda_j \text{sign}(\theta) , \lambda_j \equiv \lambda \\
\text{SCAD} & : p'_{\lambda_j}(|\theta|) = p'^S_{\lambda_j, \gamma}(|\theta|) , \lambda_j \equiv \lambda, \gamma = 3.7 \\
\text{MCP} & : p'_{\lambda_j}(|\theta|) = p'^M_{\lambda_j, \gamma}(|\theta|) , \lambda_j \equiv \lambda, \gamma = 3.7 \\
\text{ALASSO} & : p'_{\lambda_j}(|\theta|) = \lambda_j \text{sign}(\theta) , \lambda_j = \lambda |\hat{\beta}_j^{\text{cox}}|^{-1}
\end{aligned}$$

由方程式 (11) 與方程式 (14) 微分可以計算得到

$$p'^S_{\lambda, \gamma}(|\theta|) = \begin{cases} \lambda & \text{if } |\theta| \leq \lambda \\ \frac{\gamma\lambda - \theta}{\gamma - 1} & \text{if } \lambda < |\theta| \leq \gamma\lambda \\ 0 & |\theta| > \gamma\lambda \end{cases}, \quad p'^M_{\lambda, \gamma}(|\theta|) = \begin{cases} \lambda - \frac{\theta}{\gamma} & \text{if } \theta \leq \gamma\lambda \\ 0 & \text{if } \theta > \gamma\lambda \end{cases}$$

2.4 群懲戒函數

Yuan and Lin (2006) 提出在線性模式之下進行群變數選取，目的在於研究人員希望指定無重疊的變數群，可以同時被選進迴歸模式或者同時被踢除迴歸模式，例如離散型共變數，參數估計的解釋意義為對比基準水準的效應程度，若不考慮群變數挑選，LASSO 機制會將參數估計朝 0 壓縮，也就是把效應程度估計成與基準水準的效應相同；但是這樣的結論不是我們希望的，因為替換別的水準當作基準水準，會使模式選擇與參數估計的結果不一致這；群懲戒函數不對每個參數朝 0 壓縮，而是對指定群變數參數的 2-norm 朝 0 壓縮，這樣的懲戒函數會讓指定群的參數估計值同時為 0 或者同時不為 0，我們稱為群稀疏性；其次，編碼方式或者基準水準的不同皆有可能影響變數挑選的結果，因此被指定為群的變數應該為正交單範化。

令線性模式 $y = \mathbf{X}\beta + \epsilon$, $\mathbf{X}\beta = \sum_{j=1}^J \mathbf{X}_j \beta_j$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J)$, \mathbf{X}_j 為對應至第 j 群之 $n \times p_j$ 群共變數矩陣, $\beta = (\beta'_1, \beta'_2, \dots, \beta'_J)'$, β_j 是長度為 p_j 之行向量，其中 $\mathbf{X}_j^T \mathbf{X}_j = \mathbf{I}_{p_j \times p_j}$, $j = 1, \dots, J$ ，可以藉由各種正交單範轉換得到，並訂定懲戒函數定為 $\sum_{j=1}^J \sqrt{p_j} \|\beta_j\|$ ，其中 $\|\beta_j\| = \sqrt{\beta_j' \beta_j}$ 表示參數的 2-norm；藉由

最小化群 $L_{1-\text{norm}}$ 懲戒最小平方差

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left[\frac{1}{n} \|y - \mathbf{X}\beta\|^2 + \lambda \sum_{j=1}^p \sqrt{p_j} \|\beta_j\| \right] \quad (17)$$

達到群變數選取的效果；因為 $L_{1-\text{norm}}$ 在 $\beta_j = 0$ 無法定義梯度向量，因此 $\hat{\beta}$ 必須滿足 KKT 條件

$$\begin{cases} n^{-1} \mathbf{X}_j^T (y - \mathbf{X}\hat{\beta}) = \lambda \sqrt{p_j} \frac{\hat{\beta}_j}{\|\hat{\beta}_j\|} & \hat{\beta}_j \neq 0 \\ n^{-1} \|\mathbf{X}_j^T (y - \mathbf{X}\hat{\beta})\| \leq \lambda \sqrt{p_j} & \hat{\beta}_j = 0 \end{cases} \quad (18)$$

由於 $\hat{\beta}_j = 0$ 是一個 p_j 維度的集合，所以 $\hat{\beta}$ 有群稀疏性，而達到了群變數挑選與參數估計的結果；值得注意的是在線性模式之下，當 $p_j \equiv 1$ 時，方程式 (16) 等價於方程式 (18)，因此群懲戒函數可以視為懲戒函數的推廣結果

Wang et al. (2007) 將 $L_{1-\text{norm}}$ 懲戒函數改成 group SCAD(GSCAD) 懲戒函數 $\sum_{j=1}^p p_{\lambda_j}^S(\|\beta_j\|)$ ， $\lambda_j = \lambda \sqrt{p_j}$ ，而得到不錯的模式選擇結果；Wang and Leng (2008) 提出 Adaptive group LASSO(AGLASSO)，令 $\lambda_j = \lambda \sqrt{p_j} / \|\hat{\beta}_j^{\text{cox}}\|$ 也得到不錯的模式選擇結果；

2.5 正規解路徑

懲戒函數在 $\beta_j = 0$ 不可微分的性質使得最小化懲戒部分概似函數可以同時進行模式選擇與參數估計，然而此性質也使得在給定調整參數 λ 之下計算滿足 KKT 條件 (方程式 (18)) 的 $\hat{\beta}$ 有相當的難度，將 $\hat{\beta}$ 沿著 λ 描繪，即可得到正規解路徑，如圖 (1)。

Efron et al. (2004) 提出 Least Angle Regression(LARS) 演算法，藉由在線性模式之下，LASSO 機制的正規解路徑具有分段線性 (Piecewise linear) 的性質，巧妙的得到可以快速描繪出正規解路徑的演算法，然而在 Cox 比例風險模式之下，LASSO 機制的正規解路徑不具有分段線性的性質，因此在 Cox 比例風險模式之下不能應用 LARS 演算法。

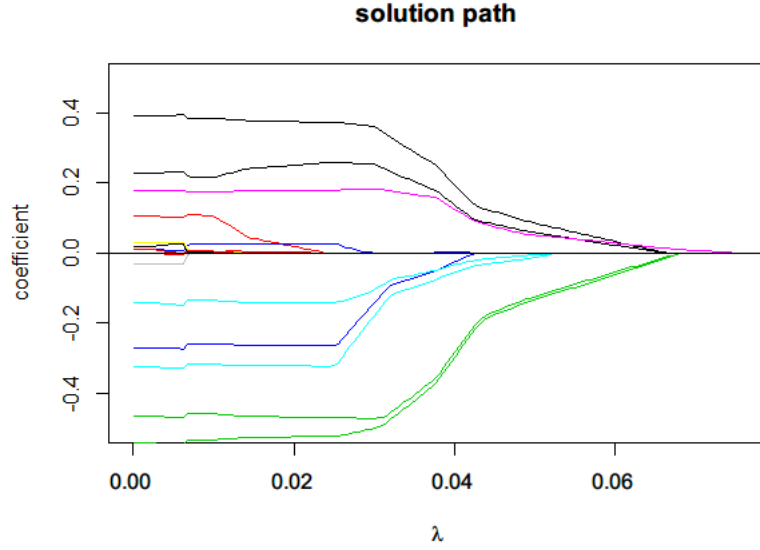


Figure 1: 正規解路徑

Friedman et al. (2010) 提出 coordinate descent algorithm，能沿著 λ 描繪出廣義線性模式的正規解路徑；Simon et al. (2011) 則將演算法推廣到可以用在 Cox 比例風險模式上；此演算法的核心概念為一次只調整參數估計的第 j 維度，即最小化單變數之懲戒 log-部分概似函數 $[\frac{1}{2n}\ell(\beta_j|\beta_{-j}) + \lambda|\beta_j|]$ ，接著不斷迭代參數的每一個維度直到滿足 KKT 條件方程式 (10)，由於單變數的懲戒函數已有公式解，因此此演算法不但能有效率地得到 Cox 比例風險模式的正規解路徑，也很容易推廣到不同的懲戒函數，如 SCAD、MCP。

Breheny and Huang (2015) 將 coordinate descent algorithm 推廣到 group descent 演算法，有效率地得到廣義線性模式的正規解路徑；與 coordinate descent(GD) algorithm 相似，GD 演算法一次只調整參數估計的第 j 群，即最小化 $[\frac{1}{2n}\ell(\beta_j|\beta_{-j}) + \lambda||\beta_j|]$ ，藉由不斷迭代每一群參數直到滿足 KKT 條件方程式 (18)，即可計算 $\hat{\beta}$ 。

Zou (2008) 則是直接將 log-部分概似函數再最大部分概似估計量展開，如方程式 (6) $\ell(\beta) \approx \ell(\hat{\beta}^{\text{cox}}) + (\beta - \hat{\beta}^{\text{cox}})^T \dot{\ell}(\hat{\beta}^{\text{cox}}) + (\beta - \hat{\beta}^{\text{cox}})^T \ddot{\ell}(\hat{\beta}^{\text{cox}})(\beta - \hat{\beta}^{\text{cox}})/2$ 如同

方程式 (8)，以懲戒二次函數逼近懲戒 log-部分概似函數

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left[\frac{1}{2n} (\beta - \hat{\beta}^{\text{cox}})^T \ddot{\ell}(\hat{\beta}^{\text{cox}}) (\beta - \hat{\beta}^{\text{cox}}) + \sum_{j=1}^p \lambda_j |\beta_j| \right] \quad (19)$$

等價於

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left[\frac{1}{2n} \|u - V\beta\|^2 + \sum_{j=1}^p \lambda_j |\beta_j| \right] \quad (20)$$

其中 $V^T V = \ddot{\ell}(\hat{\beta}^{\text{cox}})$ ， $u = V\hat{\beta}^{\text{cox}}$ ，藉此回到線性模式的懲戒函數，即可應用

LARS 演算法得出正規解路徑



3 研究方法

3.1 群懲戒 log-部分概似函數

由於資料為左截斷且右設限，若未修正估計方法，將得到有偏誤的參數估計量；令資料 $(y_1, \mathbf{x}_1, w_1, \delta_1), \dots, (y_n, \mathbf{x}_n, w_n, \delta_n)$ 為隨機樣本， y_i 為觀察到之時間、 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ 為受測者的 p 個變數、 w_i 為左截斷時間、 δ_i 為觀察到之時間的狀態，1 代表存活時間，0 代表右設限時間。令 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J)$ ， \mathbf{X}_j 為對應至第 j 群之 $n \times p_j$ 群共變數矩陣， $\beta = (\beta'_1, \beta'_2, \dots, \beta'_J)'$ ， β_j 是長度為 p_j 之行向量，其中 $\mathbf{X}_j^T \mathbf{X}_j = \mathbf{I}_{p_j \times p_j}$ ， $j = 1, \dots, J$ ；參考方程式 (3) 為考慮左截斷時間的 log-部分概似函數，將其最大化可以得到 β 的漸近不偏估計量，然而此估計量不具有稀疏性；為了讓我們提出的估計量具有稀疏性，也就是此估計量具有同時選擇模式與參數估計的效果，同時也希望能處理離散型變數，也就是希望群變數能同時被選入模式或者剔除模式，我們參考懲戒函數的概念，我們可以藉由最小化群懲戒 log-部分概似函數得到我們希望的效果

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{n} \ell^*(\beta) + \sum_{j=1}^J p_{\lambda_j}(\|\beta_j\|) \right\} \quad (21)$$

如同文獻回顧提到， $\|\beta\|$ 在 $\beta = 0$ 時無法定義梯度向量，再加上參考 KKT 條件 (方程式 (18))，可知得到的估計量能達到群變數挑選的效果；同樣地，在給定 λ 之下解方程式 (21) 並不簡單，參考將以方程式 (7) 逼近方程式 (4) 的做法，我們可以把 $\ell^*(\beta)$ 在 $\hat{\beta}^{\text{cox}}$ 泰勒展開，得到二次逼近式如方程式 (6)

$$\frac{1}{n} \ell^*(\beta) \approx \frac{1}{2n} (\beta - \hat{\beta}^{\text{cox}})^T \ddot{\ell}^*(\hat{\beta}^{\text{cox}}) (\beta - \hat{\beta}^{\text{cox}}) \quad (22)$$

