

國立中央大學

統計研究所

碩士論文

廣義伽瑪分布加速失敗時間模型之研究

研究生：李克耘

指導教授：陳玉英 博士

中華民國 九十七 年 六 月



國立中央大學圖書館

碩博士論文電子檔授權書

(95 年 7 月最新修正版)

本授權書所授權之論文全文電子檔(不包含紙本、詳備註 1 說明)，為本人於國立中央大學，撰寫之碩/博士學位論文。(以下請擇一勾選)

- ()**同意** (立即開放)
- ()**同意** (一年後開放)，原因是：_____
- ()**同意** (二年後開放)，原因是：_____
- ()**不同意**，原因是：_____

以非專屬、無償授權國立中央大學圖書館與國家圖書館，基於推動「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、公開陳列、與發行，或再授權他人以各種方法重製與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名：_____ 李克耘 _____ 學號：_____ 952205018 _____

論文名稱：_____ 廣義伽瑪分布加速失敗時間模型之研究 _____

指導教授姓名：_____ 陳玉英 博士 _____

系所：_____ 統計研究 _____ 所 博士班 碩士班

日期：民國_97_年_6_月_30_日

備註：

1. 本授權書之授權範圍僅限電子檔，紙本論文部分依著作權法第 15 條第 3 款之規定，採推定原則即預設同意圖書館得公開上架閱覽，如您有申請專利或投稿等考量，不同意紙本上架陳列，須另行加填聲明書，詳細說明與紙本聲明書請至 <http://blog.lib.ncu.edu.tw/plog/> 碩博士論文專區查閱下載。
2. 本授權書請填寫並親筆簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
3. 請加印一份單張之授權書，填寫並親筆簽名後，於辦理離校時交圖書館（以統一代轉寄給國家圖書館）。
4. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

摘要

右設限存活資料分析中的參數化加速失敗時間模型，經常假設其存活分布為韋伯分布、對數常態分布或對數邏輯司分布，當共變數存在時，又多展示為對數線性模式。本文考慮的參數化加速失敗時間模式之存活分布為廣義伽瑪分布，因為此一分布包含韋伯分布、對數常態分布及伽瑪分布等，所能涵蓋的右偏分布較為廣泛。此外，本文利用區域概似估計(local likelihood estimation)研究對數線性模式中，共變數的可能轉換。之後本文用模擬研究本文所提比較廣義的模型在各種不同加速失敗時間模型及存活分布下，估計中位存活時間的偏誤及均方誤。最後，本文藉一筆資料說明所提模型之應用。

關鍵字：右設限存活資料、加速失敗時間模型、廣義伽瑪分布、區域概似估計

Abstract

The parametric accelerated failure time(AFT) model for analyzing the right-censored survival data usually assumes Weibull, lognormal or log-logistic distribution for the lifetime variable of interest. When the covariates are presented, the AFT model is often expressed log-linear form. In this study, however we consider the parametric accelerated failure time model with lifetime which is distributed for a generalized gamma distribution, including, in particular, Weibull, lognormal and gamma distributions. In addition, we use local likelihood estimation to investigate the possible transformation of the covariates that showing in the log-linear model. A simulation study is further implemented for evaluating the bias and mean squares error of the estimated medians for a variety of AFT models and survival distributions. Finally, we use a data to illustrate the application of the model, under study.

Key word : right-censored, accelerated failure time model, generalized gamma distribution, local likelihood estimation

致謝詞

一年多來在陳玉英教授耐心的指導下，讓我在這段時間內成長許多，也讓我學習到處理事情的態度與規劃自己的時間，在此由衷的感謝老師的指導與鼓勵，另外也感謝口試委員，嵇允嬋教授與張玉媚教授，給予我的論文許多寶貴的意見，也要感謝啟軒學長與文明學長，在我有問題時能夠耐心指導我。

在撰寫論文的日子里，也感謝碩二同學們大家在課業與生活上的幫忙與解壓，也感謝我的家人，爸媽與老姊，讓我能在不需要顧慮經濟壓力下，完成我的學業。最後也感謝小福在我心情低落時幫我加油打氣。

在研究所兩年裡，從一開始害怕碩二的上台 meeting 與書報討論，到認真勇敢地面對所有事情，現在我能自信的對自己說我進步了，也不會抱著鴛鴦的心態逃避問題，最後再次感謝在這一年多裡耐心指導我的陳玉英教授，老師謝謝您，謝謝您的指導。

