國立政治大學統計學系研究所

碩士學位論文

LASSO 迴歸在 *B*-spline 基底組成之危險函數上的應用

Application of LASSO Regression in Estimating *B*-Spline-Based Hazard Functions

國立政治大學 National Chengchi University

指導教授：黃子銘　博士

研究生：林子元　撰

中 華 民 國 一 〇 六 年 二 月

摘　要

一項關於比例危險模型的重要假設為對數危險函數與共變量之間的關係是線性的，本文探討當此假設不成立時，使用 B 樣條基底函數來近似共變量的非線性函數是可行的。在估計上，本文應用了 group lasso 方法。在適當的懲罰係數之下，對於不具解釋力的共變量而言，此方法可使對應至該共變量的一組基底係數同時估為零，以避免模型難以解讀的狀況。此外，本文嘗試為所提模型發展假設檢定。考慮的檢定量除了一般的 Wald 檢定量、概似比檢定量與分數檢定量之外，尚包括了因應懲罰項而作校正的檢定量與基於拔靴法的檢定量。本文採用模擬的方法比較各檢定量的優劣。

關鍵字：比例危險模型、B 樣條、group lasso、拔靴法

國立政治大學
National Chengchi University

Abstract

A strong assumption in the Cox proportional hazards model requires linearity of the covariates on the log hazard function. However, this assumption may be violated in practice. Alternatively, it is feasible to model the nonlinear effect via a combination of *B*-spline basis functions. In estimating the basis coefficients, the group lasso is applied. By so doing, a group of coefficients can be set zero simultaneously if the corresponding covariate is not predictive. Lastly, I develop hypothesis testing regarding this model. In addition to the ordinary Wald statistic, likelihood ratio statistic, and score statistic, two other types of testing statistic are considered: one adjust for penalty function and the other one based on bootstrap samples. Simulation studies are carried out to evaluate the performance of the proposed statistics.

Keywords: proportional hazards model, *B*-splines, group lasso, bootstrap

國立政治大學
National Chengchi University

目　次

第一章　緒論 1
第一節　研究動機 1
第二節　文獻回顧 2
第三節　方法摘要 2
第二章　研究方法 4
第一節　模型架構 4
一、比例危險模型 4
二、延伸的比例危險模型 5
三、*B* 樣條近似方法 6
四、最終模型 7
五、交互作用 9
第二節　模型估計 9
一、Lasso 方法 9
二、Group lasso 方法 10
三、懲罰係數的選取 11
四、節點與階數的選取 12
五、計算 14
第三節　參數推論 16
一、基於卡方分配的參數檢定 16
二、校正的參數檢定 17
三、基於拔靴法的參數檢定 18
四、聯合信賴束 20
第三章　模擬與比較 23
第一節　與標準比例危險模型的比較 24
第二節　懲罰係數對檢定量的影響 28
第三節　拔靴法樣本的代表性 31
第四節　型一錯誤與檢定力 35
第四章　結論與建議 41
參考文獻 42

國立政治大學 National Chengchi University

表　次

表 3-1、各模型與各樣本數下β_1估計的 MSE 25
表 3-2、各模型與各樣本數下β_1估計值的標準差 25
表 3-3、各模型與各樣本數下η_2估計的 MISE 25
表 3-4、拔靴法方法二與方法三的比較 32
表 3-5、懲罰係數為 1 之下各檢定量之型一錯誤與檢定力 37
表 3-6、懲罰係數為 2 之下各檢定量之型一錯誤與檢定力 38
表 3-7、懲罰係數為 3 之下各檢定量之型一錯誤與檢定力 39

國立政治大學
National Chengchi University

圖　次

圖 3-1、標準模型對共變量函數η_2的估計.....26
圖 3-2、樣條模型對共變量函數η_2的估計.....27
圖 3-3、Wald 檢定量在各模型與各懲罰係數下的分配.....29
圖 3-4、概似比檢定量在各模型與各懲罰係數下的分配.....30
圖 3-5、分數檢定量在各模型與各懲罰係數下的分配.....30
圖 3-6、Wald 檢定量的拔靴法分配.....33
圖 3-7、概似比檢定量的拔靴法分配.....33
圖 3-8、分數檢定量的拔靴法分配.....34
圖 3-9、各檢定量在各情境下的型一錯誤與檢定力.....40

國立政治大學
National Chengchi University

第一章　緒論

第一節　研究動機

在生物統計中，存活時間為研究者時常面臨的資料。例如，在分析 HIV 感染者經歷潛伏期至 AIDS 發作的所需時間時，研究者欲瞭解哪些共變量對於發病時間是有解釋力的，這些共變量可能包括：感染者的性別、初感染年紀、病毒量、CD4 細胞數與投藥時間早晚等等，或類別型或連續型。一般來說，存活時間資料有著特殊的機率結構，因為並非所有的研究都具足夠的時間來追蹤病人的狀態，或者病人在研究中忽然失去聯繫，而造成研究者最終無法觀察到病人的結果。這樣的情況，即稱為設限(censoring)，也促成了分析上的不易。

在分析設限的存活時間資料與共變量之間的關係時，Cox (1972)的比例危險模型(proportional hazards model)為最廣泛被應用的模型。該模型描述了研究對象的危險函數(hazard function)與共變量之間的關係如下，

$$\lambda(t|\mathbf{z}) = \lambda_0(t)\exp(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) \tag{1.1}$$

其中，λ與λ_0分別為研究對象的危險函數與基線危險函數(baseline hazard function)，兩者皆為存活時間t的函數。另外，$\mathbf{z}$為研究對象的共變量，而$\boldsymbol{\beta}$為對應共變量的迴歸係數。該模型在概念上簡單，且不須界定λ_0的型式下即可估計出$\boldsymbol{\beta}$係數，只是須考慮一項重要的假設：危險函數與共變量之間具備一個對數的線性關係。然而，當這樣的假設不成立時，研究者就有可能作出錯誤的決策與推論(Keele, 2010)。因此，本文欲探討式(1.1)的延伸模型，使對數危險函數(log-hazard function)與連續型共變量之間不必然呈現線性關係，而是一個任意的平滑函數(smooth function)，並發展此延伸模型下的假設檢定。

第二節　文獻回顧

在探討 Cox 延伸模型的文獻中，一般考慮將此模型寫為一般化的可加性模型(generalized additive model)，也就是各個共變量皆是透過某個平滑函數來影響對數危險函數，其總影響即為各個函數的加總。Hastie & Tibshirani (1990)使用了類似局部得點演算法(local scoring algorithm)的方法來估計各個函數的型態。Sleeper & Harrington (1990)使用了多項式樣條(polynomial spline)來近似這些未知函數。同樣透過可加模型，Gray (1992)則採用受懲罰的樣條(penalized spline)在隨時間變化的係數(time-varying coefficient)模型上。Kooperberg, Stone, & Truong (1995)直接將危險函數寫作一個基於樣條的迴歸模型，而捨棄了危險函數的比例假設。LeBlanc & Crowley (1999)發展了自我調整迴歸樣條(adaptive regression spline)的方法來配適可加性模型的函數與交互作用。Huang & Liu (2006)則維持共變量與迴歸係數的線性組合形式，但是此線性組合係透過某個鏈結函數(link function)來影響危險函數，而在估計這個鏈結函數時，同樣地運用了樣條函數來作近似。

第三節　方法摘要

本文採用 Sleeper & Harrington (1990)所提之方法，以 *B* 樣條基底函數(*B*-spline basis function)來近似未知的共變量函數。由於一個共變量的函數須對應至多個基底係數，若是模型納入的共變量過多，則參數數量便會非常龐大。因此，在參數估計方面，本文採用具懲罰效果的最小絕對壓縮挑選機制(least absolute shrinkage and selection operator, lasso)來進行估計。Lasso 方法為 Tibshirani (1996, 1997)所提出，目的在於壓縮係數，而達到稀疏估計(sparse estimation)的效果。在近期內，lasso 成為一個相當熱門的估計方法，尤其是應用於高維度資料的分析當中。此方法運作的原理為，在最適化某個目標函數來進行參數估計時，

納入一個L_1範數(norm)的限制式，以達到壓縮係數的效果。具體而言，

$$\boldsymbol{\beta}^{\text{lasso}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\{\ell(\boldsymbol{\beta}) - \lambda\|\boldsymbol{\beta}\|_1\} \tag{1.2}$$

其中，目標函數ℓ為對數概似函數，$\|\cdot\|_1$為L_1範數，而λ為懲罰係數。在適度的懲罰之下，此估計方法將導致某些解釋力較低的共變量所對應的係數為零，進而提升模型的解讀性。

考慮到本文模型中，各個共變量乃是對應至一組基底係數，若運用傳統的 lasso 方法，可能便出現在對應至同一個共變量的基底係數中，某些係數被估成零，而某些不是零的矛盾狀況。據此，本文將選擇 group lasso 方法(Yuan & Lin, 2006)來進行估計。Group lasso 為 lasso 方法的延伸，此方法在懲罰係數時，是以組為單位，得以將同一組係數同時估為零。Group lasso 除了應用在一般的線性模型當中，也延伸至羅吉斯迴歸模型(Meier, van de Geer, & Bühlmann, 2008)與標準的 Cox 比例危險模型(Kim, Sohn, Jung, Kim, & Park, 2012)。但是對於一個以 B 樣條基底來近似共變量函數的比例危險模型，則尚未有文章詳加討論。因此，本文選擇此方法來深入研究。

進行完參數估計之後，本文嘗試提出幾個檢定方法來判斷共變量的顯著性。其中包含了基於卡方分配的檢定量、因應懲罰函數而作修正的檢定量，與基於拔靴法(bootstrap)的檢定量。最後，本文以模擬的方式比較，在不同樣本數與不同懲罰係數之下，各個檢定量犯型一錯誤(type I error)的機率與檢定力(power)。基於模擬結果，本文將對檢定量的選用給予適當建議。

第二章　研究方法

第一節　模型架構

考慮一組存活分析資料$(Y_i, \delta_i, \mathbf{z}_i)$，$1 \le i \le n$。其中，$Y$為研究者觀察到的時間，即$Y = \min\{T, C\}$，此處$T$為事件發生所歷經的時間，$C$則為設限時間；$\delta$為指標變數，定義為$\delta = I(T \le C)$，即當研究者觀察到事件發生時，其值為1，反之為0；$\mathbf{z} = (z_1 \quad \cdots \quad z_p)^{\mathrm{T}}$收集了一系列可能影響存活時間的共變量，包含$p_1$個類別共變量與$p_2$個連續共變量。

在給定一組共變量$\mathbf{z}$下，探討存活時間T的分配可以藉由密度函數(density function) $f(t|\mathbf{z})$或是分配函數(distribution function) $F(t|\mathbf{z})$，不過更常見地，則是藉由危險函數$\lambda(t|\mathbf{z})$來探討。其定義如下，

$$\lambda(t|\mathbf{z}) = \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta t | T > t, \mathbf{z})}{\Delta t} \tag{3.1}$$

$$= \frac{f(t|\mathbf{z})}{1 - F(t|\mathbf{z})} \tag{3.2}$$

也就是說，在給定一組共變量$\mathbf{z}$且事件在時間點t尚未發生的條件之下，該事件於區間$(t, t + \Delta t]$發生的風險。由於危險函數的定義在概念上相當簡易，這促成為存活時間的分配建立模型的出發點。在眾多模型當中，最為經典的便是 Cox (1972) 的比例危險模型。

一、比例危險模型

Cox 提出比例危險模型如下，

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp(\mathbf{z}^{\mathrm{T}}\boldsymbol{\beta}) \tag{3.3}$$

其中，$\lambda_0(t)$為基線危險函數，此函數被視為是一個干擾參數(nuisance parameter)，而$\boldsymbol{\beta}$為對應共變量$\mathbf{z}$的迴歸參數向量。該模型最大的特點為，並不需要事先界定λ_0的形式，便可直接估計迴歸係數。因此，這是一個半參數的模型(semiparametric

model)。在比例危險模型中，對於兩個擁有不同共變量的研究對象，兩者的危險比值並不隨時間改變，

$$\frac{\lambda(t|\mathbf{z}_1)}{\lambda(t|\mathbf{z}_2)} = \exp\{(\mathbf{z}_1^{\mathrm{T}} - \mathbf{z}_2^{\mathrm{T}})\boldsymbol{\beta}\} \tag{3.4}$$

這是此模型被稱為「比例」危險的由來。

對於$\boldsymbol{\beta}$的估計，主要係藉由最適化偏概似函數(partial likelihood function) (Cox, 1975)，

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left(\frac{\exp(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta})}{\sum_{j=1}^{n}\delta_{ij}\exp(\mathbf{z}_j^{\mathrm{T}}\boldsymbol{\beta})}\right)^{\delta_i} \tag{3.5}$$

其中，δ_{ij}為指標變數，定義為$\delta_{ij} = I(Y_j \geq Y_i)$。當$\delta_{ij} = 1$時，表示個體$j$於時間點$Y_i$時暴露在事件發生的風險中，稱此集合$\{j: \delta_{ij} = 1\}$為風險集合(risk set)。因此，偏概似函數的涵義為，在給定時間點Y_i下的風險集合與事件會發生的條件之下，對於每一個發生事件的個體i，即$\{i: \delta_i = 1\}$，將其對應機率相乘。不過，當資料出現重複的觀察時間(ties)時，偏概似函數由於需考量觀察時間的所有排列，在計算上較為繁瑣，此時可以使用 Breslow (1974)、Efron (1977)及 Kalbfleisch & Prentice (1973)等人的估計方法。

藉由最適化偏概似函數所得之最大偏概似估計量$\widehat{\boldsymbol{\beta}}_{\mathrm{PMLE}}$，在大樣本下，與一般的最大概似估計量相同，具有良好的漸近多元常態性質。具體而言，

$$\widehat{\boldsymbol{\beta}}_{\mathrm{PMLE}} \xrightarrow{D} \mathcal{N}_p\left(\boldsymbol{\beta}, \boldsymbol{\mathcal{J}}^{-1}(\widehat{\boldsymbol{\beta}}_{\mathrm{PMLE}})\right) \tag{3.6}$$

其中，$\boldsymbol{\mathcal{J}}^{-1}$為觀察的資訊矩陣(observed Fisher information matrix)的逆矩陣。這個漸近常態的性質為假設檢定的方法奠定了基礎。

二、延伸的比例危險模型

比例危險模型的假設之一為，共變量對於對數危險函數的影響是線性的。因此，當線性假設不成立時，研究者就有可能作出錯誤的結論。本文於是考慮一個

更廣義的比例危險模型，

$$\lambda(t|\mathbf{z}) = \lambda_0(t)\exp[\eta(\mathbf{z})] \tag{3.7}$$

其中，η為$\mathbf{z}$的任意的平滑函數。雖然這是比例危險模型最廣義的型態，但是只要當$\mathbf{z}$的維度$p \geq 3$時，實務上便容易受樣本數的限制而難以估計出η的真正樣貌，這個現象被稱作「維度的詛咒 (curse of dimensionality)」。

為了解決這個問題，一個可加性迴歸模型更常用來替代上述模型 (Stone, 1985)，

$$\eta(\mathbf{z}) = \sum_{j=1}^{p_1} \beta_j z_j + \sum_{j=p_1+1}^{p} \eta_j(z_j) \tag{3.8}$$