其中 $\hat{\beta}^{\text{cox}}$ 為最小化 $\ell^*(\beta)$ 得到對 β 的漸近不偏估計量，因此方程式 (21) 可以以線性模式逼近

$$\begin{aligned}\hat{\beta} &= \operatorname{argmin}_{\beta} \left[\frac{1}{2n} (\beta - \hat{\beta}^{\text{cox}})^T \ddot{\ell}^*(\hat{\beta}^{\text{cox}}) (\beta - \hat{\beta}^{\text{cox}}) + \sum_{j=1}^J p_{\lambda_j}(\|\beta_j\|) \right] \\ &= \operatorname{argmin}_{\beta} \left[\frac{1}{2n} \|u - \mathbf{V}\beta\|^2 + \sum_{j=1}^J p_{\lambda_j}(\|\beta_j\|) \right]\end{aligned}\quad (23)$$

其中 $\mathbf{V}^T \mathbf{V} = \ddot{\ell}^*(\hat{\beta}^{\text{cox}})$ ， $u = \mathbf{V}\hat{\beta}^{\text{cox}}$ ， $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_J)$ 藉此回到線性迴歸模式；參考方程式 (17)，並且參考 (Wang et al., 2007) 與 (Wang and Leng, 2008) GSCAD 與 AGLASSO 的做法，方程式 (23) 也可以考慮四個不同的懲戒函數

$$\begin{aligned}\text{GLASSO} &: p_{\lambda_j}(\|\beta_j\|) = \lambda_j \|\beta_j\|, \lambda_j = \lambda \sqrt{p_j} \\ \text{GSCAD} &: p_{\lambda_j}(\|\beta_j\|) = p_{\lambda_j, \gamma}^S(\|\beta_j\|), \lambda_j = \lambda \sqrt{p_j}, \gamma = 3.7 \quad (\text{方程式 (11)}) \\ \text{GMCP} &: p_{\lambda_j}(\|\beta_j\|) = p_{\lambda_j, \gamma}^M(\|\beta_j\|), \lambda_j = \lambda \sqrt{p_j}, \gamma = 3.7 \quad (\text{方程式 (14)}) \\ \text{AGLASSO} &: p_{\lambda_j}(\|\beta_j\|) = \lambda_j \|\beta_j\|, \lambda_j = \lambda \sqrt{p_j} \|\hat{\beta}_j^{\text{cox}}\|^{-1}\end{aligned}$$

由於以上四個懲戒函數在 $\beta = 0$ 時皆無法定義梯度向量，因此必須考慮 subgradient

$$\nabla \|\beta\| = \begin{cases} \frac{\beta}{\|\beta\|} & x \neq 0 \\ \{u : \|u\| \leq 1\} & x = 0 \end{cases}\quad (24)$$

可得到 $\hat{\beta}$ 須滿足 KKT 條件

$$\begin{cases} n^{-1} \mathbf{V}_j^T (u - \mathbf{V}\hat{\beta}) = p'_{\lambda_j}(\|\hat{\beta}_j\|) \frac{\hat{\beta}_j}{\|\hat{\beta}_j\|} & \hat{\beta}_j \neq 0 \\ n^{-1} \|\mathbf{V}_j^T (u - \mathbf{V}\hat{\beta})\| \leq \lambda_j & \hat{\beta}_j = 0 \end{cases}\quad (25)$$

$p'_{\lambda}(\theta)$ 沿用方程式 (16) 的符號，表示對應不同的懲戒函數的一階微分；將 $\hat{\beta}$ 沿著 λ 描繪，即可得到正規解路徑。然而不同於 Zou (2008) 考慮 $L_{1-\text{norm}}$ 懲戒函數而得以應用 LARS 演算法，因為我們提出的懲戒 log-部分概似函數的懲戒函數考慮群變數挑選，因此正規解路徑不具有分段線性的性質

3.2 演算法

令群變數 $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_J)$ 表示指定要同時被選入模式或者剔除模式的變數群，每群分別有 p_j 個變數， $j = 1, \dots, J$ ，利用奇異值分解

$$\frac{1}{n} \tilde{\mathbf{X}}_j^T \tilde{\mathbf{X}}_j = \mathbf{Q}_j \Lambda_j \mathbf{Q}_j^T$$

其中 \mathbf{Q}_j 與 Λ_j 分別為 $n^{-1} \tilde{\mathbf{X}}_j^T \tilde{\mathbf{X}}_j$ 的特徵向量矩陣與特徵值的對角線矩陣，則建構 $\mathbf{X}_j = \tilde{\mathbf{X}}_j \mathbf{Q}_j \Lambda_j^{-1/2}$ ，可以很快的得到以下兩個性質

$$\begin{aligned} \frac{1}{n} \mathbf{X}_j^T \mathbf{X}_j &= \mathbf{I}_{p_j \times p_j} \\ \mathbf{X}_j \beta_j &= \tilde{\mathbf{X}}_j (\mathbf{Q}_j \Lambda_j^{-1/2} \beta_j) \end{aligned} \quad (26)$$

由此可知 $\tilde{\beta}_j = \mathbf{Q}_j \Lambda_j^{-1/2} \beta_j$ 之轉換關係。應用 Cholesky 分解 $\mathbf{V}^T \mathbf{V} = \ddot{\ell}^*(\hat{\beta}^{\text{cox}})$ 即可得到 \mathbf{V} ；參考 Yuan and Lin (2006) 提出的 GD 演算法的概念，考慮 L_1 -norm 懲戒函數，最小化 $\frac{1}{2n} \|r_{-j} - \mathbf{V}_j \beta_j\|^2 + \lambda_j \|\beta_j\|$ ，即固定 β_{-j} 之下一次僅調整第 j 群參數，其中 $r_{-j} = u - \mathbf{V}_{-j} \beta_{-j}$ 為部分殘差，即固定 β_{-j} 之下的殘差值，如此等價於僅解第 j 群參數的條件方程式 (25)，藉由不斷迭代每群參數，得到給定 λ 之下的參數估計值；首先考慮 $\frac{1}{n} \mathbf{V}_j^T \mathbf{V}_j = \mathbf{I}_{p_j}$ ，在給定 β_{-j} 之下

case1: $\beta_j = 0$

$$\begin{aligned} \lambda_j &\geq n^{-1} \|\mathbf{V}_j^T (r_{-j} - \mathbf{V}_j \beta)\| \\ &= n^{-1} \|\mathbf{V}_j^T r_{-j}\| \end{aligned}$$

case2: $\beta_j \neq 0$

$$\begin{aligned} \lambda_j \frac{\beta_j}{\|\beta_j\|} &= n^{-1} \mathbf{V}_j^T (r_{-j} - \mathbf{V}_j \beta) \\ (\mathbf{V}_j^T \mathbf{V}_j = n \mathbf{I}_{p_j}) &\Rightarrow (1 + \frac{\lambda_j}{\|\beta_j\|}) \beta_j = n^{-1} \mathbf{V}_j^T r_{-j} \\ &\Rightarrow (\|\beta_j\| + \lambda_j) = n^{-1} \|\mathbf{V}_j^T r_{-j}\| \end{aligned} \quad (27)$$

Case1 : $\beta = 0 \Leftrightarrow n^{-1} \|\mathbf{V}_j^T r_{-j}\| \leq \lambda_j$

$$\text{Case2 : } \begin{cases} \|\beta_j\| = (n^{-1} \|\mathbf{V}_j^T r_{-j}\| - \lambda_j)^+ \\ \beta_j / \|\beta_j\| = \mathbf{V}_j^T r_{-j} / \|\mathbf{V}_j^T r_{-j}\| \quad (\text{方程式 (27) 中第二式與第三式相除}) \end{cases}$$

因為 $\beta_j = \frac{\beta_j}{\|\beta_j\|} \|\beta_j\|$ ，所以綜合 Case1 與 Case2 可以得到 β_j 的公式解

$$\beta_j \leftarrow S(n^{-1} \mathbf{V}_j^T r_{-j}, \lambda_j) \quad (28)$$

其中 $S(\mathbf{z}, \lambda) = (\|\mathbf{z}\| - \lambda)^+ \frac{\mathbf{z}}{\|\mathbf{z}\|}$, $(a - \lambda)^+ = (a - \lambda)I(a \geq \lambda)$ ；然而在 Cox 模式之

下， $\frac{1}{n} \mathbf{V}_j^T \mathbf{V}_j \neq \mathbf{I}_{p_j}$ ，雖然 Case1 可以得到一樣的式子判斷 β 是否等於零，但是

在 Case2 的時候沒有公式解；考慮 LQA(方程式 (13))

$$p_\lambda(\|\beta_j\|) \approx p_\lambda(\|\beta_j^{(0)}\|) + \frac{p'_\lambda(\|\beta_j^{(0)}\|)}{2\|\beta_j^{(0)}\|} (\|\beta_j\|^2 - \|\beta_j^{(0)}\|^2)$$

$$\operatorname{argmin}_\beta \left[\frac{1}{2n} \|r_{-j} - \mathbf{V}_j \beta_j\|^2 + \lambda_j \|\beta_j\| \right] \approx \operatorname{argmin}_\beta \left[\frac{1}{2n} \|r_{-j} - \mathbf{V}_j \beta_j\|^2 + \frac{\lambda_j \|\beta_j\|^2}{\|\beta_j^{(0)}\|} \right]$$

也就是 ridge regression 的形式，有公式解

$$\beta_j = (\mathbf{V}_j^T \mathbf{V}_j + \frac{n\lambda_j}{\|\beta_j^{(0)}\|} \mathbf{I}_{p_j})^{-1} \mathbf{V}_j^T r_{-j} \quad (29)$$

同理，若考慮不同的懲戒函數 $p_{\lambda_j}(\theta)$

$$\beta_j = (\mathbf{V}_j^T \mathbf{V}_j + \frac{np'_{\lambda_j}(\|\beta_j^{(0)}\|)}{\|\beta_j^{(0)}\|} \mathbf{I}_{p_j})^{-1} \mathbf{V}_j^T r_{-j} \quad (30)$$

$p'_\lambda(\theta)$ 可以參考方程式 (16)

Algorithm 3.1.

1. $\beta^{(0)}$: 起始值
 2. $r = u - \mathbf{V}\beta^{(0)}$
 3. **repeat** $r=0, \dots$
 4. **for** $j = 1, 2, \dots, J$
 5. $r_{-j} = r + \mathbf{V}_j\beta_j^{(r)}$
 6. $\beta'_j \leftarrow (\mathbf{V}_j^T \mathbf{V}_j + \frac{np'_{\lambda_j}(\|\beta_j^{(0)}\|)}{\|\beta_j^{(r)}\|} \mathbf{I}_{p_j})^{-1} \mathbf{V}_j r_{-j} I(\|\mathbf{V}_j r_{-j}\| \geq n\lambda_j)$
 7. $R \leftarrow R - \mathbf{X}_j^T (\beta'_j - \beta_j^{(r)})$
 8. $\beta_j^{(r+1)} \leftarrow \beta'_j$
 9. 執行至收斂
-

應用 (Breheny and Huang, 2015) 中寫出目標函數的 Majorization(優化) 函數，藉由最小化優化函數加上原本的懲戒函數即可處理 $\mathbf{V}^T \mathbf{V} \neq \mathbf{I}$ 情況，我們稱為 Group Majorization Descent(GMP) 演算法；相較之下，GD 演算法只能處理 $\mathbf{V}^T \mathbf{V} = \mathbf{I}$ 的情況；建構 $M(\beta)$ 的優化函數 $\tilde{M}(\beta|\tilde{\beta})$

$$\begin{aligned}
 \tilde{M}(\beta|\tilde{\beta}) &= M(\tilde{\beta}) + (\beta - \tilde{\beta})^T \nabla M(\tilde{\beta}) + \frac{1}{2} \xi (\beta - \tilde{\beta})^T (\beta - \tilde{\beta}) \\
 &\geq M(\tilde{\beta}) + (\beta - \tilde{\beta})^T \nabla M(\tilde{\beta}) + \frac{1}{2} (\beta - \tilde{\beta})^T \nabla^2 M(\tilde{\beta}) (\beta - \tilde{\beta}) \quad (31) \\
 &\geq M(\beta)
 \end{aligned}$$