僅將本文獻給摯愛的父母、親友、以及所有愛護關心我的人

李克耘 謹識於

國立中央大學統計研究所

中華民國九十七年六月

目錄

第一章 緒論	1
1.1 研究動機與目的.....	1
1.2 文獻回顧.....	2
1.2.1 加速失敗時間模型.....	2
1.2.2 廣義伽瑪分布.....	3
1.2.3 廣義伽瑪加速失敗時間模型.....	5
1.2.4 區域概似估計.....	7
第二章 統計方法	11
廣義伽瑪半參數加速失敗時間模型.....	11
第三章 模擬研究	14
3.1 模擬方法.....	14
3.2 模擬結果.....	15
第四章 實例分析	18
第五章 結論及未來研究	21
參考文獻	22

圖目錄

圖一 四種風險函數型態對應之參數圖·····	24
圖二 模擬中三種對數時間模型態·····	24
圖三 $\sigma=0.6, \lambda=0.3, x=1$ 時模擬中三種對數時間模型的機率密度函數·····	25
圖四 $\sigma=0.6, \lambda=1, x=1$ 時模擬中三種對數時間模型的機率密度函數·····	26
圖五 $\sigma=0.6, \lambda=0.3, x=2$ 時模擬中三種對數時間模型的機率密度函數·····	27
圖六 五個共變數分布之直方圖·····	28
圖七 肝硬化資料之 Kaplan-Meier 存活函數·····	29
圖八 五個共變數下加速失敗時間模型之百分位點圖·····	30
圖九 估計廣義伽瑪半參數 AFT 模型與共變數的關係·····	31
圖十 廣義伽瑪半參數 AFT 模型轉換後共變數的關係·····	32
圖十一 廣義伽瑪 AFT 模型下病人的風險函數·····	33
圖十二 共變數轉換後的廣義伽瑪 AFT 模型的 Cox-snell 殘差圖·····	33
圖十三 共變數轉換後的廣義伽瑪 AFT 模型之離差殘差圖·····	34

圖十四 肝硬化資料之對數累積風險與對數時間關係圖·····	35
圖十五 未引入共變數之 Cox 比例危險模型的 Martingale 殘 差圖·····	36
圖十六 Cox 比例風險模型的 Cox-snell 殘差圖·····	37

表目錄

表一 廣義伽瑪 AFT 模型 $\sigma=0.6$ $\lambda=0.3$ 之下，估計中位存活時間 的偏誤(均誤).....	38
表二 廣義伽瑪 AFT 模型 $\sigma=0.6$ $\lambda=1$ 之下，估計中位存活時間的 偏誤(均方誤).....	40
表三 廣義伽瑪 AFT 模型 $\sigma=0.6$ $\lambda=2$ 之下，估計中位存活時間 的偏誤(均方誤).....	42
表四 共變數轉換後兩種模型的 $-2\ln L$ 與 AIC	44
表五 共變數轉換後配適廣義伽瑪 AFT 模型之 $-2\ln L$ 值.....	44
表六 共變數轉換後之廣義伽瑪 AFT 模型，其共變數分別為 age、alb、log(bili)、log(protime)、edema 的係數 與 σ, λ 之估計值與標準差.....	45

第一章 緒論

1.1 研究動機與目的

右設限存活資料分析中的參數化加速失敗時間(AFT)模型，經常假設其存活分布為韋伯分布、對數常態分布或對數邏輯司分布。在上述分布之下，對數線性模式中的誤差分布則為極值分布、常態分布或邏輯司分布。這種特定分布的參數化 AFT 模型若與真正分布有所背離時，則其應用上將有偏差。因為實務上檢測分布的方法未必完善，資料也可能不足用於檢測，所以，本文考慮一個廣泛包含韋伯、對數常態等分布的廣義伽瑪分布(Cox et al. 2007)。此外，本文利用區域概似估計(local likelihood estimation)研究共變數引入對數線性模式的型態或轉換。因此，本文建議採用廣義伽瑪分布之下的 AFT 模型但是不需要假設對數時間與共變數之間是線性關係的廣義伽瑪半參數 AFT 模型分析右設限存活資料。

本文在第一章第二節回顧廣義伽瑪分布與區域概似估計法。第二章在上述廣義伽瑪加速失敗模型下探討引入共變數後之半參數模式。第三章進行模擬研究，針對不同存活分布、設限機率及樣本數之下，研究不同模型型態及分布下韋伯 AFT 模型、韋伯半參數 AFT 模型、廣義伽瑪 AFT 模型及廣義伽瑪半參數 AFT 模型的偏誤(bias)及均方誤(mean squares error)。第四章藉分析肝硬化病人與其危險因子的關

係，說明本文提統計方法之應用。最後，在第五章結論並且討論未來可能之研究。

1.2 文獻回顧

1.2.1 加速失敗時間模型

令具共變數 \tilde{x} 病人的風險函數為 $h(t; \tilde{x})$ ，則 $\tilde{x}=0$ (基準組) 病人的風險函數為 $h(t; 0)$ ，記作 $h_0(t)$ 。在加速失敗模型下

$$h(t; \tilde{x}) = e^{-\alpha' \tilde{x}} h_0(te^{-\alpha' \tilde{x}}) \quad (1)$$