也就是說，將各個類別共變量的影響維持為線性，而各個連續共變量的影響皆係透過某個任意平滑函數，而整體的影響是可以直接加總的。為了估計這些未知函數，本文以 B 樣條基底函數(de Boor, 1978)對這些函數作近似。

三、B 樣條近似方法

樣條是一個分段的多項式。在樣條相連接的地方，具有某些特定的連續性條件。在迴歸模型中，使用樣條來近似連續變數的非線性函數可以組成一個平滑的近似曲線(Wold, 1974)。樣條函數屬於有限維度的線性空間中，而 B 樣條函數不只構成了這個空間的基底，在計算上也具備穩定的性質。此外，樣條配適也不像多項式配適(polynomial fitting)那般容易受到離群值影響。

若變數z的值落在區間$[a,b]$上，考慮$\boldsymbol{\xi} = (\xi_1 \quad \cdots \quad \xi_s)^{\mathrm{T}}$而$a < \xi_1 < \cdots < \xi_s < b$為$[a,b]$的$s$個分割點，稱這些分割點為樣條函數的節點(knot)。在每一區間$[\xi_{i-1}, \xi_i]$當中，各個樣條皆是一個階數(order)為k的多項式（即$k-1$次多項式），且在開區間(a,b)上，各個樣條函數至少在第$k-2$階的導函數皆為連續。這些樣條的集合組成了一個維度為$s+k$的線性空間(de Boor, 1978)，且這個空間的基底可由 B 樣條基底函數$\{\mathcal{B}_t(\cdot\,|k,\boldsymbol{\xi})\}_{t=1}^{s+k}$來建構。因此，$B$ 樣條的線性組合得以組成一

條平滑曲線。

根據 Curry-Schoenberg (1966)所提出的定理，任何的樣條函數皆可被唯一一組 B 樣條基底函數的線性組合表達。本文於是以 B 樣條基底函數來近似連續共變量j的非線性函數，

$$\eta_j(z_j) \approx \sum_{t=1}^{q_j} \gamma_{jt} \mathcal{B}_t^{(j)}(z_j) \tag{3.9}$$

其中，$\mathcal{B}_t^{(j)}(\cdot) \equiv \mathcal{B}_t(\cdot \,|k_j, \boldsymbol{\xi}_j)$為給定階數$k_j$與節點向量$\boldsymbol{\xi}_j$的 B 樣條基底函數；q_j為基底函數的個數，為節點個數與階數的總和；γ_{jt}為各基底函數所對應之係數，本文將在第二節中詳細探討估計基底係數的方法。

四、最終模型

結合式(3.8)與式(3.9)，

$$\eta(\mathbf{z}) = \sum_{j=1}^{p_1} \beta_j z_j + \sum_{j=p_1+1}^{p} \sum_{t=1}^{q_j} \gamma_{jt} \mathcal{B}_t^{(j)}(z_j) \tag{3.10}$$

注意到，這是一個過度參數化(over-parameterized)的模型，因為比例風險模型並不具備截距項，可是 B 樣條基底函數的總和為一，即

$$\sum_{t=1}^{q_j} \mathcal{B}_t^{(j)}(z_j) = 1, \qquad \forall j \tag{3.11}$$

捨棄其中任何一個基底即可解決模型過度參數化的問題。此處，本文選擇捨去最後一個基底係數，即令$\gamma_{jq_j} = 0$。另外，本文設置限制式，

$$\lambda(t|\mathbf{z} = \mathbf{0}) = \lambda_0(t) \tag{3.12}$$

也就是說，當共變量$\mathbf{z}$為$\mathbf{0}$時，危險函數即等於基線危險函數。以下重設模型參數使式(3.12)成立，對於每一個共變量j，皆自基線危險函數中取出一截距項c_j來滿足下式，

$$\eta_j(0) + c_j = 0 \tag{3.13}$$

故，

$$c_j = -\sum_{t=1}^{q_j-1} \gamma_{jt} \mathcal{B}_t^{(j)}(0) \tag{3.14}$$

結合式(3.10)與式(3.14)，並以向量符號表示，得到最終模式如下，

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp[\grave{\eta}(\mathbf{z})] \tag{3.15}$$

$$\grave{\eta}(\mathbf{z}) = \eta(\mathbf{z}) + \sum_{j=p_1+1}^{p} c_j \tag{3.16}$$

$$= \mathbf{b}^{\mathrm{T}}\boldsymbol{\gamma} + \mathbf{1}^{\mathrm{T}}\mathbf{c} \tag{3.17}$$

其中，

$$\mathbf{b} = \begin{pmatrix} z_1 & \cdots & z_{p_1} & \mathcal{B}_1^{(p_1+1)}(z_{p_1+1}) & \cdots & \mathcal{B}_{q_p-1}^{(p)}(z_p) \end{pmatrix}^{\mathrm{T}} \tag{3.18}$$

$$\boldsymbol{\gamma} = \begin{pmatrix} \beta_1 & \cdots & \beta_{p_1} & \gamma_{p_1+1,1} & \cdots & \gamma_{p,q_p-1} \end{pmatrix}^{\mathrm{T}} \tag{3.19}$$

$$\mathbf{c} = \begin{pmatrix} c_{p_1+1} & \cdots & c_p \end{pmatrix}^{\mathrm{T}} \tag{3.20}$$

而$\mathbf{1} \in \mathbb{R}^{p_2}$為元素皆是 1 的向量。對應此模型的偏概似函數與對數偏概似函數分別為

$$PL(\boldsymbol{\gamma}) = \prod_{i=1}^{n} \left(\frac{\exp(\mathbf{b}_i^{\mathrm{T}}\boldsymbol{\gamma})}{\sum_{j=1}^{n} \delta_{ij} \exp(\mathbf{b}_j^{\mathrm{T}}\boldsymbol{\gamma})} \right)^{\delta_i} \tag{3.21}$$

$$\ell(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \delta_i \left(\mathbf{b}_i^{\mathrm{T}}\boldsymbol{\gamma} - \log \sum_{j=1}^{n} \delta_{ij} \exp(\mathbf{b}_j^{\mathrm{T}}\boldsymbol{\gamma}) \right) \tag{3.22}$$

由基線危險函數取出的截距項$\mathbf{c}$並不影響偏概似函數，因為截距項於分子分母間可相互抵消，所以偏概似函數為$\boldsymbol{\gamma}$的函數。估計時，採最適化對數偏概似函數得到最大偏概似估計量$\hat{\boldsymbol{\gamma}}$，之後再依式(3.14)計算各個函數的截距項即可。

五、交互作用

本文引用的模型可以再作延伸，以容納變量間的交互作用。舉例來說，若研究者考慮一個類別共變量z_1與一個連續共變量z_2，並懷疑兩者之間存在交互作用，則可以將模式寫作

$$\eta(z_1, z_2) = \beta_1 z_1 + z_1 \eta_{12}^{(1)}(z_2) + (1 - z_1)\eta_{12}^{(2)}(z_2) \tag{3.23}$$

當$z_1 = 0$時，所對應的模型為$\eta(z_1 = 0, z_2) = \eta_{12}^{(2)}(z_2)$；當$z_1 = 1$時，所對應的模型為$\eta(z_1 = 1, z_2) = \beta_1 + \eta_{12}^{(1)}(z_2)$。比較兩模型差異，可見除了類別共變量$z_1$的影響有無之外，連續共變量$z_2$的影響也因為$z_1$的有無而透過不同的函數來影響危險函數（即交互作用）。前述以 B 樣條基底函數為未知函數$\eta_{12}^{(1)}$、$\eta_{12}^{(2)}$作近似的方法仍可沿用。

若研究者考慮的是兩個連續共變量z_1、z_2之間的交互作用，則可以將其中一個共變量類別化，並套用上述方法，或者同樣利用 B 樣條基底函數來近似兩者之間的交互作用$\eta_{12}(z_1, z_2)$。此時，近似的函數為雙變量函數，可採用 O'Sullivan (1988)的張量積樣條(tensor product spline)，也就是將兩個共變量所需的基底函數相乘。具體而言，

$$\eta_{12}(z_1, z_2) = \sum_{t=1}^{q_1} \sum_{r=1}^{q_2} \gamma_{jk} \mathcal{B}_t^{(1)}(z_1) \mathcal{B}_r^{(2)}(z_2) \tag{3.24}$$

由於這兩個基底的乘積所展延(span)的空間亦包含了常數項，這導致了模型過度參數化的問題，將其中一個基底係數設為零便能確保模型的辨識性。

第二節　模型估計

一、Lasso 方法

最小絕對壓縮挑選機制(least absolute shrinkage and selection operator,

lasso) (Tibshirani, 1996, 1997)是一種同時估計參數與挑選變數的方法。其估計原理為，在最適化目標函數$\ell(\boldsymbol{\gamma})$的時候，藉由引入$\boldsymbol{\gamma}$的$L_1$範數來達到係數壓縮的效果。具體而言，

$$\hat{\boldsymbol{\gamma}}_s = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}}\{\ell(\boldsymbol{\gamma})\} \quad \text{subject to} \quad \|\boldsymbol{\gamma}\|_1 \leq s \tag{3.25}$$

此處，$\|\cdot\|_1$為L_1範數，而s為一個調諧參數(tuning parameter)。式(3.25)也可以藉拉格朗日式(Lagrangian form)來表達，

$$\hat{\boldsymbol{\gamma}}_\lambda = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}}\{\ell(\boldsymbol{\gamma}) - \lambda\|\boldsymbol{\gamma}\|_1\} \tag{3.26}$$

其中，λ為一個調諧參數，它控制了估計誤差與估計變異之間的取捨關係(trade-off)。顯然地，當$\lambda = 0$時，式(3.26)相當於沒有限制式的存在，此時所得的估計即為一般的最大偏概似估計值，雖為不偏(unbiased)估計值，但是其估計變異卻最大；當λ逐漸提升，其估計誤差將逐漸上升，而估計變異會逐漸下降。當λ大至一定程度時，所得估計值為零向量，此時估計誤差最大，而估計變異為零。

藉由引入L_1範數，所有的係數在估計時皆會受到輕重不一的懲罰(penalization)，尤其對於那些不具解釋能力的變數，在給予一個適當的λ下，其係數有可能被估為零。懲罰的目的有三：第一，藉由犧牲估計誤差來換取較小的估計變異，進而提升估計的精準度(accuracy)；第二，在模型中選取較少的變數有助於模型的解釋，此為精簡原則(principle of parsimony)；第三，在某些情況中，研究者可能面臨過多的變數，變數的數目甚至多於樣本數，此時傳統的分析方法便會失效，而 lasso 方法則不受變數個數的限制。因此，lasso 被廣泛應用於高維資料的分析當中。

二、Group lasso 方法

在某些情況中，研究者可能欲同時納入或排除一組變數，而非單一變數，例如：以多個虛擬變數(dummy variable)編碼而成的因子，或是如本文模型中以一

組基底函數來近似的共變量。Lasso 方法仍然能夠使用，只是在估計上，可能會造成同一組基底係數中，某些係數被估為零，而某些不是。這樣的矛盾現象有礙於模型的解釋。為了解決這個問題，Yuan & Lin (2006)提出了 group lasso，這個方法同時懲罰了一組變數，而非單一變數。具體而言，令$\boldsymbol{\gamma}_j$為對應係數向量$\boldsymbol{\gamma}$中的第j組係數，

$$\hat{\boldsymbol{\gamma}}_\lambda = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}}\left\{\ell(\boldsymbol{\gamma}) - \sum_{j=1}^{p} \lambda_j \left\|\boldsymbol{\gamma}_j\right\|_2\right\} \tag{3.27}$$

此處，$\|\cdot\|_2$為L_2範數。注意到，當λ_j皆相等且各組大小為一時，式(3.27)與式(3.26)是恆等的。此外，與其為每一組變數給予不同大小的懲罰係數，Meier et al. (2008)與 Yuan & Lin (2006)皆考慮將懲罰係數寫為$\lambda_j = \lambda\sqrt{m_j}$，其中，$m_j$為各組的成員數。對於類別共變量，$m_j = 1$；對於連續共變量，$m_j = q_j - 1$。因此，

$$\hat{\boldsymbol{\gamma}}_\lambda = \underset{\boldsymbol{\gamma}}{\operatorname{argmax}}\{\ell_{\text{pen}}(\boldsymbol{\gamma})\} \tag{3.28}$$

$$= \underset{\boldsymbol{\gamma}}{\operatorname{argmax}}\left\{\ell(\boldsymbol{\gamma}) - \lambda \sum_{j=1}^{p} \sqrt{m_j}\left\|\boldsymbol{\gamma}_j\right\|_2\right\} \tag{3.29}$$

三、懲罰係數的選取

懲罰係數λ的選取通常是藉由最適化某個λ的函數，例如損失函數(loss function) $\sum_{i=1}^{n} L\left(y_i, \hat{f}(\mathbf{x}_i)\right)$，其中，$\hat{f}(\mathbf{x})$為預測模型所得的配適值(fitted value)，L衡量了觀察值與配適值之間的差距。而找尋使損失函數最適化的λ最常見的方法則為 Verweij & van Houwelingen (1993)根據 Cox 模型所提出的交叉驗證法(cross-validation)。

交叉驗證的想法是，在給定某個λ之下，將樣本隨機分為K組後($1 < K \leq n$)，使用其中$K-1$組作為訓練樣本(training set)，也就是利用此組樣本來估計係數。當模型確立後，再以此模型來預測未用來建模的那組樣本，稱為測試樣本(test

set)。此步驟重複了K次，最後將每一次得到的驗證誤差加總，即為驗證函數。因此，λ的選定是基於驗證函數的最小化。當$K = n$時，此方法又稱為留一驗證(leave-one-out cross-validation)，也就是 Verweij & van Houwelingen (1993)所提出的方法。為避免留一驗證過於龐大的計算量，本文採用K折交叉驗證法（$K < n$），並引用 Bøvelstad et al. (2007)所定義的驗證函數，

$$\mathrm{CV}(\lambda) = \sum_{k=1}^{K} \left[\ell_{\mathrm{pen}}^{(-k)}\left(\hat{\boldsymbol{\gamma}}_{\lambda}^{(-k)}\right) - \ell_{\mathrm{pen}}\left(\hat{\boldsymbol{\gamma}}_{\lambda}^{(-k)}\right) \right] \tag{3.30}$$

其中，ℓ_{pen}為具懲罰項的對數偏概似函數，$\ell_{\mathrm{pen}}^{(-k)}$為捨去第 k 組樣本的具懲罰項的對數偏概似函數，而$\hat{\boldsymbol{\gamma}}_{\lambda}^{(-k)}$為捨去第 k 組樣本並固定在λ的參數估計值。

關於組數的選取，常見的組數為 2 至 10 組，但仍需依樣本數大小來判定。在找尋最適λ時，可以採取固定幾個λ來計算驗證函數，再選取其中使得驗證函數最小的λ。然而，經本文實際操作，發現直接使用 Brent (1973)方法來最小化驗證函數在運算上會較快。只是該方法容易收斂到區域極小值，須嘗試使用幾個不同的起始值來避免該狀況。