令 ξ 是 $\nabla^2 M(\tilde{\beta})$ 最大的特徵值，很明顯地，對於任意的 $\beta, \tilde{\beta}$

$$\tilde{M}(\beta|\tilde{\beta}) \geq M(\beta)$$

$$\tilde{M}(\beta|\tilde{\beta} = \beta) = M(\beta)$$

因此可以應用 Majorize-Minimization(MM) 演算法，最小化 Majorization 函數 $\tilde{M}(\beta|\tilde{\beta})$ 並且反覆迭代，即可得到 $\arg\min_{\beta} M(\beta)$ ；如同前面所述，我們的目標

為最小化 $\frac{1}{2n}||r_{-j} - \mathbf{V}_j\beta_j||^2 + \lambda_j||\beta_j||$ ，即設定

$$\begin{aligned}
M(\beta_j) &= ||r_{-j} - \mathbf{V}_j\beta_j||^2 \\
\tilde{M}(\beta_j|\tilde{\beta}_j) &= ||r_{-j} - \mathbf{V}_j\tilde{\beta}_j||^2 + (\beta_j - \tilde{\beta}_j)^T \nabla M(\tilde{\beta}_j) + \frac{\xi_j}{2}(\beta_j - \tilde{\beta}_j)^T(\beta_j - \tilde{\beta}_j) \\
&= ||r_{-j} - \mathbf{V}_j\tilde{\beta}_j||^2 - (\beta_j - \tilde{\beta}_j)^T \mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta}_j) + \frac{1}{2}\xi_j||(\beta_j - \tilde{\beta}_j)||^2
\end{aligned} \tag{32}$$

其中 ξ_j 為 $\nabla^2 M(\beta) = \mathbf{V}_j^T \mathbf{V}$ 最大的特徵值；則 $\beta_j = \operatorname{argmin}_{\beta} \frac{1}{2n} M(\beta_j|\tilde{\beta}_j)$

$+ \lambda_j||\beta_j||$ 須滿足 KKT 條件

$$\begin{aligned}
n^{-1} \mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta}_j) &= \lambda_j \frac{\beta_j}{||\beta_j||} + n^{-1} \xi_j \beta_j \quad \beta_j \neq 0 \\
n^{-1} ||\mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta}_j)|| &\leq \lambda_j \quad \beta_j = 0
\end{aligned} \tag{33}$$

如同推導方程式 (27) 與方程式 (28) 的過程，可以得到

$$\begin{aligned}
\text{Case1: } \beta &= 0 \Leftrightarrow n^{-1} ||\mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta})|| \leq \lambda_j \\
\text{Case2: } \begin{cases} n^{-1} \xi_j ||\beta_j|| &= (n^{-1} ||\mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta})|| - \lambda_j)^+ \\ \beta_j / ||\beta_j|| &= \mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta}) / ||\mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta})|| \end{cases}
\end{aligned}$$

因為 $\beta_j = \frac{n}{\xi_j} \frac{\beta_j}{||\beta_j||} n^{-1} \xi_j ||\beta_j||$ ，所以可以得到 GMD 演算法的 β_j 的公式解

$$\beta_j = \frac{n}{\xi_j} S(n^{-1} \mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta}_j) + n^{-1} \xi_j \beta_j, \lambda_j) \tag{34}$$

不斷迭代 $\operatorname{argmin}_{\beta} \tilde{M}(\beta_j|\tilde{\beta}) + \lambda_j||\beta_j||$ 即可得到 $\operatorname{argmin}_{\beta} M(\beta_j) + \lambda_j||\beta_j||$ ；然而在

GSCAD、GMCP 等懲戒函數之下 $\tilde{M}(\beta_j|\tilde{\beta}_j) + p_{\lambda}(|\beta_j|)$ 沒有公式解；參考 (Zou

and Li, 2008) 的區域線性逼近 (Local linear approximation, LLA)

$$p_{\lambda}(|\beta|) \approx p_{\lambda}(|\beta^{(0)}|) + p'_{\lambda}(|\beta^{(0)}|)(|\beta| - |\beta^{(0)}|) \tag{35}$$

$$\operatorname{argmin}_{\beta_j} [\tilde{M}(\beta_j|\tilde{\beta}_j) + p_{\lambda_j}(|\beta_j|)] \approx \operatorname{argmin}_{\beta_j} [\tilde{M}(\beta_j|\tilde{\beta}_j) + p'_{\lambda_j}(|\tilde{\beta}_j|)|\beta_j|]$$

也就是在 GSCAD、GMCP 懲戒函數之下解參數，可以以 $\lambda_j = p'_{\lambda_j}(|\tilde{\beta}_j|)$ 的

$L_{1-\text{norm}}$ 來逼近，因此，可以得到到迭代式

$$\beta_j = \frac{n}{\xi_j} S(n^{-1} \mathbf{V}_j^T(r_{-j} - \mathbf{V}_j\tilde{\beta}_j) + n^{-1} \xi_j \beta_j, p'_{\lambda_j}(|\tilde{\beta}_j|)) \tag{36}$$

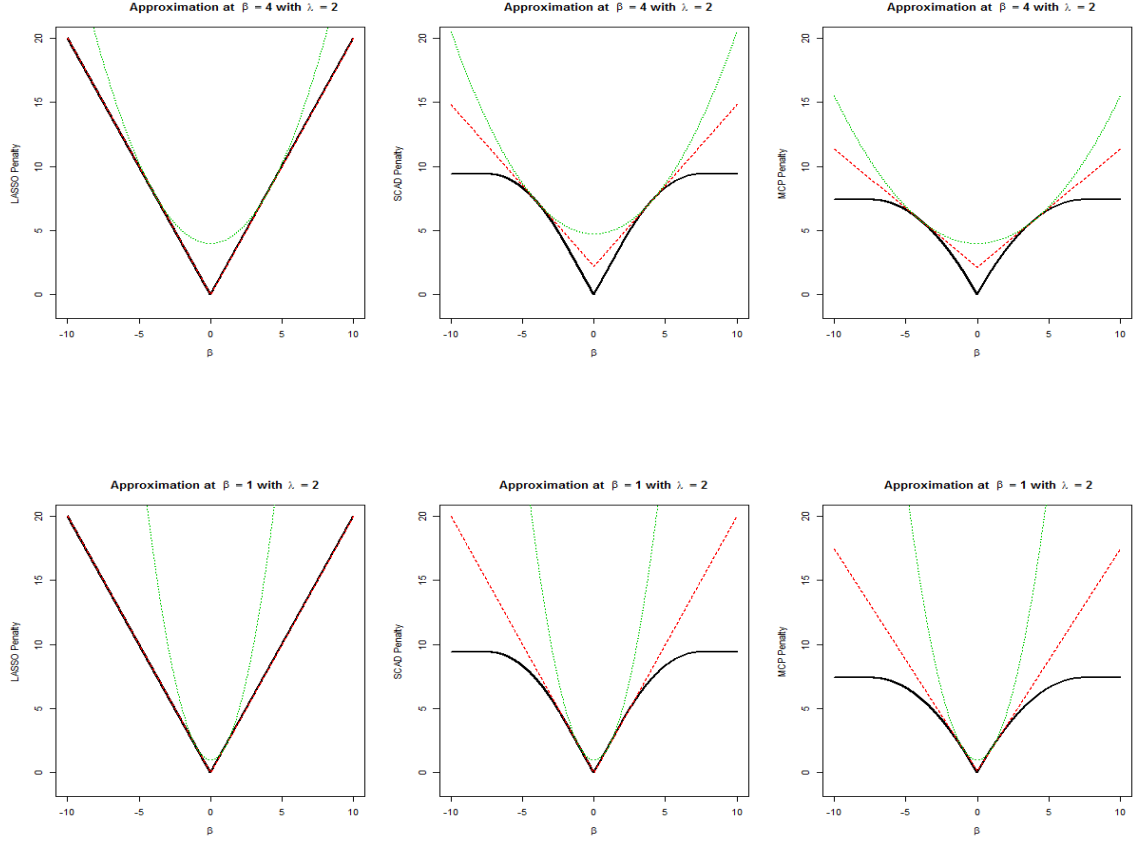


Figure 2: 黑色實線表示懲戒函數值、綠色細虛線表示 LQA 逼近值、紅色虛線表示 LLA 逼近值；由左而右分別為 LASSO、SCAD、MCP 懲戒函數

簡而言之，GMD 演算法即為不斷迭代 $\operatorname{argmin}_{\beta_j} \tilde{M}(\beta_j | \tilde{\beta}) + p'_{\lambda_j}(\|\tilde{\beta}_j\|) \|\beta_j\|$ 即可得到 $\operatorname{argmin}_{\beta_j} M(\beta_j) + p_{\lambda_j}(\|\beta_j\|)$ ， $p'_{\lambda}(\theta)$ 可以參考方程式 (16)

Algorithm 3.2. GMD 演算法

1. $\beta^{(0)}$: 起始值
2. $\xi_j = n^{-1} \mathbf{V}_j^T \mathbf{V}_j$ 最大的特徵值
3. $r = u - \mathbf{V} \beta^{(0)}$
4. **repeat** $r=0, \dots$
5. **for** $j = 1, 2, \dots, J$
6. $\beta'_j \leftarrow \frac{1}{\xi_j} S(n^{-1} \mathbf{V}_j^T r + \xi_j \beta_j, p'_{\lambda_j}(\|\beta_j\|))$
7. $R \leftarrow R - \mathbf{X}_j^T (\beta'_j - \beta_j^{(r)})$
8. $\beta_j^{(r+1)} \leftarrow \beta'_j$
9. 執行至收斂

3.3 變異數估計

根據 (Fan and Li, 2002), 區域二次逼近式為

$$p_\lambda(\|\beta_j\|) \approx p_\lambda(\|\beta_j^{(0)}\|) + \frac{p'_\lambda(\|\beta_j^{(0)}\|)}{2\|\beta_j^{(0)}\|}(\|\beta_j\|^2 - \|\beta_j^{(0)}\|^2) \quad (37)$$

因此群懲戒 log-部分概似函數又可以以區域二次逼近

$$\hat{\beta} = \operatorname{argmin}_\beta \left[\frac{1}{2n} \|u - \mathbf{V}\beta\|^2 + \sum_{j=1}^J \frac{p'_{\lambda_j}(\|\beta_j^{(0)}\|)\|\beta_j\|^2}{2\|\beta_j^{(0)}\|} \right]$$

如此一來不會遇到不可微分的問題，即可以應用牛頓-拉佛森法

$$\hat{\beta}^{r+1} = \hat{\beta}^r - \left\{ \mathbf{V}^T \mathbf{V} + nA(\hat{\beta}^r, \lambda) \right\}^{-1} \left\{ -\mathbf{V}^T(u - \mathbf{V}\hat{\beta}^r) + nb(\hat{\beta}^r, \lambda) \right\} \quad (38)$$

其中 $b(\beta, \lambda) = \left[\frac{p'_{\lambda_j}(\|\beta_j^*\|)\beta_j^*}{\|\beta_j^*\|} \right]$, $A(\beta, \lambda) = \operatorname{diag} \left\{ \frac{p'_{\lambda_j}(\|\beta_j^*\|)}{\|\beta_j^*\|} \mathbf{I}_{p_j \times p_j} \right\}$, $\beta_j^* = \beta_j + \epsilon$

根據 (Lu and Zhang, 2007))，令 β_0 為參數真值，則 $\hat{\beta}$ 的共變數矩陣對應的三

明治型式 (sandwich form) 為

$$\left\{ \nabla^2 \ell(\hat{\beta}) + nA(\hat{\beta}, \lambda) \right\}^{-1} \widehat{\operatorname{cov}}(\nabla \ell(\beta_0) + nb(\beta_0, \lambda)) \left\{ \nabla^2 \ell(\hat{\beta}) + nA(\hat{\beta}, \lambda) \right\}^{-1} \quad (39)$$

$$\widehat{\operatorname{cov}}(\nabla \ell(\beta_0) + nb(\beta_0, \lambda)) = \left\{ \nabla^2 \ell(\hat{\beta}) + nD(\hat{\beta}, \lambda) \right\} \left\{ \nabla^2 \ell(\hat{\beta}) \right\}^{-1} \left\{ \nabla^2 \ell(\hat{\beta}) + nD(\hat{\beta}, \lambda) \right\} \quad (40)$$

其中 $D(\beta, \lambda) = \operatorname{diag} \left\{ \frac{I(\beta_j=0)p'_{\lambda_j}(\|\beta_j^*\|)}{\|\beta_j^*\|} \mathbf{I}_{p_j \times p_j} \right\}$ ；結合方程式 (39) 與方程式 (40) 即