其中 $\exp(-\alpha' \tilde{x})$ 稱為加速因子 (acceleration factor)。

事實上參數化 AFT 模型，可用對數線性模型 (log-linear form) 表示

$$\log T = \mu + \alpha' \tilde{x} + \sigma \varepsilon \quad (2)$$

其中 μ 為截距參數， α 為共變數 \tilde{x} 的係數， σ 為尺度參數， ε 則為誤差項。令具共變數 \tilde{x} 病人的存活函數為 $S_T(t, \tilde{x})$ ，基準組病人的存活函數為 $S_0(t)$ ，則在加速失敗模型下 $S(t, \tilde{x}) = S_0(te^{-\alpha' \tilde{x}})$ 。

若 $P \times 100\%$ 具共變數 \tilde{x} 病人的存活時間為 $t(p; \tilde{x})$ ，則

$$t(p; \tilde{x}) = e^{-\alpha' \tilde{x}} t(p; 0)$$

其中 $t(p; 0)$ 為基準組病人的 100p 百分位存活時間。所以在 AFT 模型下，具共變數 \tilde{x} 病人與基準組病人的百分位存活時間成比例，且其比

率與共變數 x 有關。

1.2.2 廣義伽瑪分布

本文所提的存活分布是由 Cox et al. (2007) 提出的廣義伽瑪分布，令存活時間 T 服從位置參數為 β ，尺度參數為 σ 且型態參數為 λ 的廣義伽瑪分布，記作 $T \sim GG(\beta, \sigma, \lambda)$ ，則其機率密度函數為

$$f(t) = \frac{|\lambda|}{\sigma \Gamma(\lambda^{-2})} [\lambda^{-2} (e^{-\beta t})^{\lambda/\sigma}] \exp[-\lambda^{-2} (e^{-\beta t})^{\lambda/\sigma}], \quad t > 0$$

其中 $\sigma > 0$ ， λ 及 β 為實數。對應的存活函數則為

$$S(t) = \begin{cases} 1 - \Gamma[\lambda^{-2} (e^{-\beta t})^{\lambda/\sigma}; \lambda^{-2}] & \text{if } \lambda > 0 \\ \Gamma[\lambda^{-2} (e^{-\beta t})^{\lambda/\sigma}; \lambda^{-2}] & \text{if } \lambda < 0 \end{cases}$$

其中 $\Gamma(t; \gamma) = \int_0^t x^{\gamma-1} e^{-x} dx / \Gamma(\gamma)$ 為不完整伽瑪函數，亦即是平均數，變異數皆為 γ 的伽瑪隨機變數的累積分布函數。

廣義伽瑪分布為一分布族，當 $\lambda=1$ 時其為韋伯分布(尺度參數為 $e^{-\beta/\sigma}$ 且型態參數為 σ^{-1})；當 $\lambda \rightarrow 0$ 時，其對應分布為對數常態分布(位置參數為 β 且尺度參數為 σ)；當 $\lambda = \sigma$ 時，則簡化為伽瑪分布(型態參數為 λ^{-2} 且尺度參數為 $\lambda^2 e^\beta$)。此外當 $\lambda = -1$ 時，為反韋伯(Inverse Weibull)分布，而當 $\lambda = -\sigma$ 時為反伽瑪(Inverse gamma)分布。值得一提的是此一分布族具多種型態的風險函數，包含單調遞減(monotonically decreasing hazard)、單調遞增(monotonically increasing hazard)、浴缸型態(bathtub shape hazard)與弧形型態(arc shape hazard)之風險函數，對應四種型態之參數可參見圖一。

如果 $T \sim GG(\beta, \sigma, \lambda)$ 且 $b \neq 0$ ，則

$$e^a T^b \sim GG(a + \beta b, \sigma |b|, \text{sign}(b)\lambda) \quad (3)$$

其中 $\text{sign}(b) = 1$ 若 $b > 0$ ， $= -1$ 若 $b < 0$ 。令 $T_0 \sim GG(0, 1, \lambda)$ ，亦即標準廣義伽瑪分布，則其機率密度函數為

$$f_{GG(0,1,\lambda)}(t) = \frac{|\lambda|}{t\Gamma(\lambda^{-2})} [\lambda^{-2} t^\lambda] \exp[-\lambda^{-2} t^\lambda]$$

存活函數為

$$S_{GG(0,1,\lambda)}(t) = \begin{cases} 1 - \Gamma[\lambda^{-2} t^\lambda; \lambda^{-2}] & \text{if } \lambda > 0 \\ \Gamma[\lambda^{-2} t^\lambda; \lambda^{-2}] & \text{if } \lambda < 0 \end{cases}$$