四、節點與階數的選取

除了懲罰係數之外，基底函數的節點個數與階數也皆為調諧參數。本文採用最小化艾凱克訊息準則(Akaike information criterion, AIC)或貝氏訊息準則(Bayesian information criterion, BIC)來作為選取節點個數與階數的方法。這些準則評估了模型的複雜度並衡量模型擬合資料的優良性。當模型不具懲罰項時，兩個準則的定義如下：

$$\mathrm{AIC} = -2\ell(\hat{\boldsymbol{\gamma}}) + 2\left(p_1 + \sum_{j=1}^{p_2}(q_j - 1)\right) \tag{3.31}$$

$$\mathrm{BIC} = -2\ell(\hat{\boldsymbol{\gamma}}) + (\log n)\left(p_1 + \sum_{j=1}^{p_2}(q_j - 1)\right) \tag{3.32}$$

其中，$\hat{\boldsymbol{\gamma}}$為（不受懲罰的）最大偏概似估計值，而$p_1 + \sum_{j=1}^{p_2}(q_j - 1)$為模型的自由度，即參數個數。然而，當模型具懲罰項時，有效參數的個數會比名目上來的少。Moody, Hanson, & Lippmann (1992)為具懲罰項的模型提出修正的自由度，

$$\mathrm{AIC}_\lambda = -2\ell(\hat{\boldsymbol{\gamma}}_\lambda) + 2\,\mathrm{trace}\{\ddot{\boldsymbol{\ell}}(\hat{\boldsymbol{\gamma}}_\lambda)\ddot{\boldsymbol{\ell}}_{\mathrm{pen}}^{-1}(\hat{\boldsymbol{\gamma}}_\lambda)\} \tag{3.33}$$

$$\mathrm{BIC}_\lambda = -2\ell(\hat{\boldsymbol{\gamma}}_\lambda) + (\log n)\,\mathrm{trace}\{\ddot{\boldsymbol{\ell}}(\hat{\boldsymbol{\gamma}}_\lambda)\ddot{\boldsymbol{\ell}}_{\mathrm{pen}}^{-1}(\hat{\boldsymbol{\gamma}}_\lambda)\} \tag{3.34}$$

其中，$\ddot{\boldsymbol{\ell}}$與$\ddot{\boldsymbol{\ell}}_{\mathrm{pen}}$分別為$\ell$與$\ell_{\mathrm{pen}}$的二階微分，即海森矩陣(hessian matrix)，$\hat{\boldsymbol{\gamma}}_\lambda$為固定$\lambda$之下所得之（受懲罰的）最大偏概似估計值。注意到，當$\lambda = 0$時，式(3.33)、式(3.34)與式(3.31)、式(3.32)是恆等的，因為$\ddot{\boldsymbol{\ell}}(\hat{\boldsymbol{\gamma}}_\lambda)\ddot{\boldsymbol{\ell}}_{\mathrm{pen}}^{-1}(\hat{\boldsymbol{\gamma}}_\lambda) = \mathbf{I}$，此單位矩陣的維度即為$p_1 + \sum_{j=1}^{p_2}(q_j - 1)$。

採用式(3.33)或式(3.34)來決定節點個數與階數會衍生一個循環問題，也就是，計算準則時必須先選定λ。然而，在利用交叉驗證選定λ時，基底函數的型態卻須先決定，這又牽涉到了節點個數與階數。因此，本文建議先使用不具懲罰項的準則，即式(3.31)或式(3.32)，在考慮各個共變量的節點個數與階數的可能組合之後，選取其中使準則最小化的節點個數與階數來作為起始值，並將此起始值用於交叉驗證來選取λ，即式(3.30)。待λ選定之後，再將λ代入具懲罰項的準則之中，即式(3.33)或式(3.34)，並驗證使準則最小化的節點個數與階數是否與一開始選定的節點個數與階數一致。若一致，則諧調參數的估計流程結束；若不一致，則將新得的節點個數與階數作為起始值並重複上述流程，直到諧調參數的估計達到穩定為止。

節點的配置有許多方法，最簡易的方法為將節點等距離地配置在資料範圍當中，或是將節點配置於資料的分位數上，也有方法是將節點以自我調整(adaptive)的方式配置來達到某種條件(Li, Xu, Zhao, and Goh, 2005)，但是此法通常需仰賴龐大的計算。本文選擇將節點置於資料的分位數上，來確保節點與節點之間的資

料筆數相當。舉例來說，當選定的節點個數為三個時，除了資料的最小值與最大值之外，節點將放置於 25%、50%、75%的分位數上。Sleeper & Harrington (1990)的模擬發現，將節點放置於分位數上，與等距節點相比，在求算逆矩陣上會減少50%的奇異(singularity)情況。

五、計算

模型係數的估計是採最適化受懲罰的對數偏概似函數$\ell_{\text{pen}}(\boldsymbol{\gamma})$。注意到，這個目標函數包含了兩個部分，其一是對數偏概似函數$\ell(\boldsymbol{\gamma})$，如式(3.22)所定義，其二是懲罰函數，以$P(\boldsymbol{\gamma})$表示，

$$P(\boldsymbol{\gamma}) = \lambda \sum_{j=1}^{p} \sqrt{m_j}\, \|\boldsymbol{\gamma}_j\|_2 \tag{3.35}$$

對數偏概似函數是一個凹函數(concave function)，連續且在每一點皆二次可微分，其一階與二階導函數分別為

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \delta_i \left(\mathbf{b}_i - \sum_{j=1}^{n} w_{ij} \mathbf{b}_j \right) \tag{3.36}$$

$$\ddot{\boldsymbol{\ell}}(\boldsymbol{\gamma}) = -\sum_{i=1}^{n} \delta_i \left[\sum_{j=1}^{n} w_{ij} \mathbf{b}_j \mathbf{b}_j^{\mathrm{T}} - \left(\sum_{j=1}^{n} w_{ij} \mathbf{b}_j \right) \left(\sum_{j=1}^{n} w_{ij} \mathbf{b}_j^{\mathrm{T}} \right) \right] \tag{3.37}$$

其中，

$$w_{ij} = \frac{\delta_{ij} \exp\left(\mathbf{b}_j^{\mathrm{T}} \boldsymbol{\gamma}\right)}{\sum_{k=1}^{n} \delta_{ik} \exp\left(\mathbf{b}_k^{\mathrm{T}} \boldsymbol{\gamma}\right)} \tag{3.38}$$

懲罰函數也是一個連續的凹函數，但是在$\boldsymbol{\gamma}_j = \mathbf{0}$時不可微分。將受懲罰的對數偏概似函數的一階導函數以向量符號表示為$\dot{\boldsymbol{\ell}}_{\text{pen}}(\boldsymbol{\gamma}) = \left(\dot{\boldsymbol{\ell}}_{\text{pen},1}(\boldsymbol{\gamma}) \quad \cdots \quad \dot{\boldsymbol{\ell}}_{\text{pen},p}(\boldsymbol{\gamma})\right)^{\mathrm{T}}$，其中，$\dot{\boldsymbol{\ell}}_{\text{pen},j}(\boldsymbol{\gamma})$為第$j$組的一階導函數，其計算如下：

$$\dot{\boldsymbol{\ell}}_{\text{pen},j}(\boldsymbol{\gamma}) = \dot{\boldsymbol{\ell}}_j(\boldsymbol{\gamma}) - \frac{\lambda\sqrt{m_j}}{\|\boldsymbol{\gamma}_j\|_2} \boldsymbol{\gamma}_j, \qquad \boldsymbol{\gamma}_j \neq \mathbf{0} \tag{3.39}$$

此處，$\dot{\boldsymbol{\ell}}_j(\boldsymbol{\gamma})$為對應第$j$組的對數偏概似函數的一階導函數。本文將應用$\dot{\boldsymbol{\ell}}_{\text{pen}}$來解決最適化問題。

一個常用於解決最適問題的方法為牛頓法(Newton's method)，其原理為將目標函數的一階導數以泰勒展開式(Taylor series)作近似，並於起始值$\hat{\boldsymbol{\gamma}}^{(0)}$展開，藉由不斷疊代的方式來逼近實根。具體而言，$\dot{\boldsymbol{\ell}}_{\text{pen}}$的一階泰勒展開式為

$$\dot{\boldsymbol{\ell}}_{\text{pen}}(\boldsymbol{\gamma}) = \dot{\boldsymbol{\ell}}_{\text{pen}}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right) + \ddot{\boldsymbol{\ell}}_{\text{pen}}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right)\left(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}^{(0)}\right) \tag{3.40}$$

若起始值接近最大偏概似估計值，則式(3.40)應接近零，因此，將此式設為零並求解$\boldsymbol{\gamma}$，得

$$\hat{\boldsymbol{\gamma}}^{(1)} = \hat{\boldsymbol{\gamma}}^{(0)} - \ddot{\boldsymbol{\ell}}_{\text{pen}}^{-1}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right)\dot{\boldsymbol{\ell}}_{\text{pen}}\left(\hat{\boldsymbol{\gamma}}^{(0)}\right) \tag{3.41}$$

其中，$\hat{\boldsymbol{\gamma}}^{(1)}$為第一次疊代所得之估計值，將$\hat{\boldsymbol{\gamma}}^{(0)}$更新為$\hat{\boldsymbol{\gamma}}^{(1)}$並重複上述步驟，直到達到某種收斂條件為止，例如$\left|\hat{\boldsymbol{\gamma}}^{(1)} - \hat{\boldsymbol{\gamma}}^{(0)}\right| \leq 10^{-8}$。

不過牛頓法並非萬能，一個明顯的問題即是起始值的設定。倘若起始值距離最大偏概似估計值太遠，則解將難以收斂，或是將收斂至局部極值。即便牛頓法往正確的方向前進，估計上也有可能造成越過(overshoot)極值的情況發生。此外，在疊代過程中，牛頓法須應用到海森矩陣$\ddot{\boldsymbol{\ell}}_{\text{pen}}$。然而，本文考慮的函數是一個極為複雜的函數，難以計算出精確的海森矩陣。因此，本文採取另一個類似牛頓法的方法——有限內存擬牛頓法(Limited-memory Broyden–Fletcher–Goldfarb–Shanno, L-BFGS)，來解決最適問題。

L-BFGS 是擬牛頓法(quasi-Newton method)的衍生。擬牛頓法並不直接計算海森矩陣，而是利用每個步驟中的估計值與其對應的一階導函數來近似海森矩陣（或是其逆矩陣）。若以$\hat{\boldsymbol{\gamma}}^{(s)}$表示疊代第$s$次所得之估計值，並定義$\boldsymbol{\delta}^{(s)}$與$\boldsymbol{\pi}^{(s)}$如下，

$$\boldsymbol{\delta}^{(s)} = \hat{\boldsymbol{\gamma}}^{(s+1)} - \hat{\boldsymbol{\gamma}}^{(s)} \tag{3.42}$$

$$\boldsymbol{\pi}^{(s)} = \dot{\boldsymbol{\ell}}_{\text{pen}}\left(\hat{\boldsymbol{\gamma}}^{(s+1)}\right) - \dot{\boldsymbol{\ell}}_{\text{pen}}\left(\hat{\boldsymbol{\gamma}}^{(s)}\right) \tag{3.43}$$

那麼，在步驟$s+1$的海森矩陣的估計便為下列線性方程組的解，

$$\ddot{\boldsymbol{\ell}}_{\text{pen}}\left(\hat{\boldsymbol{\gamma}}^{(s+1)}\right)\boldsymbol{\delta}^{(s)} = \boldsymbol{\pi}^{(s)} \tag{3.44}$$

此為擬牛頓法的中心想法。在 L-BFGS 中，僅有數個足以代表海森矩陣的逆矩陣的向量會被保留，這可以避免在處理高維資料時消耗過多的內存，從而加快疊代的效率與速度。關於這個演算法的詳細流程，請見 Liu & Nocedal (1989)。

第三節　參數推論

一、基於卡方分配的參數檢定

一旦完成參數估計之後，研究者須依所得估計值來考量所納入模型之共變量是否具顯著的影響力。在本文所提之模型中，由於連續共變量所對應的是一組基底係數，因此在檢定連續共變量的顯著性時，須作聯合檢定(joint hypothesis testing)，也就是檢定一組基底係數是否為零；至於僅對應至一個係數的類別共變量則為個別檢定(individual hypothesis testing)。在比例危險模型當中，普遍使用的檢定統計量有三種型式：Wald 檢定量(Wald statistic)、概似比檢定量(likelihood ratio statistic)，以及分數檢定量(score statistic)。Wald 檢定量是建立在最大偏概似估計值近似常態分配的性質之下，概似比檢定量是基於兩巢狀模型(nested model)的兩倍對數概似函數比值近似卡方分配的結果，而分數檢定量則是應用了分數函數在最大偏概似估計值上會近似常態分配的特點。

為了有助於檢定統計量的表達，本文首先將參數向量$\boldsymbol{\gamma}$分解為$\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\mathrm{T}} \quad \boldsymbol{\gamma}_2^{\mathrm{T}})^{\mathrm{T}}$，其中，$\boldsymbol{\gamma}_1 \in \mathbb{R}^{m_j}$是欲檢定之共變量$j$所對應的參數向量，$\boldsymbol{\gamma}_2 \in \mathbb{R}^{m-m_j}$為其他共變量的參數向量，$m = \sum_{k=1}^{p} m_k$為總參數個數。令$\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1^{\mathrm{T}} \quad \hat{\boldsymbol{\gamma}}_2^{\mathrm{T}})^{\mathrm{T}}$為對應$\boldsymbol{\gamma}$的最大偏概似估計值。同樣地，將分數向量$\boldsymbol{\mathcal{U}}(\boldsymbol{\gamma})$與觀察的資訊矩陣$\boldsymbol{\mathcal{I}}(\boldsymbol{\gamma})$作分解，

$$\boldsymbol{\mathcal{U}}(\boldsymbol{\gamma}) = \dot{\boldsymbol{\ell}}(\boldsymbol{\gamma}) = \left(\boldsymbol{\mathcal{U}}_1^{\mathrm{T}}(\boldsymbol{\gamma}) \quad \boldsymbol{\mathcal{U}}_2^{\mathrm{T}}(\boldsymbol{\gamma})\right)^{\mathrm{T}} \tag{3.45}$$

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\gamma}) = -\ddot{\boldsymbol{\ell}}(\boldsymbol{\gamma}) = \begin{bmatrix} \boldsymbol{\mathcal{I}}_{11}(\boldsymbol{\gamma}) & \boldsymbol{\mathcal{I}}_{12}(\boldsymbol{\gamma}) \\ \boldsymbol{\mathcal{I}}_{21}(\boldsymbol{\gamma}) & \boldsymbol{\mathcal{I}}_{22}(\boldsymbol{\gamma}) \end{bmatrix} \tag{3.46}$$

其中，$\boldsymbol{\mathcal{U}}_1(\boldsymbol{\gamma}) \in \mathbb{R}^{m_j}$和$\boldsymbol{\mathcal{I}}_{11}(\boldsymbol{\gamma}) \in \mathbb{R}^{m_j \times m_j}$分別為欲檢定參數所對應的部分分數向量