可估計 $\hat{\beta}$ 的變異數

3.4 調整參數之選取

描繪出正規解路徑後，我們需要挑選 λ ，並以 $\hat{\beta}$ 做為模式選擇與參數估計的

結果，我們運用交叉驗正法 (Cross-Validation)、GCV(Generalized Cross

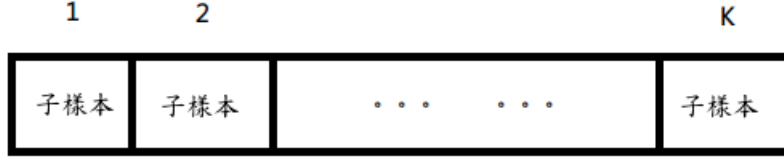


Figure 3: K-fold 交叉驗正子樣本

Validation)、BIC(Bayesian Information Criterion) 與 AIC(Akaike Information Criterion) 等四個挑選標準選擇 λ ；交叉驗正法的精神在於將原始樣本隨機切分成 K 份子樣本，如圖 (3) 顯示，將其中一份子樣本以外的樣本用來估計，並將該子樣本視為測驗資料，稱 K-fold 交叉驗正法；根據 (van Houwelingen et al., 2006) 提出在 Cox 比例風險模式之下的 K-fold 交叉驗正法為最大化

$$\widehat{CV}_k(\lambda) = \ell^*(\beta_{-k}(\lambda)) - \ell_{-k}^*(\beta_{-k}(\lambda)) \quad (41)$$

其中 ℓ_{-k}^* 表式扣除第 k 份子樣本後的樣本建構出的 log-部分概似函數， $\beta_{-k}(\lambda)$ 為最小化 $\frac{1}{n}\ell_{-k}^*(\beta) + \sum_{j=1}^J p_{\lambda_j}(|\beta_j|)$ 之解，我們選擇可以最大化 $\sum_{k=1}^K \widehat{CV}_k(\lambda)$ 之 λ 做為 K-fold 交叉驗正法的結果。另外，不論是 GSCAD、GMCP 或是 AGLASSO，由於調整參數皆降至一維，因此可以考慮挑選標準

$$\begin{aligned} \text{GCV}(\lambda) &= \ell^*(\hat{\beta}(\lambda))/(1 - \text{df}(\lambda)/n)^2 \\ \text{BIC}(\lambda) &= 2\ell^*(\hat{\beta}(\lambda))/n + \text{df}(\lambda)\log(n)/n \\ \text{AIC}(\lambda) &= 2\ell^*(\hat{\beta}(\lambda))/n + 2\text{df}(\lambda)/n \end{aligned} \quad (42)$$

其中

$$\text{df}(\lambda) = \text{trace} \left\{ \mathbf{V} \left\{ \mathbf{V}^T \mathbf{V} + nA(\hat{\beta}(\lambda), \lambda) \right\}^{-1} \mathbf{V}^T \right\}$$

我們選擇可以最小化不同挑選標準之 λ 。

4 數值模擬

我們比較有考慮左截斷時間與沒有考慮左截斷時間在 Cox 比例風險模式下進行模式選擇與參數估計的差異，並且比較不同的懲戒函數，包括 GLASSO，GSCAD，GMCP，AGLASSO。為了衡量估計的正確率，我們參考 (Tibshirani et al., 1997)，令 $MSE = (\hat{\beta}(\lambda) - \beta_{\text{true}})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}(\lambda) - \beta_{\text{true}})$ 的中位數作為衡量標準，其中 $\mathbf{X}^T \mathbf{X}$ 為共變數的樣本協方差矩陣 (Covariance Matrix)；此外我們也比較了不同挑選標準的結果，包含 5-fold 交叉驗正法、GCV、BIC 與 AIC。

設定生存時間與截斷時間皆服從 Cox 比例風險模式， $\lambda_i(t) = \lambda_0(t)e^{\mathbf{x}_i\beta}$ ，並令隱變數 $(\mathbf{Z}_1^*, \mathbf{Z}_2^*, \dots, \mathbf{Z}_{10}^*)$ 為平均值 0，且 \mathbf{Z}_i^* 與 \mathbf{Z}_j^* 之協方差為 $0.5^{|i-j|}$ 之高斯分佈， \mathbf{X}^* 為受測者的共變數，其中 \mathbf{X}_j^* , $j=1, \dots, 5$ 為離散型隨機變數，水準 0, 1, 2 分別為 \mathbf{Z}_j^* 小於 $\Phi(1/3)$ ，大於 $\Phi(2/3)$ 與中間， Φ 為 $N(0, 1)$ 的分位數， $\mathbf{X}_j^* = \mathbf{Z}_j^* I(|\mathbf{Z}_j^*| \leq 2) + 2I(|\mathbf{Z}_j^*| > 2)\text{sign}(\mathbf{Z}_j^*)$, $j=6, \dots, 10$ 。令 $\lambda_0(t) = 2t$ ，訂定

$$(i) \quad \text{存活時間 } \mathbf{T}^* = \sqrt{-\log(U_1)e^{-\mathbf{X}^*\beta_{\text{true}}}}, \quad U_1 \sim \text{Uniform}(0, 1)$$

$$\mathbf{x}_i^* \beta_{\text{true}} = 0.3I(\mathbf{x}_{i1}^* \neq 2) - 0.35I(\mathbf{x}_{i3}^* \neq 2) + 0.2\mathbf{x}_{i6}^*$$

$$(ii) \quad \text{左截斷時間 } \mathbf{W}^* = \sqrt{-\log(U_2)e^{-\mathbf{X}^*\beta_w}}, \quad U_2 \sim \text{Uniform}(0, 1)$$

$$\mathbf{x}_i^* \beta_w = -10\mathbf{x}_i^* \beta_{\text{true}} - 2.5[(\mathbf{x}_{i2}^* = 2) + (\mathbf{x}_{i4}^* = 2) + (\mathbf{x}_{i5}^* = 2)] - 1.5[\mathbf{x}_{i7}^* - \mathbf{x}_{i8}^* + \mathbf{x}_{i9}^* - \mathbf{x}_{i10}^*]$$

(iii) 右設限時間 $C^* = W + U$, $U \sim \text{Uniform}(0, c)$ 因為資料為左截斷且右設限，所以蒐集到的樣本為 (Y, X, W, Δ) 。在此設定下，資料截斷率 = 66%、資料設限率 (調整 c) = 10% 並且在樣本數: $N=400, 800, 1200$ 下重複抽樣 100 次。

數值模擬結果請參考附錄一，表 1 至表 4 分別為 CV、GCV，BIC，AIC 模式選擇之結果，左右欄位分別呈現在四種群懲戒函數之下，有無考慮左截斷時間正確的 0 個數與不正確的 0 個數；表 5 至表 8 分別為 CV、GCV，BIC，AIC

參數估計偏誤之結果，左右欄位分別呈現在四種群懲戒函數之下，有無考慮左截斷時間正確的參數估計偏誤 ($\hat{\beta} - \beta_{\text{true}}$)；表 9 至表 12 分別為 CV、GCV，BIC，AIC 推論估計量的變異數之結果，呈現在四種群懲戒函數之下，有考慮左截斷時間參數估計的樣本變異與參數估計變異數之估計值；由於參數估計應具有 Oracle 性質，也就是參數估計量應建進收斂至高斯分佈，因此定信賴區間為 $(\hat{\beta} \pm 1.96\hat{\sigma}_A SE)$ ，表 13 至表 16 分別為 CV、GCV，BIC，AIC 參數估計之信賴區間結果，左右欄位分別呈現在四種群懲戒函數之下，有無考慮左截斷時間，由參數估計與參數估計變異之估計值建構出的 95% 信賴區間覆蓋參數真值的覆蓋率。

由表 1 至表 4 看出如果沒有考慮左截斷時間的模式選擇，模式選擇的結果隨著樣本數增加而越來越差，相反地，有考慮左截斷時間的模式選擇，模式選擇的結果隨著樣本數增加而越來越好，因為沒有考慮左截斷時間的模式選擇，參數估計值的期望值會收斂至某個不同於參數真值 β_{true} 的錯誤參值 β_{wrong} ，又因為懲罰 log-部分概似函數擁有選擇一致性之性質 (selection consistency)，導致沒有考慮左截斷時間的模式選擇會一致地挑選在 β_{wrong} 之下的 Cox 比例風險模式；相反地，有考慮左截斷時間的模式選擇，可以一致地挑選在 β_{true} 之下的 Cox 比例風險模式，且隨著樣本數增加，模式選擇的表現越來越好。

由表 5 至表 8 可以看出，沒有考慮左截斷時間的模式選擇偏誤比有考慮左截斷時間的模式選擇大很多，而有考慮左截斷時間的模式選擇的偏誤卻隨著樣本數增加而越來越小，沒有考慮左截斷時間的模式選擇因為參數估計值的期望值隨著樣本數變大而收斂至 β_{wrong} ，所以當然不會是 β_{true} 的漸近不偏估計量，相反地從表 5 至表 8 可以看出考慮左截斷時間的模式選擇，參數估計看似有漸近不偏的性質。

由表 9 至表 12 可以看出，在章節 3.3 提出的變異數估計值是合理的，且隨著樣本數增加有越來越準的趨勢

表 13 至表 16 顯示我們建構的 95% 信賴區間，雖然我們由表 9 至表 12 可以看出變異數估計是合理的，但是由表 5 至表 8 看到有些挑選標準的偏誤偏大，造成偏誤越大，信賴區間覆蓋 β_{true} 的比例越低，偏誤越小，信賴區間覆蓋的覆蓋率隨著樣本數增加而越來越靠近 95%。

雖然我們估計的偏誤很大，但我們模式挑選的結果還算不錯，表 1 至表 4 顯示隨著樣本數越大，選到正確模式的期望次數越多，為了改進參數估計的結果，我們捨棄群懲戒函 log-部分概似函數的同時進行模式選擇與參數估計的優點，改為先進行模式選擇再進行參數估計，也就是將參數估計不為零的變數挑進模式裡，再將被挑進模式變數當重新計算一次最大概似估計量，結果請參考附錄二；表 17 至表 20 顯示參數估計的偏誤結果，表 21 至表 24 變異數估計的結果，表 25 至表 28 顯示由參數估計與變異數估計建構的信賴區間覆蓋參數真值的比例。

5 實例分析

5.1 女性乳癌資料

我們分析美國癌症登記及老人醫療保險之中罹患乳癌的病人與其資料；我們從此資料庫蒐集與記錄 1970 年之後被診斷出罹患乳癌的病人的資料，希望能藉由此資料庫挑選出對於診斷出罹患乳癌後的存活時間有效應的變數，並且估計有效應的變數對於存活時間的影響程度。然而此筆資料從 1986 才開始記錄病人因罹患乳癌而死亡的時間，使得此筆資料只有收入 1986 年後還活著的病人，而 1970 年至 1986 年間因乳癌死亡之病人的數據無法被蒐集到，因此定義 T^* 為診斷出罹患乳癌後之後的存活時間，是為我們有興趣之事件的存活時間，並定義左截斷時間 W^* 為病人被診斷出罹患乳癌的日期至 1986 年開始記錄死亡時間的日期，我們蒐集到的資料只能觀察到存活時間大於左截斷時間 W^* 的存活時間 T 與左截斷時間 W ，再加上資料實驗期間會有病人離開實驗或者其他因素導致無法觀察到診斷出罹患乳癌後的存活時間，也就是只能觀測到右設限時間，定義右設限時間為病人從 1980 年至離開實驗的時間 C ，則我們的資料有 $Y = \min(T, W + C)$, $\Delta = I(T \leq W + C)$ ， Y 代表診斷出罹患乳癌後的存活時間或者右設限時間，分別以 $\Delta = 1$ 或者 0 表示。

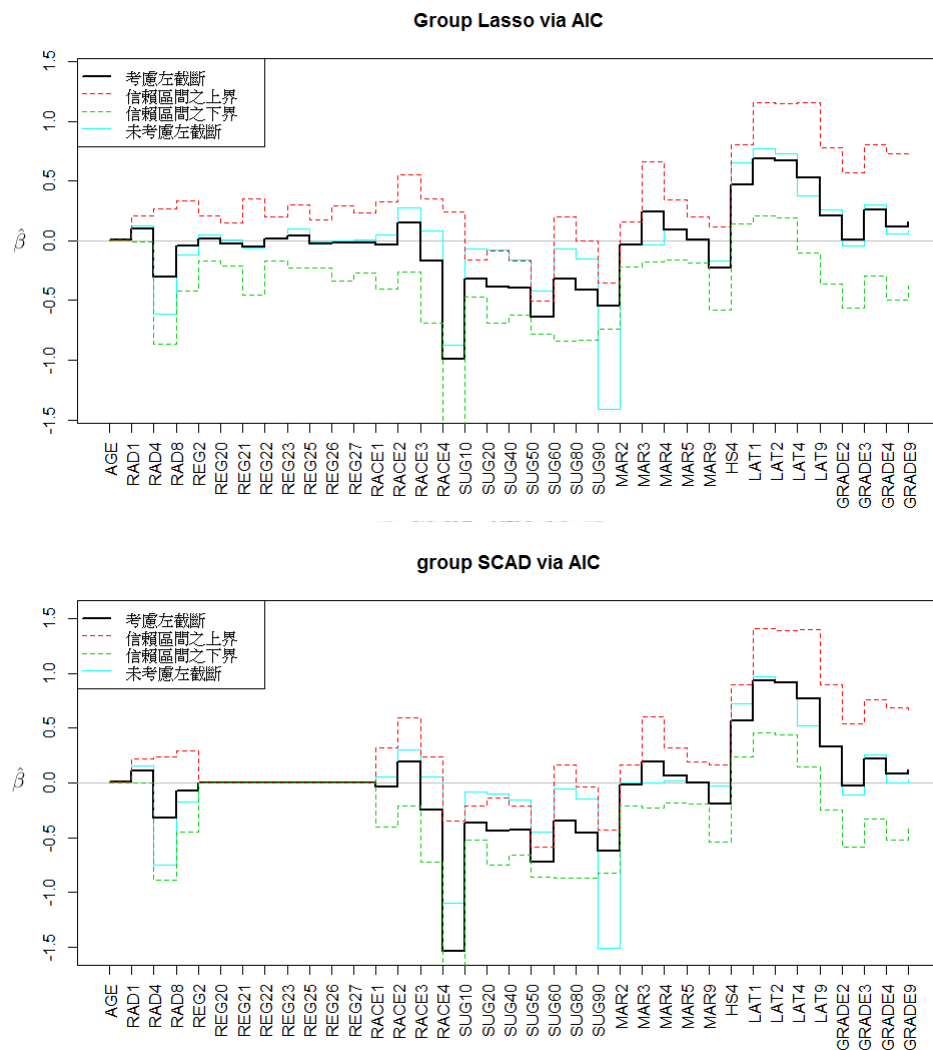
我們從此筆資料庫中挑選出乳癌第四期的病患進行分析，蒐集到 2297 筆樣本，因乳癌死亡的人數有 1539 人，剩下的 758 人包含非乳癌致死與未死亡的受測者，這些人觀察不到診斷出乳癌後的存活時間，只能觀察到他們的右設限時間，因此設限比率約 33%；根據 Rubin (1997) 的研究結果，與診斷出罹患乳癌後的存活時間混雜的潛在因素包含：年齡、種族、婚姻狀態、乳癌細胞偏側、腫瘤大小、癌細胞擴散情況與居住地區；再加上我們好奇放射性治療、手術方式是否對於乳癌的治療有影響，所以也把這兩項因素加入分析。然而有些變數的

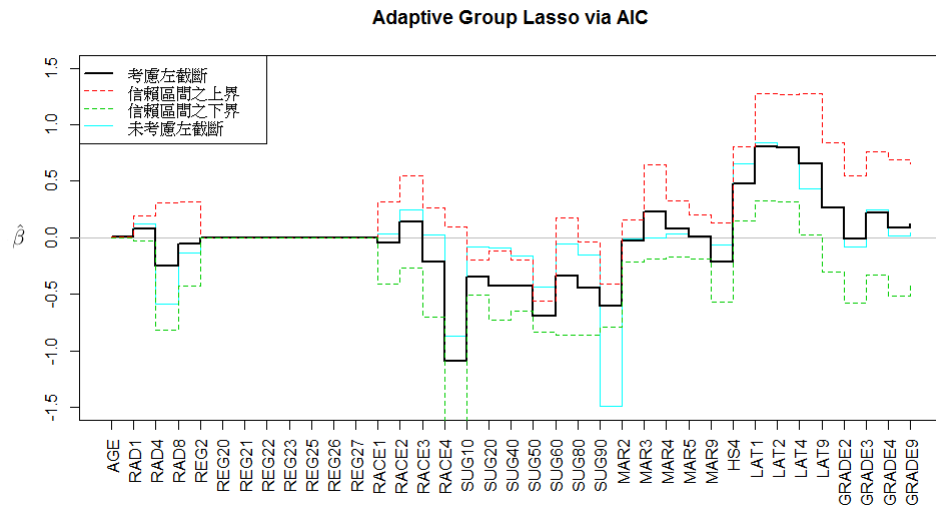
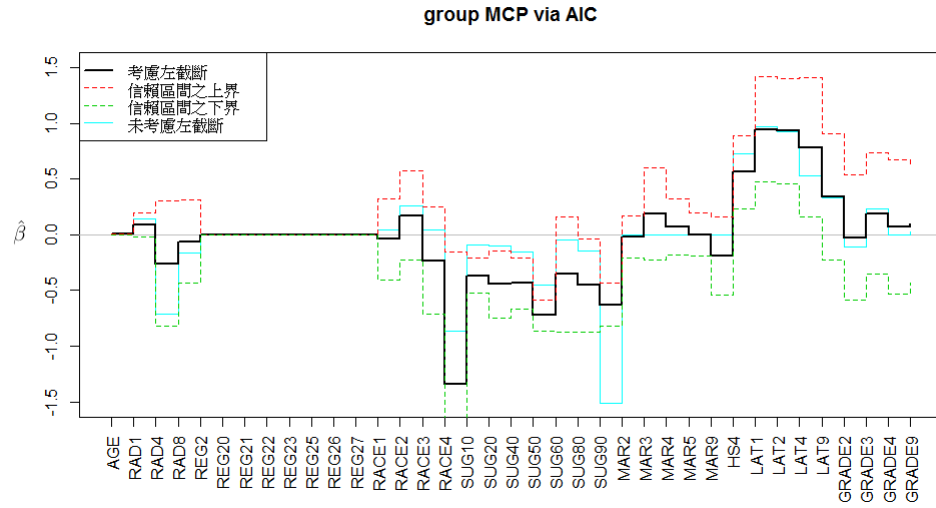
某些水準，樣本數太少，例如種族部分只有兩位西班牙人、腫瘤位置只有一邊的病患只有兩人，等等，如此會造成參數估計值變異太大使得模式選擇結果不穩定，或者演算法難以收斂，因此我們將樣本數小於 10 的水準從資料中剔除，最後剩餘 2265 筆樣本，變數有年齡 (視為連續型)、放射性治療 (RAD, 4 個水準)、居住地區 (REG, 9 個水準)、人種 (RACE, 5 個水準)、手術方式 (SUG, 8 個水準)、婚姻狀態 (MAR, 6 個水準)、癌細胞擴散情況 (HS, 2 個水準)、乳癌細胞偏側情況 (LAT, 5 個水準)、腫瘤大小 (GRADE, 5 個水準) 等九個變數。

因為數值模擬結果看出 AIC 參數估計的偏誤較小，所以我們顯示使用 GLASSO、GSCAD、GMCP、AGLASSO 四種不同懲戒函數與 AIC 挑選標準的模式選擇與參數估計結果，由於我們將有被挑入模式的變數重新再計算一次最大概似估計量，數值沒有太大差異，所以圖表呈現的是同時進行模式選擇與參數估計的結果。橫軸為變數的名稱，縱軸為參數估計值，粗黑線顯示有考慮左截斷時間的參數估計值，若參數估計值等於零，代表沒有被選入模式，也就是對於診斷出罹患乳癌後的存活時間沒有效應；紅色虛線與綠色虛線分別為考慮左截斷時間之下，由方程式 (39) 與方程式 (40) 估計之標準差建構的 95% 信賴區間的上下界，藍色細線則顯示未考慮左截斷時間之下的模式選擇與參數估計的結果；可以看出有無考慮左截斷時間對於參數估計是有相當影像的，而且模式挑選結果也有差，例如 GMCP 的模式選擇結果，考慮左截斷時間之下 MAR 有被選入模式中，但沒有考慮左截斷時間的話 MAR 就有被選入模式之中。離散型變數的水準請參考附錄三表 29 至 36

於我們使用四種懲戒函數，只有 Group Lasso 的模式挑選結果不一樣，而事實上 Group Lasso 已經被證實會多挑選到事實上沒有效應的變數。其餘三個懲戒函數挑選模式的結果皆一致，除了 REG(居住地區) 以外的變數皆有挑選到，

所以我們認為住在哪個不同居住地區的人對於罹患乳癌後的存活時間是沒有效應的變數。人種的部分，RACE4(亞洲人) 的參數估計值比其他水準都來的低，且信賴區間沒有覆蓋 0，因此我們認為亞洲人因罹患乳癌後的存活時間較其他人種長(死亡風險低)。由於 LAT(源發部位腫瘤偏側) 的基準水準是非成對器官，且所有水準的參數估計值皆為正數，我們認為如果源發部位是成對器官，存活時間較短’(風險較高)。





6 結論

本文主要研究如何在 Cox 比例風險模式之下分析盛行倖存資料，同時進行模式選擇與參數估計。傳統倖存資料只考慮到設限的問題，而盛行倖存資料除了要考慮資料右設限的問題，還需要考慮到資料左截斷的問題，以免得到錯誤的參數估計結果。在 Cox 比例風險模式之下進行模式選擇且考慮資料左截斷造成樣本偏誤的問題，我們參考 Tibshirani et al. (1997) 提出的懲戒 log-部分概似函數，我們將目標函數改成 Wang et al. (1993) 提出修正過後的 log-部分概似函數 (方程式 (3))，並且參考 (Zou, 2008) 得到近似結果；為了處理離散型的變數，我們也參考 (Yuan and Lin, 2006; Wang et al., 2007; Wang and Leng, 2008) 的群懲戒函數，得到我們的群懲戒 log-部分概似函數 (方程式 (21))；演算法部分我們主要參考 (Yuan and Lin, 2006; Breheny and Huang, 2015)，並以 Fan and Li (2002) 提出的 LQA、Zou and Li (2008) 提出的 LLA 近似群懲戒函數，推出演算法 3.1 與演算法 3.2；參數估計的變異根據 Lu and Zhang (2007)，可以藉由 LQA 逼近計算；調整參數的挑選我們考慮 CV、GCV、BIC、AIC 等四種挑選標準，然而這四種挑選標準在 Cox 比例風險模式之下的理論性值並不完備，我們認為這是使數值模擬的模式挑選結果的效率較差且偏誤較大的主要原因，改進挑選標準也是我們未來將會努力的方向之一。

7 附錄

7.1 附錄一

Table 1: 5-fold CV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；Corr 代表正確的 0 個數，Incorr 代表不正確的 0 個數

n	method	Corr.(10)	Incorr.(0)	MSE	Corr.(10)	Incorr.(0)	MSE
400	MLE	0	0	0.05(0.02)	0	0	0.085(0.03)
	GLASSO	7.06	1.56	0.029(0.013)	6.14	3.18	0.049(0.014)
	GSCAD	7.67	1.95	0.036(0.016)	6.42	3.57	0.059(0.03)
	GMCP	8.37	2.42	0.037(0.018)	7.1	3.88	0.058(0.03)
	AGLASSO	8.35	2.05	0.033(0.015)	7.07	4.14	0.053(0.02)
800	MLE	0	0	0.022(0.01)	0	0	0.059(0.016)
	GLASSO	5	0.25	0.015(0.009)	2.24	1.44	0.044(0.01)
	GSCAD	6.57	0.33	0.017(0.012)	3.25	2.32	0.054(0.016)
	GMCP	7.09	0.41	0.017(0.012)	3.07	2.34	0.055(0.016)
	AGLASSO	7.52	0.42	0.015(0.011)	3.05	2.4	0.052(0.013)
1200	MLE	0	0	0.015(0.005)	0	0	0.05(0.015)
	GLASSO	4.88	0.04	0.011(0.005)	0.8	0.63	0.041(0.01)
	GSCAD	6.69	0.08	0.01(0.008)	1.43	1.43	0.051(0.013)
	GMCP	7.49	0.18	0.01(0.007)	1.52	1.58	0.05(0.013)
	AGLASSO	7.48	0.14	0.01(0.006)	1.01	1.41	0.047(0.012)