與部分觀察的資訊矩陣。此外，令觀察的資訊矩陣的逆矩陣為

$$\mathcal{I}^{-1}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathcal{I}^{11}(\boldsymbol{\gamma}) & \mathcal{I}^{12}(\boldsymbol{\gamma}) \\ \mathcal{I}^{21}(\boldsymbol{\gamma}) & \mathcal{I}^{22}(\boldsymbol{\gamma}) \end{bmatrix} \tag{3.47}$$

定義了這些符號之後，對於一個檢定問題$H_0: \boldsymbol{\gamma}_1 = \mathbf{0}$，Wald 檢定量、概似比檢定量以及分數檢定量的形式可以如下表達，

$$\chi^2_{\mathrm{W}} = \hat{\boldsymbol{\gamma}}_1^{\mathrm{T}}[\mathcal{I}^{11}(\hat{\boldsymbol{\gamma}})]^{-1}\hat{\boldsymbol{\gamma}}_1 \tag{3.48}$$

$$\chi^2_{\mathrm{LR}} = -2[\ell(\hat{\boldsymbol{\gamma}}_0) - \ell(\hat{\boldsymbol{\gamma}})] \tag{3.49}$$

$$\chi^2_{\mathrm{SC}} = \mathcal{U}_1^{\mathrm{T}}(\hat{\boldsymbol{\gamma}}_0)\mathcal{I}^{11}(\hat{\boldsymbol{\gamma}}_0)\mathcal{U}_1(\hat{\boldsymbol{\gamma}}_0) \tag{3.50}$$

其中，$\hat{\boldsymbol{\gamma}}_0$為在虛無假設成立下之最大偏概似估計量。當樣本數與不同死亡時間的觀察值趨近無窮大時，在虛無假設之下，這些統計量將近似卡方分配，其自由度為虛無假設與對立假設兩模型之間參數個數的差。然而，由於本文所提的估計方法，在最適化對數偏概似函數時，因為引入了懲罰項，所得到的估計值已不再是漸近不偏，這將對這些檢定量的分配造成偏移。本文將在第三章中詳細討論這個現象。

二、校正的參數檢定

在具懲罰項的偏概似函數中，Gray (1994)與 Therneau, Grambsch, & Pankratz (2003)提出三種因應懲罰項而作校正的檢定量，並發展出這些檢定量的極限分配。首先，Gray (1994)提議使用下式來估計參數估計的共變異矩陣，

$$\mathbf{V}(\hat{\boldsymbol{\gamma}}) = \mathcal{I}_{\mathrm{pen}}^{-1}(\hat{\boldsymbol{\gamma}})\mathcal{I}(\hat{\boldsymbol{\gamma}})\mathcal{I}_{\mathrm{pen}}^{-1}(\hat{\boldsymbol{\gamma}}) \tag{3.51}$$

此處，$\mathcal{I}_{\mathrm{pen}}(\boldsymbol{\gamma}) = -\ddot{\boldsymbol{\ell}}_{\mathrm{pen}}(\boldsymbol{\gamma})$為受懲罰的觀察資訊矩陣。此外，本文再定義矩陣的分解，

$$\mathbf{V}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{V}_{11}(\boldsymbol{\gamma}) & \mathbf{V}_{12}(\boldsymbol{\gamma}) \\ \mathbf{V}_{21}(\boldsymbol{\gamma}) & \mathbf{V}_{22}(\boldsymbol{\gamma}) \end{bmatrix} \tag{3.52}$$

$$\mathcal{I}_{\mathrm{pen}}^{-1}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathcal{I}_{\mathrm{pen}}^{11}(\boldsymbol{\gamma}) & \mathcal{I}_{\mathrm{pen}}^{12}(\boldsymbol{\gamma}) \\ \mathcal{I}_{\mathrm{pen}}^{21}(\boldsymbol{\gamma}) & \mathcal{I}_{\mathrm{pen}}^{22}(\boldsymbol{\gamma}) \end{bmatrix} \tag{3.53}$$

其中，$\mathbf{V}_{11}(\boldsymbol{\gamma}) \in \mathbb{R}^{m_j \times m_j}$和$\mathcal{I}_{\mathrm{pen}}^{11}(\boldsymbol{\gamma}) \in \mathbb{R}^{m_j \times m_j}$分別為欲檢定參數所對應的部分共變

異估計與部分受懲罰的資訊逆矩陣，並定義$\boldsymbol{\mathcal{U}}_{\text{pen},1}(\boldsymbol{\gamma}) \in \mathbb{R}^{m_j}$為欲檢定參數所對應的部分受懲罰的分數向量。三種校正後的檢定量如下，

$$\chi^2_{\text{W,adj}} = \hat{\boldsymbol{\gamma}}_1^{\text{T}}\left[\boldsymbol{\mathcal{I}}^{11}_{\text{pen}}(\hat{\boldsymbol{\gamma}})\right]^{-1}\hat{\boldsymbol{\gamma}}_1 \tag{3.54}$$

$$\chi^2_{\text{LR,adj}} = -2\left[\ell_{\text{pen}}(\hat{\boldsymbol{\gamma}}_0) - \ell_{\text{pen}}(\hat{\boldsymbol{\gamma}})\right] \tag{3.55}$$

$$\chi^2_{\text{SC,adj}} = \boldsymbol{\mathcal{U}}^{\text{T}}_{\text{pen},1}(\hat{\boldsymbol{\gamma}}_0)\boldsymbol{\mathcal{I}}^{11}_{\text{pen}}(\hat{\boldsymbol{\gamma}}_0)\boldsymbol{\mathcal{U}}_{\text{pen},1}(\hat{\boldsymbol{\gamma}}_0) \tag{3.56}$$

在虛無假設之下，三種檢定量會漸近$\sum_{i=1}^{m_j} e_i Z_i^2$的分配。其中，$Z_i$為標準常態變數，而$e_i$為矩陣$\mathbf{E}$的特徵值，

$$\mathbf{E} = \lim_{n\to\infty}\left[\boldsymbol{\mathcal{I}}^{11}_{\text{pen}}(\boldsymbol{\gamma})\right]^{-1}\mathbf{V}_{11}(\hat{\boldsymbol{\gamma}}) \tag{3.57}$$

顯而易見地，在未受懲罰的模型中，$\mathbf{E}$將化約為維度為m_j的單位矩陣，

$$\left[\boldsymbol{\mathcal{I}}^{11}_{\text{pen}}(\boldsymbol{\gamma})\right]^{-1}\mathbf{V}_{11}(\hat{\boldsymbol{\gamma}}) = \left[\boldsymbol{\mathcal{I}}^{11}(\boldsymbol{\gamma})\right]^{-1}\boldsymbol{\mathcal{I}}^{11}(\boldsymbol{\gamma}) = \mathbf{I}_{m_j} \tag{3.58}$$

對於一個單位矩陣，$e_i = 1$。也就是說，三種檢定量在未受懲罰時會漸近自由度為m_j的卡方分配，這與傳統的檢定方法相符。值得注意的是，對於一個常見的檢定，$H_0: \boldsymbol{\gamma}_1 = \mathbf{0}$，分數檢定量無法被定義，因為本文考慮的懲罰函數無法在$\boldsymbol{\gamma}_1 = \mathbf{0}$上微分，故其受懲罰的分數向量與資訊矩陣皆無法被計算。除了校正的分數檢定量之外，本文將在第三章中詳細討論另外兩個校正過的統計檢定量犯型一錯誤的機率與檢定力。

三、基於拔靴法的參數檢定

除了基於卡方分配的檢定量（以下稱「未校正的檢定量」）與校正的檢定量之外，本文亦嘗試使用拔靴法來建立假設檢定的方法。針對未校正的檢定量，因為懲罰項的出現使得其真實分配未知，而拔靴法得以利用有限的樣本資料，經過多次的重複抽樣(resampling)之後，重新建立起足以代表母體樣本分配的新樣本。具體而言，本文根據每一個重複抽樣的樣本來計算未校正的檢定量，藉此近似未校正檢定量真實的分配，並取拔靴法分配的第$100 \times (1-\alpha)$分位數來作為拒絕的

臨界值。此為基於拔靴法的參數檢定方法。一般而言，拔靴法所提供的近似會比常用的極限近似來得精確，所以拔靴法由 Efron (1979)提出之後，即被大量地應用於統計分析當中。

令$F(t|\mathbf{z})$為存活時間T的分配函數，且$\hat{F}$為對應F的估計，而$G(t)$為設限時間C的分配函數，且$\hat{G}$為對應G的估計。對於一組存活分析資料$(Y_i, \delta_i, \mathbf{z}_i)$，$1 \le i \le n$，Burr (1994)考慮以下三種重複抽樣的方法：

方法一、直接對資料$(Y_i, \delta_i, \mathbf{z}_i), 1 \le i \le n$抽樣，每次抽樣時，每一個樣本點都有相同的機率被抽到（即抽出放回）。直到重複抽樣的樣本其大小為n時，停止。

方法二、對於每一個i，產生服從$\hat{F}(t|\mathbf{z}_i)$的T_i^*與服從$\hat{G}$的C_i^*，再計算$Y_i^* = \min\{T_i^*, C_i^*\}$與$\delta_i^* = I(T_i^* \le C_i^*)$，則得到重複抽樣樣本為$(Y_i^*, \delta_i^*, \mathbf{z}_i)$，$1 \le i \le n$。

方法三、對於每一個i，產生服從$\hat{F}(t|\mathbf{z}_i)$的T_i^*。若是$\delta_i = 0$，令$C_i^* = Y_i$。若是$\delta_i = 1$，產生服從$\hat{G}(t|C > Y_i) = P(C \le t|C > Y_i)$的$C_i^*$，也就是來自$[\hat{G}(t) - \hat{G}(Y_i)]/[1 - \hat{G}(Y_i)]$的分配。在計算$Y_i^* = \min\{T_i^*, C_i^*\}$與$\delta_i^* = I(T_i^* \le C_i^*)$之後，得到重複抽樣樣本為$(Y_i^*, \delta_i^*, \mathbf{z}_i)$，$1 \le i \le n$。

Burr (1994)指出，依據方法一所得的重複抽樣樣本來計算相關的統計量時，其分配為非條件的(unconditional)，也就是說，該分配是對共變量與設限時間的聯合分配平均而得。而依據方法二所得的重複抽樣樣本來計算相關的統計量時，其分配將受限於共變量$\mathbf{z}$。若是共變量的分配與模型參數無關，那麼共變量可被視為是一個輔助性統計量(ancillary statistic)。也就是說，所有的參數推論都應受限於輔助性統計量$\mathbf{z}$上。此外，Burr (1994)將輔助性統計量的原則加以推廣。若設限時間的分配也與模型參數無關，那麼參數推論除了受限於共變量之外，也應受限於設限時間。據此，方法三的抽樣方式即考慮了存活分析資料其特殊的機率結構。

根據 Burr (1994)的模擬，應用方法一所得的重複抽樣樣本來建立參數的信

賴區間時，雖然會最接近預定的覆蓋率(coverage probability)，但是其區間長度過長，反而不如方法二與方法三所建立的信賴區間。據此，在第三章中，本文將採用模擬的方式來判斷方法二或是方法三何種較適合本文所提出的模型。此外，在估計G時，本文採用的是乘積極限估計量(product-limit estimator) (Kaplan & Meier, 1958)，具體而言，

$$\hat{G}(t) = 1 - \prod_{j:y_j^c \le t} \left(1 - \frac{c_j}{n_j}\right) \tag{3.59}$$

其中，y_j^c為發生設限的時間點，n_j為在時間點y_j^c前所有可能被設限的人數，c_j為在時間點y_j^c時的設限人數。至於在估計F時，首先注意到F與累積危險函數Λ_0的關係式，

$$F(t|\mathbf{z}) = 1 - S(t|\mathbf{z}) \tag{3.60}$$

$$= 1 - \exp\left(-\int_0^t \lambda(s|\mathbf{z})ds\right) \tag{3.61}$$

$$= 1 - \left[\exp\left(-\int_0^t \lambda_0(s)ds\right)\right]^{\exp[\dot{\eta}(\mathbf{z})]} \tag{3.62}$$

$$= 1 - \left[\exp\left(-\Lambda_0(t)\right)\right]^{\exp[\dot{\eta}(\mathbf{z})]} \tag{3.63}$$

此處，本文採用 Breslow (1972)估計式來估計Λ_0，

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \frac{I(Y_i \le t)\delta_i}{\sum_{j=1}^{n} \delta_{ij}\, e^{\hat{\eta}(\mathbf{z}_j)}} \tag{3.64}$$

以此估計式代入式(3.63)，即得$F(t|\mathbf{z})$之估計式。

四、聯合信賴束

本節試圖以拔靴法為共變量函數$\dot{\eta}_j(z)$建構一個聯合的信賴束(simultaneous confidence band)。令$\hat{\dot{\eta}}_j(z)$為對應$\dot{\eta}_j(z)$的估計，而$\mathrm{se}\left(\hat{\dot{\eta}}_j(z)\right)$為此估計式的標準誤，則$100(1-\alpha)\%$信賴束可由此建構(Lenhoff et al., 1999)，

$$\hat{\eta}_j(z) \pm C_\alpha \cdot \mathrm{se}\left(\hat{\eta}_j(z)\right) \tag{3.65}$$

其中，C_α的選取應滿足

$$P\left(\max_z \left\{\frac{\left|\hat{\eta}_j(z) - \grave{\eta}_j(z)\right|}{\mathrm{se}\left(\hat{\eta}_j(z)\right)}\right\} \le C_\alpha\right) = 1 - \alpha \tag{3.66}$$

由於函數$\grave{\eta}_j$為真實的函數，無法直接確認其形態。此時可藉拔靴法的原理，將觀察到的樣本視為母體，所估計的$\hat{\eta}_j(z)$視為是真實的函數，並對此母體重複抽樣B次大小皆相同的樣本。在第b次的抽樣中，令$\hat{\eta}_j^b(z)$為對應$\hat{\eta}_j(z)$的估計，而$\mathrm{se}\left(\hat{\eta}_j^b(z)\right)$為此估計式的標準誤，則式(3.66)可藉下式來作近似，

$$\frac{1}{B}\sum_{b=1}^{B} I\left(\max_z \left\{\frac{\left|\hat{\eta}_j^b(z) - \hat{\eta}_j(z)\right|}{\mathrm{se}\left(\hat{\eta}_j^b(z)\right)}\right\} \le C_\alpha\right) = 1 - \alpha \tag{3.67}$$

以下詳細敘述如何計算估計式$\hat{\eta}_j$的標準誤。在估計$\grave{\eta}_j$時，本文引入了截距項c_j，利用式(3.11)，可將$\grave{\eta}_j$如下表示，

$$\grave{\eta}_j(z) = \sum_{t=1}^{q_j} (\gamma_{jt} + c_j)\mathcal{B}_t^{(j)}(z_j) \tag{3.68}$$

$$= \grave{\mathbf{b}}_j^{\mathrm{T}} \grave{\boldsymbol{\gamma}}_j \tag{3.69}$$

其中，$\grave{\boldsymbol{\gamma}}_j = (\gamma_{j1} + c_j \quad \cdots \quad \gamma_{j,q_j-1} + c_j \quad c_j)^{\mathrm{T}}$為加入截距項後的係數向量，而$\grave{\mathbf{b}}_j$為第$j$組完整的基底函數。因為截距項為該組基底係數的線性組合，如式(3.14)所示，故$\grave{\boldsymbol{\gamma}}_j$與$\boldsymbol{\gamma}_j$之間的關係仍為線性，

$$\grave{\boldsymbol{\gamma}}_j = \mathbf{A}\boldsymbol{\gamma}_j \tag{3.70}$$

具體而言，

$$\mathbf{A}_{q_j \times (q_j - 1)} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0}^{\mathrm{T}} \end{bmatrix} - \begin{bmatrix} \mathcal{B}_1^{(j)}(0)\mathbf{1} & \cdots & \mathcal{B}_{q_j-1}^{(j)}(0)\mathbf{1} \end{bmatrix} \tag{3.71}$$