Table 2: GCV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；Corr 代表正確的 0 個數，Incorr 代表不正確的 0 個數

n	method	Corr.(10)	Incorr.(0)	MSE	Corr.(10)	Incorr.(0)	MSE
400	MLE	0	0	0.05(0.02)	0	0	0.085(0.03)
	GLASSO	9.94	4.77	0.044(0.006)	9.85	4.94	0.046(0.004)
	GSCAD	9.94	4.77	0.044(0.006)	9.85	4.94	0.046(0.005)
	GMCP	9.96	4.79	0.046(0.005)	9.95	4.96	0.048(0.009)
	AGLASSO	9.98	4.76	0.044(0.006)	9.85	4.95	0.046(0.004)
800	MLE	0	0	0.022(0.01)	0	0	0.059(0.016)
	GLASSO	9.96	4.2	0.042(0.007)	9.93	4.96	0.047(0.002)
	GSCAD	9.97	4.22	0.042(0.007)	9.93	4.96	0.047(0.002)
	GMCP	9.99	4.28	0.041(0.007)	9.93	4.97	0.047(0.002)
	AGLASSO	10	4.29	0.042(0.007)	9.93	4.97	0.047(0.002)
1200	MLE	0	0	0.015(0.005)	0	0	0.05(0.015)
	GLASSO	9.97	3.46	0.038(0.007)	9.76	4.95	0.046(0.002)
	GSCAD	9.95	3.38	0.038(0.007)	9.76	4.95	0.046(0.002)
	GMCP	9.98	3.58	0.037(0.007)	9.76	4.95	0.046(0.003)
	AGLASSO	10	3.62	0.038(0.007)	9.76	4.95	0.046(0.002)

Table 3: BIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；Corr 代表正確的 0 個數，Incorr 代表不正確的 0 個數

n	method	Corr.(10)	Incorr.(0)	MSE	Corr.(10)	Incorr.(0)	MSE
400	MLE	0	0	0.05(0.02)	0	0	0.085(0.03)
	GLASSO	8.61	2.2	0.032(0.01)	7.6	3.96	0.045(0.008)
	GSCAD	8.82	2.19	0.032(0.011)	7.64	3.97	0.047(0.012)
	GMCP	9.65	3.62	0.041(0.012)	9.02	4.84	0.058(0.022)
	AGLASSO	9.28	2.41	0.032(0.011)	8	4.36	0.047(0.01)
800	MLE	0	0	0.022(0.01)	0	0	0.059(0.016)
	GLASSO	8.41	0.88	0.022(0.01)	6.24	3.58	0.043(0.005)
	GSCAD	8.95	0.88	0.021(0.011)	6.11	3.59	0.044(0.007)
	GMCP	9.39	1.08	0.02(0.011)	6.55	3.86	0.044(0.007)
	AGLASSO	9.4	1.06	0.02(0.011)	6.47	4.17	0.045(0.006)
1200	MLE	0	0	0.015(0.005)	0	0	0.05(0.015)
	GLASSO	7.85	0.4	0.016(0.008)	5.1	3.05	0.04(0.006)
	GSCAD	8.89	0.4	0.014(0.008)	5.01	3.15	0.042(0.008)
	GMCP	9.39	0.48	0.014(0.008)	5.48	3.44	0.042(0.008)
	AGLASSO	9.48	0.44	0.013(0.008)	5.27	3.98	0.043(0.007)

Table 4: AIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；Corr 代表正確的 0 個數，Incorr 代表不正確的 0 個數

n	method	Corr.(10)	Incorr.(0)	MSE	Corr.(10)	Incorr.(0)	MSE
400	MLE	0	0	0.05(0.02)	0	0	0.085(0.03)
	GLASSO	4.47	0.46	0.025(0.011)	2.97	1.61	0.05(0.017)
	GSCAD	5.85	0.51	0.03(0.016)	3.44	2.25	0.063(0.025)
	GMCP	7.29	1.15	0.041(0.02)	4.81	3.09	0.078(0.03)
	AGLASSO	6.15	0.57	0.027(0.014)	3.63	2.78	0.058(0.021)
800	MLE	0	0	0.022(0.01)	0	0	0.059(0.016)
	GLASSO	3.74	0	0.014(0.007)	1.27	0.9	0.043(0.011)
	GSCAD	6.09	0.02	0.013(0.01)	1.79	1.85	0.053(0.015)
	GMCP	6.74	0.04	0.013(0.009)	2	2.22	0.052(0.015)
	AGLASSO	6.25	0.04	0.013(0.008)	1.85	1.88	0.048(0.012)
1200	MLE	0	0	0.015(0.005)	0	0	0.05(0.015)
	GLASSO	3.77	0	0.01(0.005)	0.61	0.57	0.04(0.01)
	GSCAD	6.13	0	0.009(0.005)	1.01	1.77	0.05(0.013)
	GMCP	6.83	0	0.009(0.005)	1.23	2.11	0.049(0.012)
	AGLASSO	6.34	0	0.009(0.005)	0.98	1.76	0.045(0.011)

Table 5: 5-fold CV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1} BIAS	β_{1_2} BIAS	β_{3_1} BIAS	β_{3_2} BIAS	β_6 BIAS	β_{1_1} BIAS	β_{1_2} BIAS	β_{3_1} BIAS	β_{3_2} BIAS	β_6 BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-183	-175	223	215	-104	-264	-256	319	310	-177
	GSCAD	-164	-154	207	197	-79	-272	-267	319	322	-181
	GMCP	-182	-163	221	209	-89	-274	-273	328	325	-182
	AGLASSO	-168	-158	211	202	-95	-274	-266	333	328	-183
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-108	-95	145	150	-66	-229	-231	293	295	-167
	GSCAD	-63	-47	103	106	-28	-238	-241	298	301	-170
	GMCP	-69	-52	106	112	-29	-237	-239	296	302	-164
	AGLASSO	-91	-74	124	132	-52	-243	-243	298	302	-168
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-87	-87	128	128	-49	-218	-225	287	292	-159
	GSCAD	-32	-31	79	80	-11	-222	-235	295	303	-154
	GMCP	-40	-38	84	83	-11	-225	-228	294	302	-153
	AGLASSO	-61	-61	105	105	-31	-218	-224	287	290	-155

Table 6: GCV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
		BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-298	-298	350	350	-182	-299	-300	350	350	-198
	GSCAD	-298	-298	350	350	-182	-299	-300	350	350	-198
	GMCP	-300	-300	350	350	-181	-300	-300	350	350	-198
	AGLASSO	-297	-297	350	350	-182	-299	-300	350	350	-199
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-292	-290	350	350	-171	-300	-300	350	350	-200
	GSCAD	-292	-290	350	350	-170	-300	-300	350	350	-200
	GMCP	-292	-290	350	350	-167	-300	-300	350	350	-200
	AGLASSO	-291	-289	350	350	-170	-300	-300	350	350	-200
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-284	-284	347	347	-154	-300	-300	350	350	-198
	GSCAD	-283	-283	346	346	-151	-300	-300	350	350	-198
	GMCP	-285	-285	347	347	-145	-300	-300	350	350	-198
	AGLASSO	-284	-284	347	348	-151	-300	-300	350	350	-198

Table 7: BIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
		BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-235	-226	296	294	-131	-286	-284	341	339	-191
	GSCAD	-227	-217	288	284	-113	-285	-283	341	339	-190
	GMCP	-240	-228	306	299	-120	-296	-298	350	350	-191
	AGLASSO	-215	-204	281	276	-119	-289	-287	341	339	-191
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-192	-180	256	258	-106	-278	-277	341	339	-191
	GSCAD	-173	-160	245	246	-75	-277	-275	342	340	-190
	GMCP	-168	-154	244	245	-75	-276	-274	343	341	-189
	AGLASSO	-162	-148	236	239	-89	-284	-281	343	341	-192
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-157	-155	221	225	-83	-266	-269	340	341	-186
	GSCAD	-125	-125	209	211	-43	-268	-271	340	341	-184
	GMCP	-123	-122	205	208	-45	-266	-270	341	342	-184
	AGLASSO	-120	-118	195	199	-61	-277	-281	344	344	-187

Table 8: AIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
		BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-128	-113	157	148	-69	-241	-233	298	288	-170
	GSCAD	-84	-66	121	107	-23	-244	-236	307	297	-166
	GMCP	-80	-64	115	105	-28	-242	-231	305	295	-168
	AGLASSO	-96	-81	123	113	-47	-247	-238	304	293	-171
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-92	-79	122	128	-56	-225	-223	286	289	-166
	GSCAD	-34	-19	65	70	-13	-238	-237	305	306	-167
	GMCP	-36	-21	68	72	-14	-236	-237	304	304	-166
	AGLASSO	-55	-40	81	87	-33	-235	-234	296	299	-170
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-74	-73	110	111	-42	-218	-225	291	294	-161
	GSCAD	-18	-17	50	50	-5	-236	-244	319	319	-162
	GMCP	-20	-19	53	53	-6	-234	-242	318	318	-161
	AGLASSO	-40	-39	71	71	-20	-228	-236	310	311	-164

Table 9: 5-fold CV 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；此表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	112	97(64)	122	100(66)	151	106(85)	163	110(89)	75	61(29)
	GSCAD	167	86(68)	177	88(70)	212	88(87)	232	92(91)	111	58(31)
	GMCP	155	73(69)	172	75(71)	190	72(85)	219	74(88)	107	48(34)
	AGLASSO	141	84(67)	157	86(69)	180	88(87)	195	91(90)	98	51(33)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	90	93(19)	105	95(20)	123	112(33)	125	116(35)	54	52(6)
	GSCAD	131	91(21)	140	94(22)	170	107(38)	168	111(40)	70	51(6)
	GMCP	122	90(23)	132	92(24)	157	103(42)	161	107(44)	66	50(6)
	AGLASSO	108	91(21)	122	93(22)	136	101(43)	135	105(45)	62	49(8)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	82	78(8)	75	79(8)	105	97(10)	106	101(11)	46	43(3)
	GSCAD	103	77(8)	99	78(8)	141	94(17)	141	98(18)	51	41(2)
	GMCP	103	75(13)	99	76(14)	136	91(23)	131	94(24)	48	40(3)
	AGLASSO	93	76(11)	88	77(11)	122	92(21)	120	95(22)	48	40(2)

Table 10: GGV 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；
此表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	20	1(13)	15	1(14)	0	0(0)	0	0(0)	45	15(30)
	GSCAD	20	1(13)	15	1(14)	0	0(0)	0	0(0)	47	15(30)
	GMCP	0	0(0)	0	0(0)	0	0(0)	0	0(0)	48	15(30)
	AGLASSO	21	3(20)	18	3(20)	0	0(0)	0	0(0)	49	14(29)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	29	13(33)	32	13(34)	0	4(22)	0	4(23)	39	24(25)
	GSCAD	31	13(33)	35	13(34)	0	2(18)	0	3(19)	42	24(25)
	GMCP	32	12(32)	35	12(33)	0	0(0)	0	0(0)	45	24(25)
	AGLASSO	37	12(32)	41	12(33)	0	1(12)	0	1(13)	44	23(25)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	36	25(37)	37	25(37)	19	5(22)	17	5(23)	38	34(16)
	GSCAD	38	28(38)	38	28(39)	21	5(22)	19	5(23)	44	34(16)
	GMCP	37	20(35)	38	20(35)	17	5(22)	15	5(23)	47	33(16)
	AGLASSO	40	21(35)	41	22(36)	15	4(20)	12	4(20)	46	31(18)

Table 11: BIC 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；此
表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	85	83(68)	97	85(70)	99	73(86)	109	75(89)	64	56(32)
	GSCAD	100	82(69)	113	84(70)	115	74(86)	129	77(89)	88	56(31)
	GMCP	122	31(57)	141	31(58)	133	18(52)	154	19(54)	90	50(33)
	AGLASSO	109	75(69)	123	77(70)	129	65(83)	141	67(87)	81	52(32)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	80	83(33)	87	86(33)	98	86(55)	97	89(57)	49	48(10)
	GSCAD	102	83(32)	113	85(33)	114	86(55)	116	89(57)	71	47(10)
	GMCP	102	78(37)	113	80(38)	120	79(57)	122	82(59)	67	46(10)
	AGLASSO	103	80(34)	114	83(35)	126	77(58)	125	79(60)	61	47(10)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	76	74(17)	73	75(17)	96	83(35)	95	86(36)	43	41(2)
	GSCAD	96	74(17)	92	75(17)	118	83(35)	115	86(36)	57	39(2)
	GMCP	96	73(19)	92	74(19)	121	80(38)	117	82(39)	54	39(2)
	AGLASSO	93	73(17)	90	74(17)	122	80(37)	119	83(38)	49	39(2)