此處，$\mathbf{I}$為維度$q_j - 1$的單位矩陣，$\mathbf{0}$為維度$q_j - 1$的零向量，而$\mathbf{1}$為維度q_j的向量，其元素皆為 1。令$\grave{\boldsymbol{\gamma}}_j$的估計共變異矩陣為$\boldsymbol{\Sigma}_j(\hat{\boldsymbol{\gamma}})$，則

$$\boldsymbol{\Sigma}_j(\hat{\boldsymbol{\gamma}}) = \mathbf{A}\mathbf{V}_j(\hat{\boldsymbol{\gamma}})\mathbf{A}^{\mathrm{T}} \tag{3.72}$$

其中，$\mathbf{V}_j(\hat{\boldsymbol{\gamma}})$為估計共變異矩陣$\mathbf{V}(\hat{\boldsymbol{\gamma}})$中對應第$j$組基底係數的部分，而$\mathbf{V}(\hat{\boldsymbol{\gamma}})$如式(3.51)所定義。因此，估計式$\hat{\eta}_j$的標準誤便以下式來作計算。

$$\mathrm{se}\left(\hat{\eta}_j(z)\right) = \left(\grave{\mathbf{b}}_j^{\mathrm{T}}\boldsymbol{\Sigma}_j(\hat{\boldsymbol{\gamma}})\grave{\mathbf{b}}_j\right)^{1/2} \tag{3.73}$$

國立政治大學
National Chengchi University

第三章　模擬與比較

本章將採用模擬的方式，在第一節比較 B 樣條比例危險模型與標準比例危險模型在估計上的誤差。在第二節呈現 Wald 檢定量、概似比檢定量及分數檢定量在懲罰函數出現時，懲罰係數的輕重將如何使檢定量的分配偏移。第三節檢視當傳統的檢定方法失效時，在特定的觀察樣本下以拔靴法生成的檢定量其分配是否具足夠的代表性。最後，第四節討論在進行參數推論時，校正過的檢定量與基於拔靴法的檢定量兩者的犯型一錯誤的機率α與檢定力$1-\beta$。

在產生資料時，本文以伯努利分配$\mathrm{Ber}(p=0.5)$來產生類別共變量，以均勻分配$\mathrm{Uni}(-\pi,\pi)$來產生連續共變量。關於設限時間的生成，則是採用指數分配$\mathrm{Exp}(\theta)$。其中，率參數θ的選取實際上是以試誤法(trial and error method)來決定，原則是選取使資料的設限率(censoring rate)不超過 10%的θ。此外，本文令基線危險函數λ_0為 1；在這個設定下，存活時間亦將服從率參數為$\exp[\eta(\mathbf{z})]$的指數分配。

在選取共變量j所對應的函數$\grave{\eta}_j$時，則以一個週期為2π的函數來生成。具體而言，本文以傅立葉級數(Fourier series)的基底$\{1, \cos x, \sin x, \cos 2x, \sin 2x, \dots\}$與一組隨機生成的係數$\{a_{j0}, a_{j1}, b_{j1}, a_{j2}, b_{j2}, \dots\}$的線性組合來代表$\grave{\eta}_j$，

$$\grave{\eta}_j(z) = a_{j0} + \sum_{n=1}^{\infty}\left(a_{jn}\sin nz + b_{jn}\cos nz\right) \tag{4.1}$$

且對每一個j，生成的$\grave{\eta}_j$皆滿足$\grave{\eta}_j = 0$的限制。此外，本文以函數的絕對值於區間$(-\pi,\pi)$上的平均值(average function value, AFV)來為所生成的函數分類，

$$\mathrm{AFV}(\grave{\eta}_j) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\grave{\eta}_j(z)\right|dz \tag{4.2}$$

當 AFV 愈接近零，表示基底係數愈小，對危險函數的影響也就愈不顯著；反之，其影響則愈顯著。這個分類將在第四節中有助於分析各個檢定量的檢定力。

第一節　與標準比例危險模型的比較

本節比較 B 樣條比例危險模型（下稱樣條模型）與標準比例危險模型（下稱標準模型）在估計上的誤差。首先，本文考慮以下五個模型，

$$\dot{\eta}^{(k)}(\mathbf{z}) = \beta_1 z_1 + \dot{\eta}_2^{(k)}(z_2), \qquad k = 1, 2, 3, 4, 5 \tag{4.3}$$

其中，z_1為類別共變量，與之對應的係數為$\beta_1 = 0.3$；z_2為連續共變量，與之對應的函數為$\dot{\eta}_2^{(k)}$，且$\mathrm{AFV}\left(\dot{\eta}_2^{(k)}\right) = 0.1k$。在使用樣條模型時，本文統一為各個函數$\dot{\eta}_2^{(k)}$配置 3 個節點，設階數為 4。模式如下，

$$\dot{\eta}^{(k)}(\mathbf{z}) = \beta_1 z_1 + \sum_{t=1}^{7}\left(\gamma_{2t}^{(k)} + c_2\right)\mathcal{B}_t^{(k)}(z_2), \qquad \forall k \tag{4.4}$$

而懲罰係數則依各個模擬資料利用交叉驗證作估計。因此，在樣條模型中，共有 6 個參數須作估計。在使用標準模型作估計時，本文加入z_2的高次方項來配適$\dot{\eta}_2^{(k)}$的非線性特質，其項次最高取至三次方。模式如下，

$$\dot{\eta}^{(k)}(\mathbf{z}) = \beta_1 z_1 + \sum_{t=1}^{3}\beta_{2t}^{(k)} z_2^t \ , \qquad \forall k \tag{4.5}$$

因此，在標準模型中，共有 4 個參數須作估計。至於樣本數的大小，本文考慮 20 倍、40 倍與 60 倍的樣條模型所需的參數個數，即$n = 120, 240, 360$。

表 3-1 與表 3-2 呈現了在模擬一千次之下，β_1估計的均方誤差(mean square error, MSE)與β_1估計值的標準差。就β_1估計的誤差而言，其誤差並不受z_2效果大小而有所影響，且隨著樣本數提高，其誤差下降是可預期的。在兩模型之間，誤差的差異不大。但就β_1估計的變異程度而言，在樣條模型中則顯得較穩定，這暗示了懲罰函數具穩定估計量的作用，而其所犧牲的不偏性也不至於使估計值離得真值太遠。表 3-3 呈現$\dot{\eta}_2^{(k)}$估計的積分均方誤差(mean integrated square error, MISE)。在樣條模型中，其誤差皆大幅低於標準模型的誤差。這顯示，即便標準模型容納了z_2的高次方項來考慮$\dot{\eta}_2^{(k)}$的非線性影響，在估計上與樣條模型相比仍

是不夠的。

圖 3-1 與圖 3-2 為一千次模擬當中的其中一次估計，圖 3-1 為以標準模型估計，而圖 3-2 為以樣條模型估計，兩者估計皆係建立在相同的模擬資料上。可以發現，樣條模型對非線性函數的配適較佳，因為其較能反映函數的曲折。在圖 3-2 中，灰色部分為以拔靴法建立的聯合信賴束，方法如第二章第三節所述。

表 3-1、各模型與各樣本數下β_1估計的 MSE

n	120		240		360	
k	標準	樣條	標準	樣條	標準	樣條
1	0.034	0.030	0.016	0.016	0.011	0.011
2	0.035	0.032	0.016	0.016	0.011	0.011
3	0.035	0.031	0.016	0.016	0.011	0.011
4	0.034	0.031	0.016	0.016	0.011	0.011
5	0.033	0.031	0.018	0.017	0.011	0.011

表 3-2、各模型與各樣本數下β_1估計值的標準差

n	120		240		360	
k	標準	樣條	標準	樣條	標準	樣條
1	0.1115	0.0989	0.0765	0.0767	0.0638	0.0631
2	0.1124	0.1044	0.0773	0.0768	0.0650	0.0645
3	0.1117	0.1029	0.0761	0.0757	0.0633	0.0638
4	0.1110	0.1024	0.0767	0.0760	0.0632	0.0628
5	0.1103	0.1057	0.0814	0.0782	0.0633	0.0635

表 3-3、各模型與各樣本數下$\hat{\eta}_2^{(k)}$估計的 MISE

n	120		240		360	
k	標準	樣條	標準	樣條	標準	樣條
1	0.244	0.085	0.130	0.072	0.092	0.057
2	0.449	0.266	0.328	0.184	0.285	0.139
3	1.444	0.568	1.287	0.263	1.268	0.180
4	0.862	0.562	0.785	0.258	0.747	0.177
5	2.279	0.688	2.121	0.392	2.075	0.301

圖 3-1、標準模型對共變量函數$\dot{\eta}_2^{(k)}$的估計

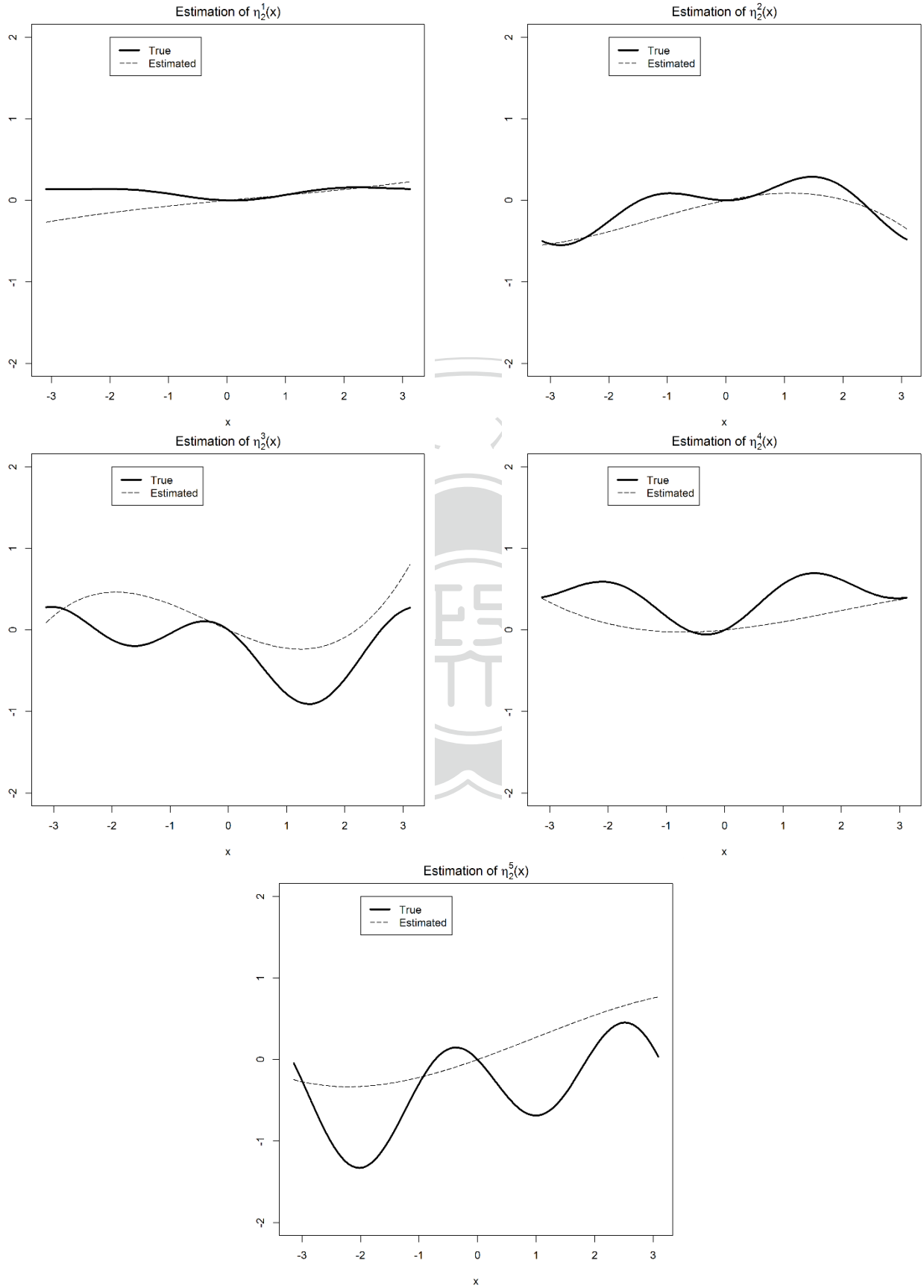
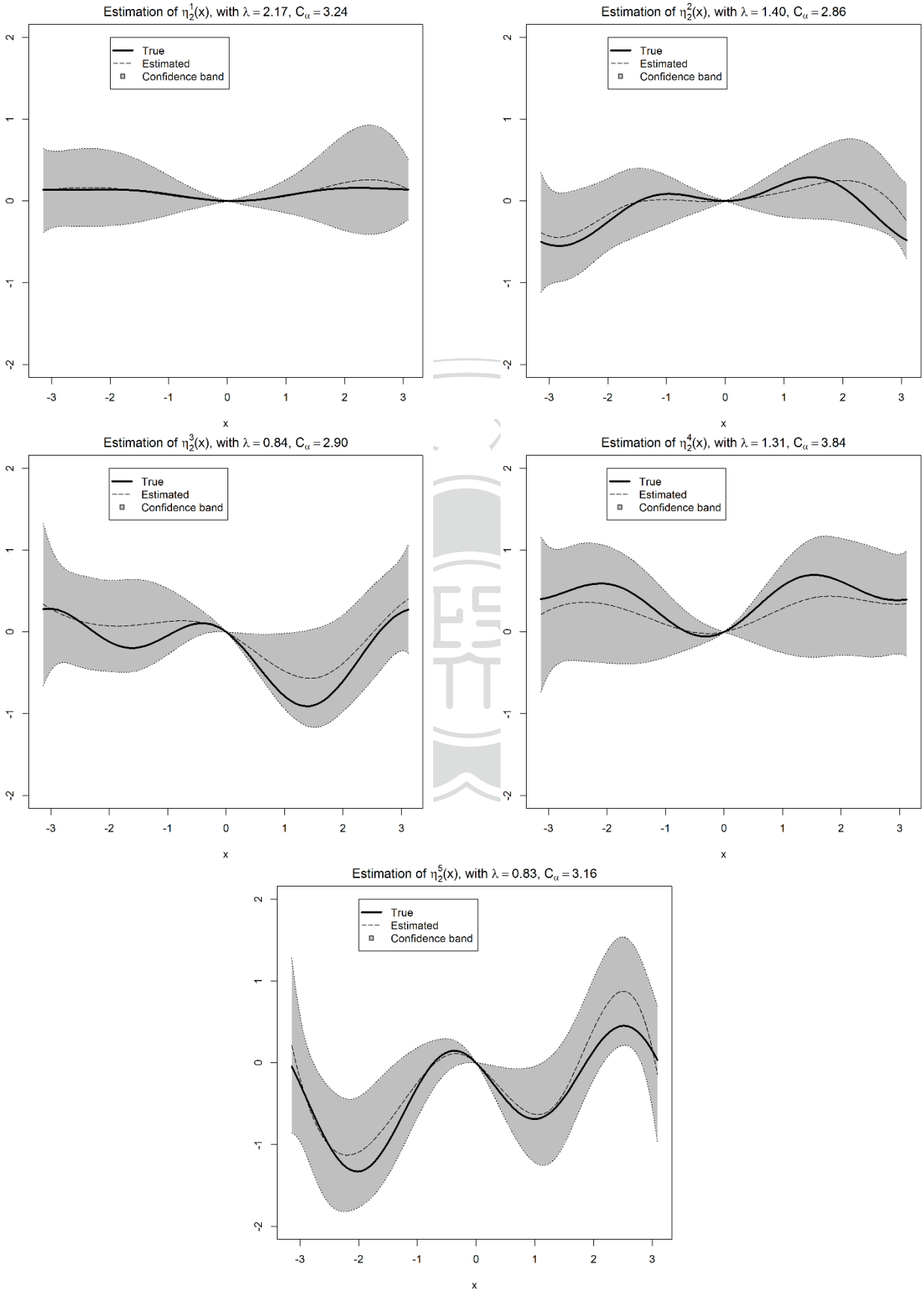


圖 3-2、樣條模型對共變量函數$\hat{\eta}_2^{(k)}$的估計



第二節　懲罰係數對檢定量的影響

本節討論懲罰係數將如何對三種檢定量的分配造成影響。當沒有懲罰函數出現時，各個檢定量在虛無假設下應近似卡方分配，其自由度為虛無假設與對立假設兩模型之間參數個數的差。但是當懲罰函數出現時，檢定量的分配便會隨懲罰係數的輕重而偏移。

為詳細呈現此現象，本文考慮以下三個模型：

$$\dot{\eta}^{(k)}(\mathbf{z}) = \dot{\eta}_1^{(k)}(z_1) + \dot{\eta}_2(z_2), \qquad k = 0, 1, 2 \tag{4.6}$$

其中，z_1與z_2皆為連續共變量，與之對應的函數為$\dot{\eta}_1^{(k)}$與$\dot{\eta}_2$，且$\mathrm{AFV}\left(\eta_1^{(0)}\right) = 0$、$\mathrm{AFV}\left(\dot{\eta}_1^{(1)}\right) = 0.3$、$\mathrm{AFV}\left(\dot{\eta}_1^{(2)}\right) = 0.5$，而$\mathrm{AFV}(\dot{\eta}_2) = 0.3$。考慮檢定$H_0: \boldsymbol{\gamma}_1 = \mathbf{0}$，即共變量$z_1$的參數為零。當虛無假設為真，模型即為$k = 0$；當虛無假設為偽，兩個對立假設下的模型分別為$k = 1, 2$。在近似各個函數時，本文統一配置 3 個節點，設階數為 4，共計 12 個參數。樣本數則定為 30 倍的參數個數，即$n = 360$。此外，本文考慮懲罰係數$\lambda = 0, 1, 2, 3$。在各個懲罰係數下，進行 500 次的模擬。在每一次模擬中，各個樣本的 Wald 檢定量、概似比檢定量及分數檢定量皆被記錄下來。

圖 3-3 至圖 3-5 為三個檢定量的分配，可以發現當懲罰係數為 0 時，三者在虛無假設下的分配的確近似自由度為 6 的卡方分配，在對立假設下的分配也都與虛無假設下的分配有所區隔，這表示此時應用卡方檢定具有一定的檢定力，也能將型一誤差控制住。然而，隨著懲罰係數加重，Wald 檢定量有縮小的趨勢，虛無假設下的分配尤為明顯。此時，若斷然執行卡方檢定，將造成不易拒絕虛無假設的情形，而顯得過於保守。概似比檢定量也呈現了類似的問題，雖然其縮小的幅度不如 Wald 檢定量嚴重。至於分數檢定量，其分配受懲罰係數影響最輕，大致上都能夠維持卡方分配的形狀，但須注意檢定量仍有輕微上升的現象。

若以交叉驗證的方法來估計懲罰係數，則模型$k = 0$的懲罰係數約落在 1.5，

模型$k=1$與模型$k=2$的懲罰係數約落在 1.0。此時，若採用傳統的檢定，則更有可能做出錯誤的決策。

圖 3-3、Wald 檢定量在各模型與各懲罰係數下的分配

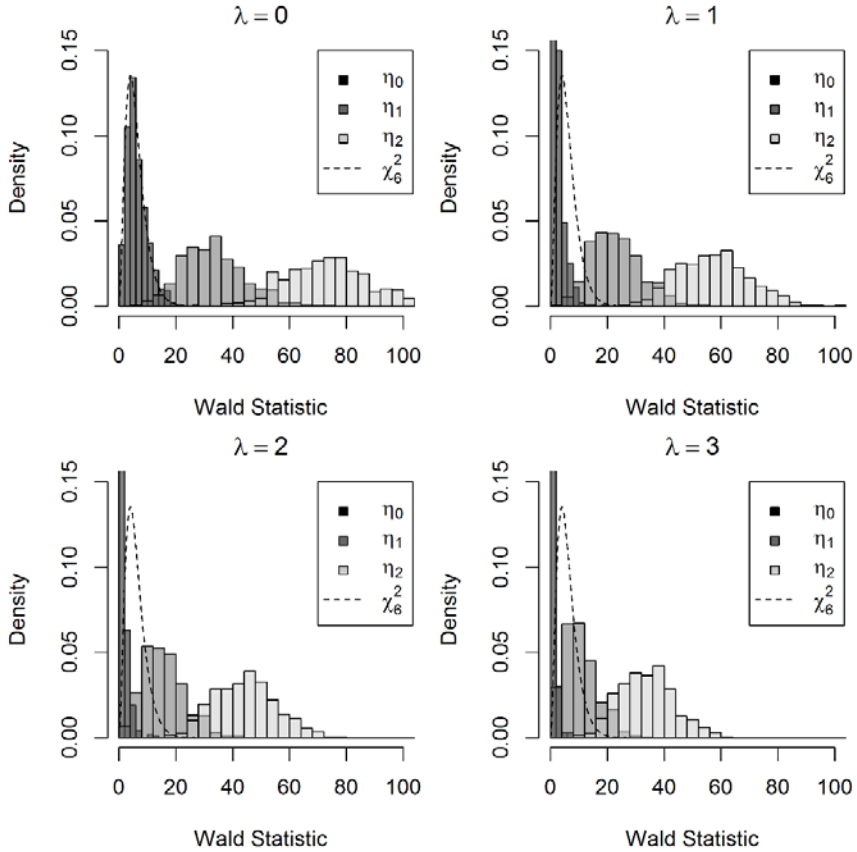


圖 3-4、概似比檢定量在各模型與各懲罰係數下的分配

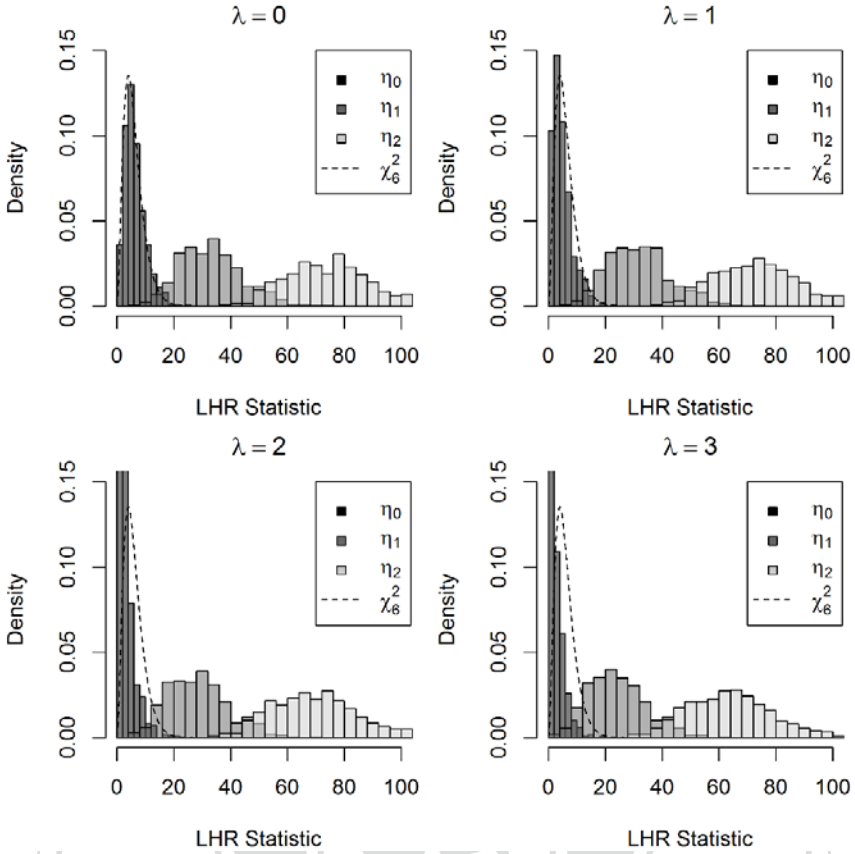
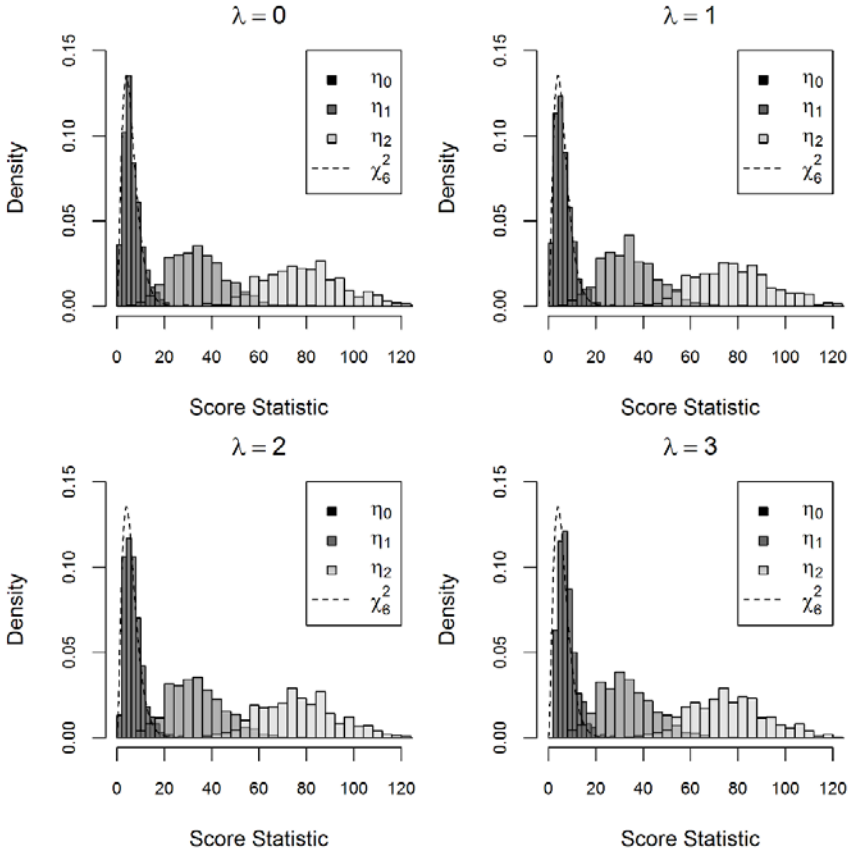


圖 3-5、分數檢定量在各模型與各懲罰係數下的分配



第三節　拔靴法樣本的代表性

本節以拔靴法對觀察到的特定樣本重複抽樣，抽樣方法採第三章所提的方法二與方法三，除了探討何種抽樣方法較適合本文所提之模型外，我們也檢視依拔靴法樣本計算的檢定量所形成的分配是否仍具代表性。首先，本文以虛無假設下的模型$\dot{\eta}^{(0)}(\mathbf{z})$來隨機生成一組大小為$n = 360$的樣本，並對此樣本重複抽樣$B = 500$次。在每一組樣本中，執行$H_0: \boldsymbol{\gamma}_1 = \mathbf{0}$的檢定，並記錄相關的檢定量。當拔靴法確實能夠模擬來自$\dot{\eta}^{(0)}(\mathbf{z})$的樣本時，依拔靴法樣本計算的檢定量其分配應相當接近前一節虛無假設下的分配形狀（以下稱前節之虛無假設下的檢定量分配為「模擬分配」，以拔靴法所生之檢定量分配為「拔靴法分配」）。

首先，本文比較方法二與方法三的優劣。表 3-4 呈現兩抽樣方法所得分配之各項統計量的比較，其中，Q 為百分位數。可以發現，不管懲罰係數的輕重為何，方法三的拔靴法分配其統計量皆較為接近模擬分配的統計量。若是採用拔靴法來作檢定並將型一誤差控制在$\alpha = 0.05$時，則必須關心拔靴法分配的第 95 個百分位數。此時，仍是方法三的百分位數皆較為接近模擬分配的百分位數。據此，在後續的分析當中，本文將採用方法三來作為標準的抽樣方法。

圖 3-6 至圖 3-8 為各檢定量以方法三抽樣的拔靴法分配，圖中虛線部分為以核密度估計(kernel density estimation)近似的模擬分配。在 Wald 檢定量分配中，拔靴法的檢定量幾乎都貼近虛線，除了在較重的懲罰係數下，中段部分的代表性略嫌薄弱。在概似比檢定量分配中，其分配則隨著懲罰係數增加而在前段部分出現高峰，此時，分數檢定量也有類似右偏的現象。大體而言，這些偏移並不太嚴重，因此本文決定採用拔靴法來近似這些檢定量的分配，並在第四節探討基於拔靴法的檢定量其表現優劣。

表 3-4、拔靴法方法二與方法三的比較

懲罰係數	統計量	Wald 檢定量			概似比檢定量			分數檢定量		
		模擬分配	方法二	方法三	模擬分配	方法二	方法三	模擬分配	方法二	方法三
0	25Q	3.81	3.48	3.50	3.80	3.47	3.52	3.83	3.49	3.51
	中位數	5.58	5.31	5.75	5.52	5.39	5.75	5.62	5.36	5.78
	平均數	6.45	5.98	6.40	6.39	5.96	6.34	6.52	6.05	6.47
	75Q	8.33	7.91	8.50	8.18	7.85	8.40	8.42	7.97	8.61
	95Q	13.75	12.37	14.16	13.64	12.34	13.60	14.01	12.53	14.33
1	25Q	0.91	0.72	0.73	2.32	2.02	1.95	3.74	3.51	3.45
	中位數	1.95	1.64	1.87	3.99	3.64	4.00	5.66	5.28	5.63
	平均數	2.51	2.26	2.43	4.77	4.36	4.62	6.43	5.96	6.34
	75Q	3.34	3.24	3.52	6.44	6.17	6.52	8.37	7.74	8.45
	95Q	7.06	6.14	6.50	11.78	10.00	10.80	13.79	12.10	13.70
2	25Q	0.19	0.11	0.13	1.19	1.03	0.97	4.09	3.57	3.46
	中位數	0.66	0.49	0.59	2.61	2.26	2.61	6.33	5.30	5.53
	平均數	1.16	0.98	1.07	3.38	2.99	3.20	6.93	5.94	6.29
	75Q	1.45	1.41	1.45	4.65	4.28	4.55	8.82	7.80	8.53
	95Q	4.09	3.25	3.67	9.33	7.80	8.37	14.16	12.22	13.21
3	25Q	0.04	0.01	0.03	0.66	0.43	0.75	5.14	3.60	3.47
	中位數	0.19	0.16	0.20	1.68	1.62	1.84	7.04	5.30	5.57
	平均數	0.58	0.50	0.57	2.46	2.25	2.52	7.84	5.94	6.27
	75Q	0.80	0.66	0.72	3.44	3.30	3.72	9.43	7.78	8.52
	95Q	2.35	1.90	2.22	7.44	6.52	7.01	14.96	12.42	13.23