Table 12: AIC 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；此表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	118	123(46)	132	126(47)	161	158(54)	174	164(56)	72	75(12)
	GSCAD	158	123(44)	175	127(45)	209	151(60)	228	157(62)	91	72(14)
	GMCP	162	107(56)	176	110(57)	218	113(82)	232	116(84)	93	68(17)
	AGLASSO	138	123(44)	152	126(45)	186	147(63)	202	153(65)	84	70(17)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	86	98(2)	97	100(3)	106	122(5)	112	127(5)	51	54(3)
	GSCAD	110	96(2)	122	98(3)	143	119(13)	148	124(13)	54	52(4)
	GMCP	108	95(2)	121	98(3)	139	118(18)	145	122(18)	55	51(4)
	AGLASSO	98	96(2)	110	98(3)	122	118(18)	131	122(18)	56	51(3)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	80	79(2)	74	80(2)	97	99(3)	97	102(3)	45	44(2)
	GSCAD	94	78(2)	87	79(2)	122	97(3)	122	101(3)	45	42(3)
	GMCP	94	78(2)	88	79(2)	121	97(3)	120	100(3)	45	41(3)
	AGLASSO	88	78(2)	82	79(2)	108	97(3)	108	101(3)	47	42(3)

Table 13: 5-fold CV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	67%	70%	59%	60%	69%	29%	36%	26%	26%	19%
	GSCAD	57%	57%	47%	48%	56%	22%	24%	20%	23%	11%
	GMCP	50%	51%	39%	39%	56%	18%	21%	15%	15%	13%
	AGLASSO	57%	59%	51%	50%	59%	15%	19%	12%	12%	13%
800	MLE	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	81%	84%	77%	73%	71%	31%	29%	25%	26%	14%
	GSCAD	79%	76%	70%	72%	86%	29%	22%	21%	22%	16%
	GMCP	83%	79%	76%	76%	84%	30%	26%	21%	22%	18%
	AGLASSO	81%	80%	80%	81%	76%	26%	25%	22%	24%	14%
1200	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	77%	83%	74%	77%	79%	22%	18%	13%	11%	5%
	GSCAD	86%	86%	73%	74%	88%	26%	21%	16%	13%	13%
	GMCP	84%	87%	77%	79%	88%	22%	21%	16%	16%	12%
	AGLASSO	80%	84%	77%	79%	84%	26%	22%	17%	15%	8%

Table 14: GCV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	1%	1%	0%	0%	13%	1%	1%	0%	0%	2%
	GSCAD	1%	1%	0%	0%	13%	1%	1%	0%	0%	2%
	GMCP	0%	0%	0%	0%	13%	0%	0%	0%	0%	2%
	AGLASSO	2%	2%	0%	0%	12%	1%	1%	0%	0%	1%
800	MLE	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	2%	4%	0%	0%	7%	0%	0%	0%	0%	0%
	GSCAD	2%	4%	0%	0%	9%	0%	0%	0%	0%	0%
	GMCP	2%	3%	0%	0%	9%	0%	0%	0%	0%	0%
	AGLASSO	2%	5%	0%	0%	8%	0%	0%	0%	0%	0%
1200	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	1%	3%	0%	0%	3%	0%	0%	0%	0%	0%
	GSCAD	2%	3%	1%	0%	8%	0%	0%	0%	0%	0%
	GMCP	2%	3%	0%	0%	7%	0%	0%	0%	0%	0%
	AGLASSO	3%	3%	0%	0%	6%	0%	0%	0%	0%	0%

Table 15: BIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	50%	53%	36%	39%	53%	18%	19%	13%	14%	7%
	GSCAD	52%	53%	38%	41%	52%	18%	19%	14%	14%	7%
	GMCP	23%	23%	11%	9%	46%	1%	1%	0%	0%	5%
	AGLASSO	51%	53%	35%	37%	53%	11%	14%	8%	9%	6%
800	MLE	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	45%	57%	37%	39%	43%	6%	8%	1%	4%	0%
	GSCAD	50%	57%	36%	40%	62%	6%	8%	1%	3%	2%
	GMCP	54%	60%	37%	41%	63%	8%	10%	2%	3%	3%
	AGLASSO	58%	65%	40%	41%	57%	6%	7%	2%	3%	1%
1200	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	53%	52%	31%	33%	48%	2%	1%	0%	0%	1%
	GSCAD	61%	66%	30%	33%	76%	2%	1%	0%	1%	3%
	GMCP	64%	68%	37%	37%	77%	4%	4%	1%	1%	3%
	AGLASSO	64%	70%	46%	44%	66%	2%	2%	0%	1%	1%

Table 16: AIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	88%	87%	86%	87%	90%	54%	61%	50%	55%	23%
	GSCAD	83%	89%	80%	83%	86%	37%	46%	43%	46%	21%
	GMCP	76%	78%	62%	62%	85%	33%	39%	28%	29%	21%
	AGLASSO	87%	88%	80%	80%	88%	34%	40%	36%	40%	23%
800	MLE	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	87%	89%	85%	83%	83%	30%	28%	25%	28%	11%
	GSCAD	90%	85%	85%	84%	93%	24%	21%	13%	15%	14%
	GMCP	89%	85%	86%	84%	94%	28%	23%	16%	19%	12%
	AGLASSO	92%	90%	88%	86%	85%	29%	22%	20%	22%	11%
1200	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	82%	86%	81%	83%	85%	22%	17%	10%	9%	5%
	GSCAD	90%	90%	82%	88%	91%	19%	13%	7%	5%	13%
	GMCP	91%	90%	83%	87%	88%	19%	14%	5%	5%	13%
	AGLASSO	88%	90%	86%	87%	86%	22%	17%	8%	8%	5%

7.2 附錄二

Table 17: 5-fold CV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
		BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-72	-56	94	78	-36	-234	-217	289	276	-161
	GSCAD	-92	-74	131	119	-35	-244	-235	299	288	-170
	GMCP	-124	-97	160	142	-52	-252	-248	309	303	-174
	AGLASSO	-87	-80	126	113	-47	-262	-243	315	311	-171
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-16	-4	26	30	-17	-203	-205	263	265	-152
	GSCAD	-16	-2	26	33	-9	-213	-216	278	277	-162
	GMCP	-19	-6	36	42	-11	-219	-217	276	284	-159
	AGLASSO	-19	-4	35	42	-14	-230	-229	283	290	-160
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-11	-11	26	18	-10	-201	-209	265	272	-146
	GSCAD	-9	-9	26	23	-3	-208	-218	282	289	-149
	GMCP	-13	-12	32	25	-4	-207	-216	288	296	-150
	AGLASSO	-12	-10	30	26	-5	-208	-213	278	282	-148

Table 18: GCV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
		BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-294	-295	350	350	-151	-294	-296	350	350	-191
	GSCAD	-294	-295	350	350	-151	-294	-296	350	350	-191
	GMCP	-300	-300	350	350	-150	-300	-300	350	350	-191
	AGLASSO	-290	-289	350	350	-151	-294	-296	350	350	-194
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-253	-248	335	337	-111	-297	-296	350	350	-197
	GSCAD	-253	-248	340	340	-111	-297	-296	350	350	-197
	GMCP	-258	-251	350	350	-111	-297	-296	350	350	-199
	AGLASSO	-257	-252	345	345	-116	-297	-296	350	350	-199
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-197	-203	325	326	-61	-300	-300	350	350	-193
	GSCAD	-187	-193	325	326	-60	-300	-300	350	350	-193
	GMCP	-216	-220	326	327	-63	-300	-300	350	350	-193
	AGLASSO	-210	-216	331	332	-70	-300	-300	350	350	-193

Table 19: BIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
		BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-99	-82	166	157	-48	-241	-231	314	298	-172
	GSCAD	-101	-85	161	152	-46	-243	-229	308	295	-173
	GMCP	-211	-195	288	278	-64	-294	-296	350	350	-178
	AGLASSO	-110	-94	178	169	-47	-261	-250	325	315	-181
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-39	-22	83	78	-27	-227	-222	304	304	-172
	GSCAD	-39	-19	81	76	-20	-224	-223	306	307	-172
	GMCP	-50	-29	95	92	-20	-231	-231	318	314	-175
	AGLASSO	-44	-24	99	97	-22	-250	-250	322	320	-181
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-22	-22	54	52	-15	-203	-211	308	311	-161
	GSCAD	-20	-20	52	49	-8	-206	-215	311	314	-162
	GMCP	-20	-21	58	57	-8	-213	-222	317	319	-166
	AGLASSO	-17	-18	55	53	-7	-239	-251	330	330	-172

Table 20: AIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表呈現參數估計的偏誤值 ($\times 1000$)

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
		BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS	BIAS
400	MLE	-3	11	-3	-16	5	-209	-206	260	249	-144
	GLASSO	-23	-8	17	1	-5	-199	-193	249	237	-146
	GSCAD	-21	-3	18	0	0	-212	-204	265	255	-153
	GMCP	-43	-23	63	47	-9	-225	-207	286	273	-156
	AGLASSO	-22	-3	23	7	-5	-220	-207	275	260	-158
800	MLE	-6	8	3	15	-7	-3	11	-3	-16	5
	GLASSO	-8	5	6	16	-11	-196	-196	249	261	-148
	GSCAD	-6	9	4	16	-6	-203	-203	258	264	-153
	GMCP	-6	7	8	16	-7	-207	-206	265	271	-154
	AGLASSO	-8	6	7	19	-9	-205	-205	256	265	-153
1200	MLE	-5	-5	18	13	-4	-204	-211	268	272	-144
	GLASSO	-9	-8	22	17	-7	-201	-209	265	271	-145
	GSCAD	-6	-6	19	15	-3	-204	-211	284	288	-149
	GMCP	-6	-6	19	17	-2	-206	-215	289	292	-151
	AGLASSO	-7	-6	18	16	-4	-204	-213	282	287	-150

Table 21: 5-fold CV 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；此表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	182	145(5)	194	148(6)	240	183(12)	262	190(11)	102	82(4)
	GSCAD	193	145(5)	203	149(7)	251	185(13)	274	191(12)	110	82(4)
	GMCP	191	145(5)	209	149(7)	245	186(13)	273	192(12)	118	82(4)
	AGLASSO	191	145(5)	205	149(7)	244	185(12)	267	191(11)	115	82(4)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	108	100(3)	115	102(3)	139	125(5)	146	130(5)	59	56(2)
	GSCAD	112	100(2)	121	102(3)	151	125(5)	155	130(5)	58	56(2)
	GMCP	115	100(3)	123	102(3)	156	126(6)	165	130(5)	58	56(2)
	AGLASSO	110	100(2)	121	102(3)	159	126(6)	161	130(6)	60	56(2)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	88	81(2)	81	82(2)	102	101(3)	108	105(3)	50	46(1)
	GSCAD	89	81(2)	82	82(2)	112	101(3)	116	105(3)	49	46(1)
	GMCP	94	81(2)	90	82(2)	123	101(3)	119	105(3)	48	46(1)
	AGLASSO	90	81(2)	86	82(2)	119	101(3)	119	105(3)	47	46(1)

Table 22: GGV 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；
此表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	60	146(6)	47	151(7)	0	192(13)	0	198(12)	99	82(4)
	GSCAD	60	146(6)	47	151(7)	0	192(13)	0	198(12)	99	82(4)
	GMCP	0	146(6)	0	151(7)	0	192(13)	0	198(12)	100	82(4)
	AGLASSO	72	146(6)	81	151(7)	0	192(13)	0	198(12)	102	82(4)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	126	100(3)	138	104(3)	84	133(6)	77	138(6)	95	56(2)
	GSCAD	126	100(3)	138	104(3)	71	133(6)	72	138(6)	95	56(2)
	GMCP	119	100(3)	137	104(3)	0	134(6)	0	138(6)	95	56(2)
	AGLASSO	122	100(3)	134	104(3)	53	133(6)	45	138(6)	96	56(2)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	157	82(2)	150	83(2)	109	108(4)	105	112(4)	73	46(1)
	GSCAD	159	82(2)	153	83(2)	109	108(4)	105	112(4)	74	46(1)
	GMCP	148	82(2)	144	83(2)	107	108(4)	103	112(4)	71	46(1)
	AGLASSO	151	82(2)	144	83(2)	92	108(4)	91	112(5)	80	46(1)

Table 23: BIC 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；此
表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	191	145(5)	200	148(7)	239	186(14)	260	192(13)	107	82(4)
	GSCAD	192	145(5)	199	148(7)	240	186(14)	259	192(12)	106	82(4)
	GMCP	173	146(6)	197	150(7)	182	191(13)	211	197(12)	106	82(4)
	AGLASSO	194	145(5)	204	149(7)	240	187(14)	261	193(12)	110	82(4)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	127	100(3)	134	102(3)	188	127(6)	194	131(6)	61	56(2)
	GSCAD	128	100(3)	138	102(3)	188	127(6)	195	132(6)	65	56(2)
	GMCP	139	100(3)	149	102(3)	197	127(6)	202	132(6)	63	56(2)
	AGLASSO	133	100(3)	141	102(3)	201	127(6)	206	132(6)	61	56(2)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	100	81(2)	95	82(2)	148	101(3)	151	105(3)	46	46(1)
	GSCAD	101	81(2)	95	82(2)	148	101(3)	150	105(4)	48	46(1)
	GMCP	104	81(2)	99	82(2)	157	102(4)	157	105(4)	46	46(1)
	AGLASSO	100	81(2)	95	82(2)	154	101(4)	156	105(4)	44	46(1)

Table 24: AIC 挑選標準；比較參數估計的樣本標準差與標準差估計的差異；此表呈現參數估計的樣本變異與變異的估計值 ($\times 1000$)

n	method	β_{1_1}		β_{1_2}		β_{3_1}		β_{3_2}		β_6	
		SE	ASE	SE	ASE	SE	ASE	SE	ASE	SE	ASE
400	MLE	126	133(4)	144	137(5)	172	167(8)	198	172(8)	69	67(3)
	GLASSO	154	144(5)	168	148(6)	199	180(10)	222	187(9)	85	82(4)
	GSCAD	154	144(5)	168	148(6)	208	180(10)	222	187(9)	87	82(4)
	GMCP	170	145(5)	183	148(6)	239	182(11)	255	189(11)	86	82(4)
	AGLASSO	149	144(5)	168	148(6)	211	181(10)	230	187(9)	88	82(4)
800	MLE	84	93(2)	101	96(2)	106	117(4)	116	121(4)	43	46(1)
	GLASSO	95	100(2)	104	102(3)	110	125(5)	123	129(4)	58	56(2)
	GSCAD	96	100(2)	106	102(3)	115	125(5)	123	129(4)	56	56(2)
	GMCP	96	100(2)	107	102(3)	118	125(5)	126	129(5)	56	56(2)
	AGLASSO	96	100(2)	105	102(3)	117	125(5)	128	129(5)	56	56(2)
1200	MLE	80	76(1)	74	77(1)	98	95(3)	98	98(3)	39	38(1)
	GLASSO	87	81(2)	76	82(2)	96	101(3)	106	105(3)	50	46(1)
	GSCAD	85	81(2)	79	82(2)	99	101(3)	104	105(3)	50	46(1)
	GMCP	85	81(2)	79	82(2)	99	101(3)	103	105(3)	49	46(1)
	AGLASSO	84	81(2)	79	82(2)	98	101(3)	103	105(3)	49	46(1)

Table 25: 5-fold CV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	68%	68%	58%	58%	81%	34%	40%	24%	26%	29%
	GSCAD	60%	60%	46%	46%	75%	26%	30%	20%	22%	22%
	GMCP	51%	53%	40%	38%	64%	20%	24%	15%	15%	18%
	AGLASSO	59%	60%	49%	48%	70%	16%	20%	12%	12%	18%
800	MLE	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	GLASSO	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	91%	90%	92%	89%	93%	42%	45%	40%	39%	21%
	GSCAD	89%	87%	89%	87%	96%	41%	43%	31%	34%	21%
	GMCP	91%	86%	86%	83%	94%	38%	44%	32%	31%	19%
1200	AGLASSO	92%	87%	85%	84%	95%	32%	32%	29%	30%	20%
	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	97%	95%	93%	93%	88%	31%	24%	24%	22%	15%
	GSCAD	95%	95%	92%	91%	89%	30%	24%	22%	18%	13%
	GMCP	94%	92%	89%	92%	83%	30%	26%	20%	19%	14%
	AGLASSO	95%	92%	90%	91%	88%	30%	23%	22%	21%	14%

Table 26: GCV 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	0%	1%	0%	0%	21%	1%	1%	0%	0%	4%
	GSCAD	0%	1%	0%	0%	21%	1%	1%	0%	0%	4%
	GMCP	0%	0%	0%	0%	21%	0%	0%	0%	0%	4%
	AGLASSO	1%	1%	0%	0%	19%	1%	1%	0%	0%	3%
800	MLE	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	12%	11%	3%	3%	48%	1%	1%	0%	0%	2%
	GSCAD	12%	11%	2%	2%	48%	1%	1%	0%	0%	2%
	GMCP	11%	10%	0%	0%	48%	1%	1%	0%	0%	1%
	AGLASSO	11%	10%	1%	1%	45%	1%	1%	0%	0%	1%
1200	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	31%	30%	4%	5%	82%	0%	0%	0%	0%	5%
	GSCAD	35%	34%	4%	5%	82%	0%	0%	0%	0%	5%
	GMCP	25%	24%	4%	5%	82%	0%	0%	0%	0%	5%
	AGLASSO	27%	26%	4%	4%	76%	0%	0%	0%	0%	5%

Table 27: BIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	58%	60%	40%	39%	74%	22%	25%	14%	15%	20%
	GSCAD	57%	59%	40%	40%	75%	21%	25%	15%	15%	19%
	GMCP	21%	23%	10%	8%	69%	1%	1%	0%	0%	14%
	AGLASSO	53%	55%	36%	35%	70%	13%	16%	9%	9%	12%
800	MLE	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	86%	83%	71%	70%	90%	31%	37%	17%	18%	16%
	GSCAD	84%	82%	71%	69%	90%	32%	36%	16%	17%	16%
	GMCP	79%	77%	66%	64%	90%	29%	31%	11%	13%	14%
	AGLASSO	82%	80%	64%	62%	92%	21%	22%	10%	11%	12%
1200	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	94%	91%	81%	82%	87%	40%	37%	16%	15%	10%
	GSCAD	94%	91%	81%	82%	85%	39%	37%	15%	14%	10%
	GMCP	93%	90%	78%	79%	87%	37%	34%	12%	13%	10%
	AGLASSO	94%	91%	79%	80%	90%	26%	22%	8%	7%	11%

Table 28: AIC 挑選標準；左欄位為考慮左截斷時間之結果，右欄位為未考慮左截斷時間之結果；此表顯示由參數估計值與變異估計值建構的 95% 信賴區間覆蓋參數真值的比例

n	method	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6	β_{1_1}	β_{1_2}	β_{3_1}	β_{3_2}	β_6
400	MLE	95%	96%	93%	91%	93%	86%	85%	79%	86%	64%
	GLASSO	86%	86%	84%	83%	94%	60%	65%	51%	55%	42%
	GSCAD	87%	87%	81%	84%	91%	50%	52%	41%	44%	37%
	GMCP	77%	78%	62%	62%	91%	34%	40%	29%	28%	33%
	AGLASSO	87%	87%	79%	80%	91%	42%	44%	36%	39%	27%
800	MLE	96%	92%	100%	97%	98%	79%	76%	69%	69%	23%
	GLASSO	96%	93%	99%	96%	92%	48%	48%	44%	42%	20%
	GSCAD	93%	90%	99%	96%	95%	45%	48%	40%	42%	20%
	GMCP	93%	90%	98%	95%	94%	44%	47%	38%	37%	19%
	AGLASSO	95%	92%	98%	95%	95%	45%	48%	44%	41%	20%
1200	MLE	98%	95%	92%	95%	91%	64%	63%	48%	56%	14%
	GLASSO	97%	97%	96%	93%	91%	31%	23%	24%	21%	15%
	GSCAD	97%	96%	96%	93%	87%	31%	25%	23%	21%	14%
	GMCP	96%	95%	94%	93%	86%	31%	25%	23%	20%	15%
	AGLASSO	96%	94%	95%	93%	90%	32%	25%	22%	20%	12%

7.3 附錄三

Table 29: 放射性治療

RAD	樣本數	水準意義
RAD0	1430	無
RAD1	761	光束輻射
RAD4	28	光束輻射”結合”放射性植入物”或者”放射性同位素
RAD8	46	建議服射治療但未知結果治療方式

Table 30: 居住地區

REG	樣本數	水準意義
REG1	317	未知
REG2	405	康乃狄克
REG20	478	底特律
REG21	56	夏威夷
REG22	409	愛荷華
REG23	111	新墨西哥
REG25	282	西雅圖
REG26	79	猶他
REG27	131	亞特蘭大

Table 31: 人種

RACE	樣本數	水準意義
RACE0	54	未知
RACE1	1974	白人
RACE2	171	黑人
RACE3	57	其他
RACE4	9	亞洲人

Table 32: 手術方式

SUG	樣本數	水準意義
SUG9	724	見 http://seer.cancer.gov/archive/manuals/AppendD.pdf
SUG10	325	
SUG20	64	
SUG40	135	
SUG50	552	
SUG60	17	
SUG80	37	
SUG90	411	



Table 33: 婚姻狀況

MAR	樣本數	水準意義
MAR1	217	單身
MAR2	969	已婚
MAR3	42	分居
MAR4	154	離婚
MAR5	819	喪偶
MAR9	64	未知

Table 34: 癌細胞擴散情況

HS	樣本數	水準意義
HS2	76	區域
HS4	2186	擴散

Table 35: 源發腫瘤偏側位置

LAT	樣本數	水準意義
LAT0	69	非成對器官
LAT1	1029	源發於右側器官
LAT2	1051	源發於左側器官
LAT4	44	雙側乳房接被侵犯，單一源發部位
LAT9	72	成對器官，但未知偏側

Table 36: 腫瘤分化情況

GRADE	樣本數	水準意義
GRADE1	31	完全分化
GRADE2	182	大部分分化
GRADE3	387	少部分分化
GRADE4	73	未分化
GRADE9	1592	未知



References

- Breheny, P. and Huang, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing* **25**, 173–187.
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics* **24**, 2350–2383.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B* **34**, 187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics* **32**, 407–499.
- Fan, J. and Li, R. (2002). Variable selection for cox’s proportional hazards model and frailty model. *The Annals of Statistics* **30**, 74–99.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**, 1–22.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, 2nd edition.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.

- Lu, W. and Zhang, H. H. (2007). Variable selection for proportional odds model. *Statistics in Medicine* **26**, 3771–3781.
- Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine* **127**, 757–763.
- Simon, N., Friedman, J., Hastie, T., Tibshirani, R., et al. (2011). Regularization paths for cox’s proportional hazards model via coordinate descent. *Journal of Statistical Software* **39**, 1–13.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267–288.
- Tibshirani, R. et al. (1997). The lasso method for variable selection in the cox model. *Statistics in Medicine* **16**, 385–395.
- Tsai, W.-Y., Jewell, N. P., and Wang, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74**, 883–886.
- van Houwelingen, H. C., Bruinsma, T., Hart, A. A. M., van’t Veer, L. J., and Wessels, L. F. A. (2006). Cross-validated cox regression on microarray gene expression data. *Statistics in Medicine* **25**, 3201–3216.
- Wang, H. and Leng, C. (2008). A note on adaptive group lasso. *Computational Statistics and Data Analysis* **52**, 5277–5286.
- Wang, L., Chen, G., and Li, H. (2007). Group scad regression analysis for microarray time course gene expression data. *Bioinformatics* **23**, 1486–1494.

- Wang, M.-C. (1989). A semiparametric model for randomly truncated data. *Journal of the American Statistical Association* **84**, 742–748.
- Wang, M.-C., Brookmeyer, R., and Jewell, N. P. (1993). Statistical models for prevalent cohort data. *Biometrics* **49**, 1–11.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 49–67.
- Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38**, 894–942.
- Zhang, H. H. and Lu, W. (2007). Adaptive lasso for cox’s proportional hazards model. *Biometrika* **94**, 691–703.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429.
- Zou, H. (2008). A note on path-based variable selection in the penalized proportional hazards model. *Biometrika* **95**, 241–247.
- Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics* **36**, 1509.