國立政治大學 National Chengchi University

圖 3-6、Wald 檢定量的拔靴法分配

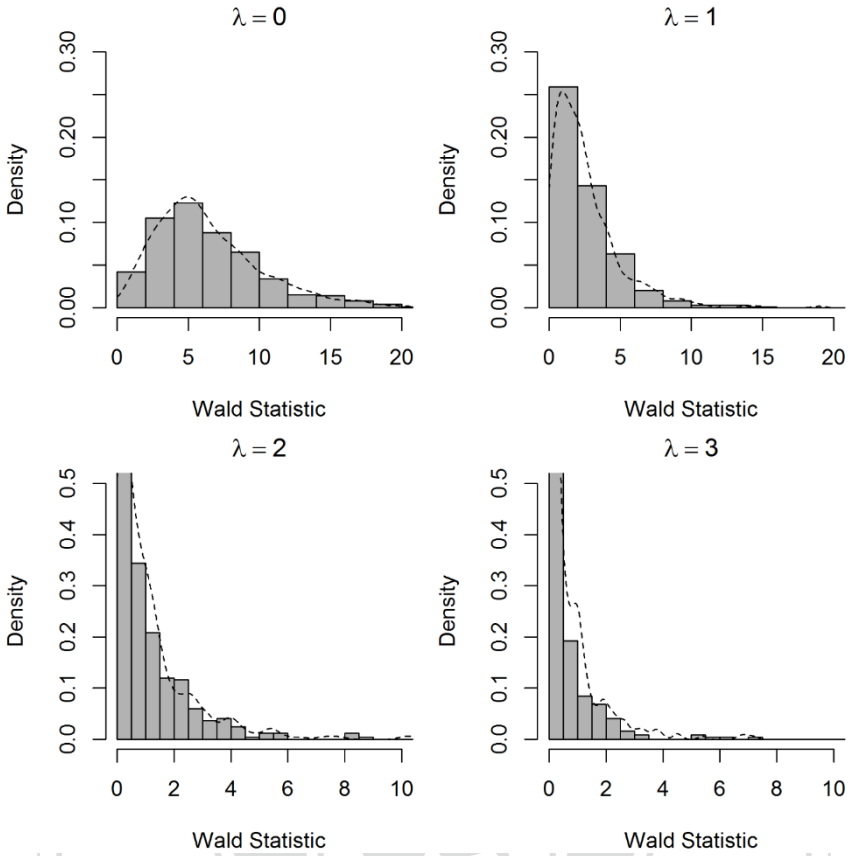


圖 3-7、概似比檢定量的拔靴法分配

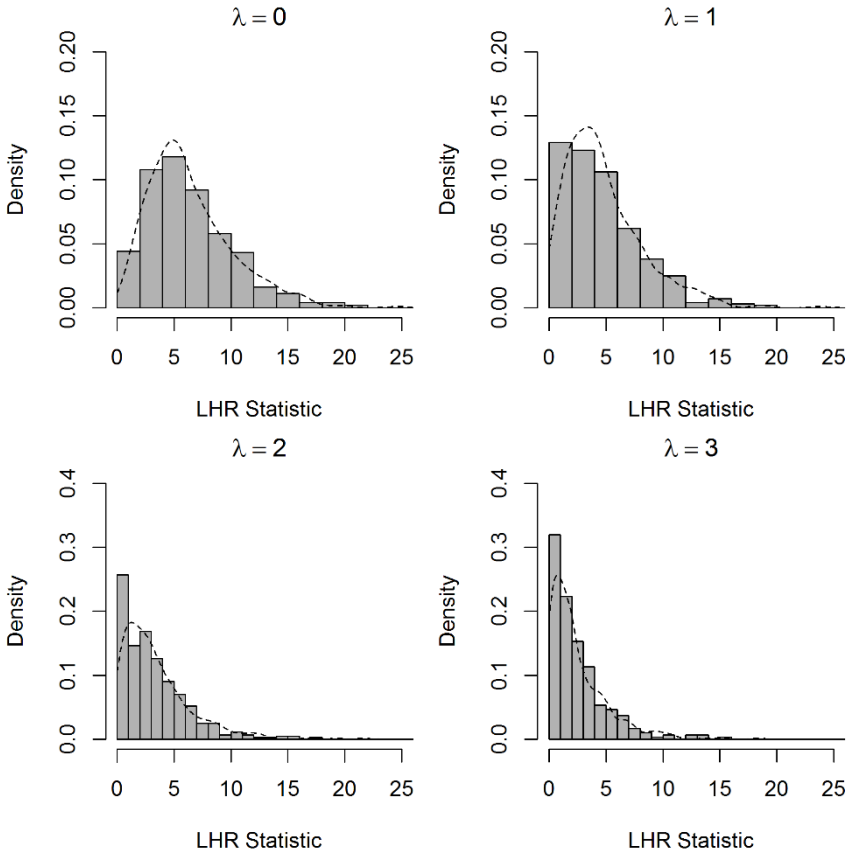
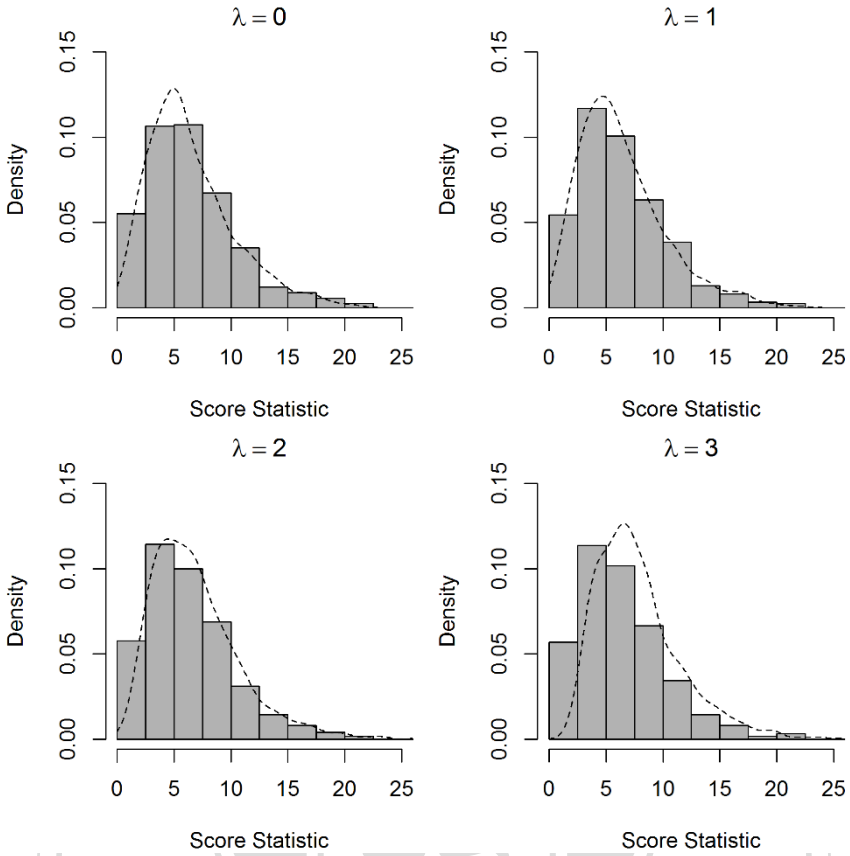


圖 3-8、分數檢定量的拔靴法分配



National Chengchi University

第四節　型一錯誤與檢定力

本節比較各個檢定量的型一錯誤與檢定力，考慮的檢定量共有 Wald 檢定量、概似比檢定量、分數檢定量，各分為未校正、校正（但是校正的分數檢定量不在其中）。未校正的檢定量即為建立在卡方分配的傳統方法，而校正的檢定量則考慮了懲罰項的影響，在計算檢定量時皆以受懲罰的分數向量或資訊矩陣來計算，但其分配已不再近似於卡方分配。針對未校正的檢定量，因為不確定其分配為何，則採用拔靴法來近似其真正的分配，並取拔靴法分配的第$100 \times (1-\alpha)$百分位數來作為拒絕的臨界值。此外，拔靴法抽樣的次數也被列入比較當中。

考慮以下六個模型：

$$\dot{\eta}^{(k)}(\mathbf{z}) = \dot{\eta}_1^{(k)}(z_1) + \dot{\eta}_2(z_2), \qquad k = 0,1,2,3,4,5 \tag{4.7}$$

其中，z_1與z_2皆為連續共變量，與之對應的函數為$\dot{\eta}_1^{(k)}$與$\dot{\eta}_2$，且$\mathrm{AFV}\left(\dot{\eta}_1^{(k)}\right) = 0.1k$，而$\mathrm{AFV}(\dot{\eta}_2) = 0.3$。考慮檢定$H_0: \boldsymbol{\gamma}_1 = \mathbf{0}$，即共變量$z_1$的參數為零。當虛無假設為真，模型即為$k = 0$；當虛無假設為偽，五個對立假設下的模型分別為$k = 1,2,3,4,5$。在近似各個函數時，本文統一配置 3 個節點，設階數為 4，共計 12 個參數。樣本數則定為 10 倍、20 倍與 30 倍的參數個數，即$n = 120, 240, 360$。此外，本文考慮不同的懲罰係數$\lambda = 1,2,3$與不同的拔靴法抽樣次數$B = 50, 100$。在各種情境的組合之下，進行 500 次的模擬。在每一次模擬中，各個樣本的檢定量皆被記錄下來。

表 3-5 至表 3-7 為各種樣本數與懲罰係數之下的模擬結果。首先，我們討論各個檢定量是否將型一錯誤控制得宜。在未校正的檢定量方面，Wald 與概似比檢定量皆將型一錯誤控制在$\alpha = 0.05$之下，但 Wald 檢定量顯得比概似比檢定量來得保守，此外，兩者犯型一錯誤的機率皆隨著懲罰係數提升而下降。至於分數檢定量，則較無法控制犯型一錯誤的機率，尤其是在較重的懲罰係數下時，但不管懲罰係數為何，增加樣本數皆有助於降低犯型一錯誤的機率。在校正的檢定量方面，Wald 與概似比檢定量兩者犯型一錯誤的機率皆不相上下，雖然都在$\alpha =$

0.05之下，但其犯型一錯誤的機率同樣地隨著懲罰係數提升而下降。這暗示了不管校正與否，在較重的懲罰係數下，應用 Wald 與概似比檢定皆可能過於保守。在拔靴法檢定量方面，抽樣次數增加在小樣本中似乎有助於降低型一錯誤，這點在懲罰係數較重時尤為明顯。但整體而言，拔靴法檢定量較難以控制型一錯誤，且隨著懲罰係數提升，益加無法控制。只有在樣本數大時，其型一錯誤才可能降至 0.10 之下，其中，又以分數檢定量較得以控制。

根據對各檢定量控制型一錯誤的分析，本文選擇三個表現較佳的檢定量，進一步分析他們的檢定力。此處考慮的檢定量為未校正的概似比檢定量、校正的 Wald 檢定量，與基於拔靴法的分數檢定量，並將其檢定結果繪製如圖 3-9，以利比較。在各懲罰係數與樣本數的組合之下，拔靴法檢定量皆具最高的檢定力，然而有時不一定能夠控制住型一錯誤的機率，尤其是在小樣本與高懲罰係數之下，其犯型一錯誤的機率已高達 0.2 以上，此時應增加拔靴法的抽樣次數，除卻這樣較不理想的情境，其在檢定力的表現皆優於另外兩者。至於校正的 Wald 檢定量與未校正的概似比檢定量，又以未校正的概似比檢定表現較佳，因為兩者犯型一錯誤的機率皆相差不遠，但是未校正的概似比檢定在各情境的檢定力皆大幅高於校正的 Wald 檢定量。

國立政治大學 National Chengchi University

表 3-5、懲罰係數為 1 之下各檢定量之型一錯誤與檢定力

檢定量	$\lambda = 1$	型一錯誤			檢定力														
	AFV_1, AFV_2	(0.0, 0.3)			(0.1, 0.3)			(0.2, 0.3)			(0.3, 0.3)			(0.4, 0.3)			(0.5, 0.3)		
	n	120	240	360	120	240	360	120	240	360	120	240	360	120	240	360	120	240	360
Wald	拔靴法50	0.094	0.058	0.084	0.128	0.212	0.294	0.320	0.562	0.798	0.612	0.934	0.992	0.894	0.998	1.000	0.972	1.000	1.000
	拔靴法100	0.078	0.060	0.090	0.126	0.176	0.278	0.274	0.522	0.740	0.564	0.918	0.980	0.858	0.994	1.000	0.962	1.000	1.000
	校正	0.014	0.016	0.020	0.022	0.058	0.100	0.070	0.280	0.568	0.220	0.762	0.958	0.516	0.968	1.000	0.778	0.996	1.000
	未校正	0.000	0.004	0.006	0.000	0.010	0.032	0.010	0.128	0.382	0.060	0.524	0.890	0.194	0.922	1.000	0.496	0.992	1.000
概似比	拔靴法50	0.082	0.062	0.084	0.122	0.190	0.266	0.300	0.554	0.788	0.598	0.932	0.986	0.884	0.998	1.000	0.972	1.000	1.000
	拔靴法100	0.072	0.054	0.080	0.114	0.166	0.254	0.268	0.518	0.738	0.548	0.916	0.980	0.846	0.992	1.000	0.958	1.000	1.000
	校正	0.014	0.012	0.018	0.024	0.060	0.102	0.064	0.286	0.564	0.226	0.766	0.960	0.528	0.968	1.000	0.786	0.996	1.000
	未校正	0.022	0.024	0.038	0.038	0.090	0.176	0.150	0.400	0.678	0.390	0.886	0.978	0.734	0.990	1.000	0.930	0.998	1.000
分數	拔靴法50	0.070	0.052	0.070	0.110	0.178	0.258	0.292	0.520	0.788	0.576	0.926	0.984	0.892	0.996	1.000	0.974	1.000	1.000
	拔靴法100	0.060	0.054	0.066	0.110	0.150	0.240	0.260	0.492	0.738	0.530	0.918	0.974	0.842	0.992	1.000	0.960	1.000	1.000
	未校正	0.106	0.060	0.054	0.158	0.194	0.256	0.360	0.548	0.762	0.646	0.944	0.982	0.904	0.998	1.000	0.980	1.000	1.000

National Chengchi University

表 3-6、懲罰係數為 2 之下各檢定量之型一錯誤與檢定力

檢定量	$\lambda = 2$	型一錯誤			檢定力														
	AFV_1, AFV_2	(0.0, 0.3)			(0.1, 0.3)			(0.2, 0.3)			(0.3, 0.3)			(0.4, 0.3)			(0.5, 0.3)		
	n	120	240	360	120	240	360	120	240	360	120	240	360	120	240	360	120	240	360
Wald	拔靴法50	0.168	0.094	0.086	0.226	0.232	0.282	0.384	0.538	0.758	0.578	0.920	0.984	0.818	0.998	1.000	0.950	1.000	1.000
	拔靴法100	0.134	0.098	0.090	0.202	0.222	0.278	0.350	0.532	0.746	0.590	0.914	0.982	0.844	0.994	1.000	0.964	1.000	1.000
	校正	0.000	0.006	0.010	0.002	0.018	0.044	0.018	0.114	0.358	0.064	0.466	0.838	0.156	0.872	0.998	0.390	0.980	1.000
	未校正	0.000	0.000	0.000	0.000	0.002	0.006	0.000	0.014	0.116	0.000	0.178	0.612	0.016	0.586	0.976	0.048	0.914	1.000
概似比	拔靴法50	0.160	0.072	0.080	0.202	0.208	0.258	0.348	0.530	0.750	0.574	0.918	0.984	0.820	0.998	1.000	0.950	1.000	1.000
	拔靴法100	0.122	0.078	0.078	0.186	0.196	0.262	0.334	0.516	0.738	0.576	0.914	0.976	0.850	0.992	1.000	0.964	1.000	1.000
	校正	0.000	0.006	0.008	0.000	0.018	0.046	0.018	0.118	0.368	0.068	0.466	0.840	0.166	0.872	0.998	0.404	0.980	1.000
	未校正	0.006	0.008	0.012	0.012	0.050	0.084	0.044	0.238	0.534	0.154	0.720	0.956	0.372	0.962	1.000	0.658	0.996	1.000
分數	拔靴法50	0.134	0.068	0.072	0.182	0.186	0.250	0.324	0.512	0.756	0.544	0.910	0.980	0.806	0.998	1.000	0.940	1.000	1.000
	拔靴法100	0.108	0.072	0.074	0.168	0.188	0.256	0.316	0.488	0.728	0.558	0.908	0.976	0.836	0.990	1.000	0.958	1.000	1.000
	未校正	0.200	0.088	0.086	0.264	0.216	0.270	0.462	0.544	0.748	0.700	0.936	0.980	0.922	0.998	1.000	0.984	1.000	1.000

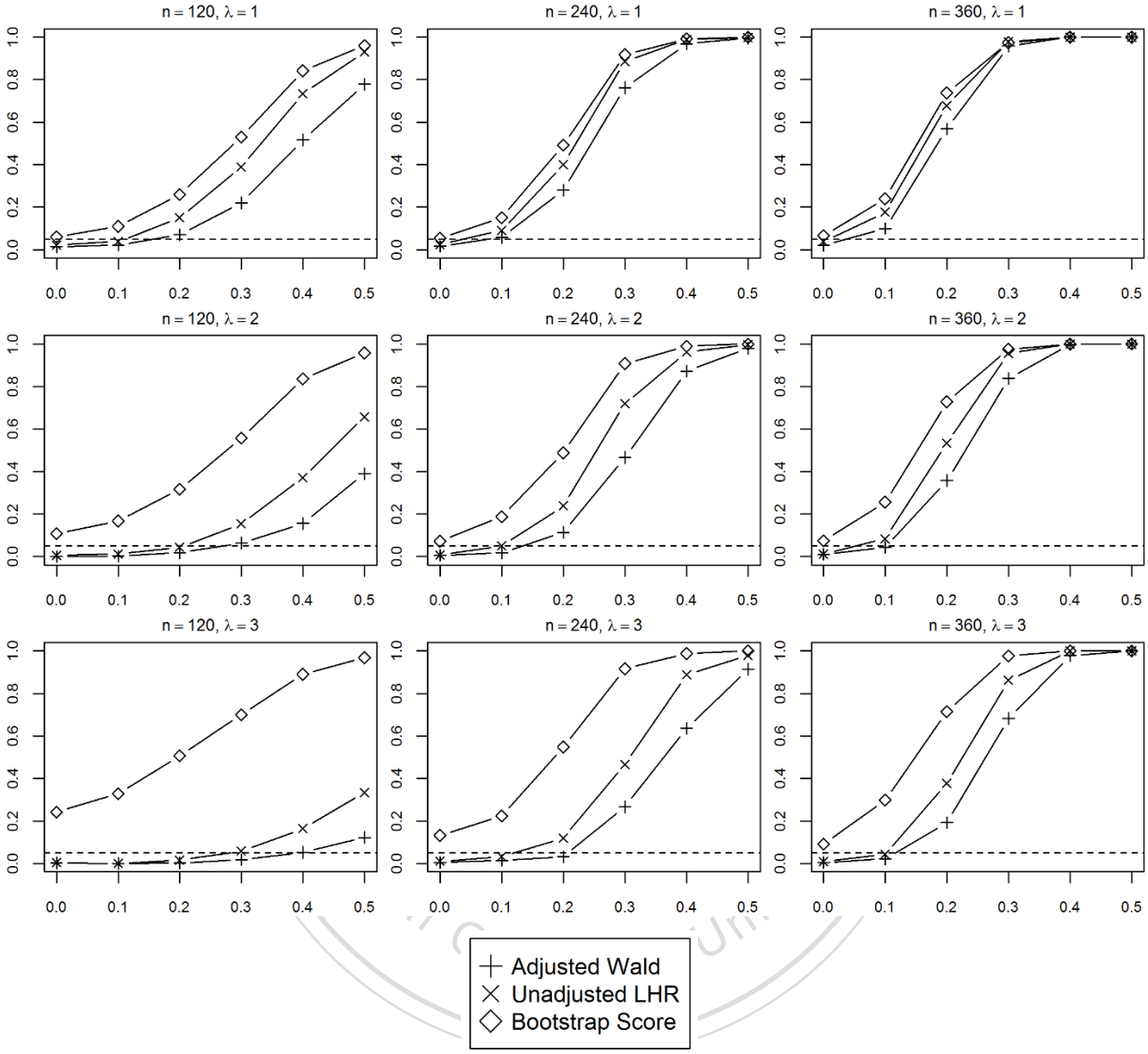
Chengchi Univ

表 3-7、懲罰係數為 3 之下各檢定量之型一錯誤與檢定力

檢定量	$\lambda = 3$	型一錯誤			檢定力														
	AFV_1, AFV_2	(0.0, 0.3)			(0.1, 0.3)			(0.2, 0.3)			(0.3, 0.3)			(0.4, 0.3)			(0.5, 0.3)		
	n	120	240	360	120	240	360	120	240	360	120	240	360	120	240	360	120	240	360
Wald	拔靴法50	0.590	0.166	0.134	0.638	0.322	0.318	0.762	0.566	0.742	0.798	0.864	0.978	0.900	0.992	1.000	0.942	0.996	1.000
	拔靴法100	0.278	0.152	0.136	0.370	0.302	0.358	0.524	0.588	0.746	0.736	0.920	0.976	0.886	0.992	1.000	0.968	1.000	1.000
	校正	0.002	0.004	0.004	0.000	0.014	0.022	0.002	0.032	0.192	0.018	0.266	0.682	0.052	0.636	0.978	0.122	0.914	1.000
	未校正	0.000	0.000	0.000	0.000	0.002	0.002	0.000	0.002	0.024	0.000	0.026	0.308	0.000	0.206	0.846	0.002	0.614	0.992
概似比	拔靴法50	0.548	0.140	0.112	0.628	0.282	0.284	0.766	0.544	0.736	0.798	0.862	0.976	0.900	0.988	1.000	0.946	0.998	1.000
	拔靴法100	0.274	0.144	0.110	0.340	0.262	0.318	0.504	0.554	0.726	0.712	0.918	0.976	0.902	0.992	1.000	0.964	1.000	1.000
	校正	0.000	0.004	0.002	0.000	0.012	0.022	0.002	0.034	0.192	0.018	0.264	0.678	0.054	0.644	0.980	0.124	0.914	1.000
	未校正	0.004	0.010	0.008	0.000	0.032	0.044	0.016	0.120	0.376	0.058	0.466	0.862	0.164	0.888	1.000	0.334	0.980	1.000
分數	拔靴法50	0.514	0.134	0.094	0.596	0.264	0.288	0.704	0.524	0.724	0.798	0.852	0.974	0.884	0.988	1.000	0.928	0.998	1.000
	拔靴法100	0.242	0.132	0.092	0.328	0.224	0.298	0.506	0.548	0.714	0.700	0.916	0.976	0.890	0.988	1.000	0.968	1.000	1.000
	未校正	0.466	0.160	0.120	0.502	0.310	0.320	0.676	0.618	0.742	0.834	0.932	0.980	0.964	0.998	1.000	0.990	1.000	1.000

National Chengchi University

圖 3-9、各檢定量在各情境下的型一錯誤與檢定力



第四章　結論與建議

本文討論了當比例危險模型的線性假設不成立時，使用 B 樣條基底函數來近似共變量的函數是可行的。根據本文的模擬結果，在估計一個非線性的共變量函數時，與標準的比例危險模型相比，B 樣條基底模型能夠大幅降低估計上的誤差。就估計的變異程度而言，B 樣條基底模型因為引入了懲罰函數，模擬結果亦顯示該模型的估計值具較高的穩定性。

此外，本文亦探討了 B 樣條基底模型下的假設檢定。本文考慮的檢定量除了傳統的 Wald 檢定量、概似比檢定量與分數檢定量之外，還考慮了 Gray (1994)所提出的，因應懲罰函數而作修正的檢定量，與基於拔靴法樣本建立的檢定量。模擬結果顯示，雖然基於拔靴法的檢定量擁有較高的檢定力，但是在樣本數小與懲罰係數高的時候，將無法有效控制住型一錯誤。此時，提高重複抽樣次數將使犯型一錯誤的機率降低，但是此方法需要龐大的計算量。在這些檢定量中，校正的檢定量表現是最差的，本文猜測這是因為 group lasso 考慮的懲罰函數並非一個良態函數(well-behaved function)所致。由於該函數並非每一點皆可微分，所以用來計算檢定量的受懲罰的分數矩陣與資訊矩陣並非在每一點皆可定義。而在參數較多的情況下，計算逆矩陣的近似也會招致許多計算誤差。

關於 lasso 估計量的檢定，Lockhart, Taylor, Tibshirani, & Tibshirani (2014)在迴歸模型的架構下發展了一個共變異檢定量。該檢定量服從一個率參數為 1 的指數分配。文末，Lockhart et al. (2014)亦將此檢定量應用在一般的線性迴歸模型中，包含了羅吉斯迴歸與標準的比例危險模型，其模擬結果顯示共變異檢定量一樣適用於一般線性模型（但是此結果尚缺乏理論基礎）。由於本文考慮的是一個 B 樣條基底模型，與標準的比例危險模型迥異，經模擬測試（結果未於本文中顯示），發現此檢定量的確服從指數分配，但是其率參數則不如該文所述。因此，在 B 樣條基底模型中發展共變異檢定量來取代須高度計算的拔靴法檢定量為未來可進行之方向。

參考文獻

[1] Bøvelstad, H. M., Nygård, S., Størvold, H. L., Aldrin, M., Borgan, Ø., Frigessi, A., and Lingjærde, O. C. (2007). Predicting survival from microarray data—a comparative study. *Bioinformatics* **23** (16), 2080-2087.

[2] Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Prentice Hall.

[3] Breslow, N. E. (1972). Contribution to the discussion of paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B* **34**, 216-217.

[4] Breslow, N. E. and Crowley, J. (1974). A large-sample study of the life table and product limit estimates under random censorship. *Annals of Statistics* **2**, 437-454.

[5] Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *Journal of the American Statistical Association* **89**, 1290-1302.

[6] Cox, D. R. (1972). Regression models and life-table (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187-220.

[7] Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

[8] Curry, H. B. and Schoenberg, I. J. (1966). On Pólya Frequency Functions IV: The fundamental spline functions and their limits. *Journale d'Analyse Mathématique* **17**, 71-107.

[9] de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

[10] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association* **72**, 557-565.

[11] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1-26.

[12] Gray, R. J. (1992). Flexible methods for analyzing survival data using splines,

國立政治大學 National Chengchi University

with applications to breast cancer prognosis. *Journal of the American Statistical Association* **87**, 942-951.

[13] Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics* **50**, 640-652.

[14] Hastie, T. and Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* **46**, 1005-1016.

[15] Huang, J. Z. and Liu, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics* **62**, 793-802.

[16] Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60**, 267-278.

[17] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481.

[18] Keele, L. (2010). Proportionally difficult: testing for nonproportional hazards in Cox models. *Political Analysis* **18**, 189-205.

[19] Kim, J., Sohn, I., Jung, S. H., Kim, S., and Park, C. (2012). Analysis of survival data with group Lasso. *Communications in Statistics—Simulation and Computation* **41**, 1593-1605.

[20] Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995). Hazard regression. *Journal of the American Statistical Association* **90**, 78-94.

[21] LeBlanc, M. and Crowley, J. (1999). Adaptive regression splines in the Cox model. *Biometrics* **55**, 204-213.

[22] Lenhoff, M. W., Santner, T. J., Otis, J. C., Peterson, M. G. E., Williams, B. J., and Backus, S. I. (1999). Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data. *Gait & Posture* **9** (1), 10-17.

[23] Li, W., Xu, S., Zhao, G., and Goh, L. P. (2005). Adaptive knot placement in *B*-

國立政治大學 National Chengchi University

spline curve approximation. *Computer-Aided Design* **37**, 791-797.

[24] Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming* **45**, 503-528.

[25] Lockhart, R., Taylor, J., Tibshirani, R., and Tibshirani, R. J. (2014). A significance test for the lasso (with discussion). *Annals of Statistics* **42** (2), 413-468.

[26] Meier, L., van de Geer, S., and Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **70**, 53-71.

[27] Moody, J. E., Hanson, S. J., and Lippmann, R. P. (1992). The effective number of parameters: an analysis of generalization and regularization in nonlinear learning system. *Advances in Neural Information Processing System* **4**, 847-854.

[28] O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM Journal on Scientific and Statistical Computing* **9**, 531-542.

[29] Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association* **85**, 941-949.

[30] Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics* **13**, 689-705.

[31] Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* **12** (1), 156-175.

[32] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society Series B-Methodological* **58** (1), 267-288.

[33] Tibshirani, R. (1997). The LASSO method for variable selection in the Cox model. *Statistics in Medicine* **16** (4), 385-395.

國立政治大學 National Chengchi University

[34] Verweij, P. J. M. and van Houwelingen, H. C. (1993). Cross-validation in survival analysis. *Statistics in Medicine* **12**, 2305-2314.

[35] Wold, S. (1974). Spline functions in data analysis. *Technometrics* **16**, 1-11.

[36] 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* **86** (1), 49-67.

國立政治大學
National Chengchi University