得知

$$T = e^\beta T_0^\sigma \sim GG(\beta, \sigma, \lambda) \quad (4)$$

所以，

$$\log T = \beta + \sigma \log T_0 \quad (5)$$

此外，

$$T^{\lambda/\sigma} \sim GG\left(\frac{\beta\lambda}{\sigma}, |\lambda|, |\lambda|\right) \quad (6)$$

亦即 $T^{\lambda/\sigma}$ 服從尺度參數為 $\lambda^2 e^{\beta\lambda/\sigma}$ 及型態參數為 λ^2 的伽瑪分布。

令 $t_{GG(\beta,\sigma,\lambda)}(p)$ 為 T 的 100p 百分位存活時間，則由(5)得知

$$\log[t_{GG(\beta,\sigma,\lambda)}(p)] = \beta + \sigma \log[t_{GG(0,1,\lambda)}(p)]$$

其中

$$t_{GG(0,1,\lambda)}(p) = \begin{cases} [\lambda^2 \Gamma^{-1}(p; \lambda^{-2})]^{1/\lambda} & \text{if } \lambda > 0 \\ [\lambda^2 \Gamma^{-1}(1-p; \lambda^{-2})]^{1/\lambda} & \text{if } \lambda < 0 \end{cases}$$

所以，在廣義伽瑪分布下可獲得 AFT 模型。

1.2.3 廣義伽瑪加速失敗時間模型

令 T_i 是共變數為 \underline{x}_i 病人之存活時間， $i=1,2,\dots,n$ 。在廣義伽瑪加速失敗時間模型下，考慮對數線性模型

$$\log T_i = \mu + \eta_i + \sigma \varepsilon_{i\lambda} \quad (7)$$

其中 μ 為截距參數， $\eta_i = \alpha \underline{x}_i$ ， $\varepsilon_{i\lambda}$ 之分布為 $\log GG(0,1,\lambda)$ 。此一模式中的參數 $\theta = [\mu, \alpha, \sigma, \lambda]'$ 的估計可藉由最大概似估計法(maximum likelihood estimation)求得。令 δ_i 為對應於 T_i 的設限指標，則觀察右設限資料後，參數 θ 的右設限的概似函數為

$$L(\theta; \underline{x}, \delta) = \prod_{i=1}^N [f_i(t_i)]^{\delta_i} [S_i(t_i)]^{1-\delta_i}$$

其中 $f_i(t_i)$ 及 $S_i(t_i)$ 為 T_i 的機率密度函數及存活函數，分別為

$$f_i(t_i) = \frac{|\lambda|}{\sigma t \Gamma(\lambda^{-2})} [\lambda^{-2} (e^{-(\mu + \alpha' \underline{x}_i)} t)^{\lambda/\sigma}] \exp[-\lambda^{-2} (e^{-(\mu + \alpha' \underline{x}_i)} t)^{\lambda/\sigma}], \quad t > 0$$

$$S_i(t_i) = \begin{cases} 1 - \Gamma[\lambda^{-2} (e^{-(\mu + \alpha' \underline{x}_i)} t)^{\lambda/\sigma}; \lambda^{-2}] & \text{if } \lambda > 0 \\ \Gamma[\lambda^{-2} (e^{-(\mu + \alpha' \underline{x}_i)} t)^{\lambda/\sigma}; \lambda^{-2}] & \text{if } \lambda < 0 \end{cases}$$

則對數概似函數為

$$l(\theta) = \ln[L(\theta; \underline{x}, \delta)] = \sum_{i=1}^N \delta_i \ln[f_i(t_i)] + (1 - \delta_i) \ln[S_i(t_i)]$$

$l(\theta)$ 為 θ 的非線性函數，因此我們使用數值方法求 θ 的最大概似估計量，計作 $\hat{\theta}$ 。

由對數概似函數得到觀察的訊息矩陣(observed information matrix)

為

$$i(\theta) = \begin{bmatrix} -\frac{\partial^2 l}{\partial \mu^2} & -\frac{\partial^2 l}{\partial \mu \partial \alpha_1} & \cdots & -\frac{\partial^2 l}{\partial \mu \partial \alpha_k} & -\frac{\partial^2 l}{\partial \mu \partial \sigma} & -\frac{\partial^2 l}{\partial \mu \partial \lambda} \\ -\frac{\partial^2 l}{\partial \mu \partial \alpha_1} & -\frac{\partial^2 l}{\partial \alpha_1^2} & \cdots & -\frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_k} & -\frac{\partial^2 l}{\partial \alpha_1 \partial \sigma} & -\frac{\partial^2 l}{\partial \alpha_1 \partial \lambda} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -\frac{\partial^2 l}{\partial \mu \partial \alpha_k} & -\frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_k} & \cdots & -\frac{\partial^2 l}{\partial \alpha_k^2} & -\frac{\partial^2 l}{\partial \alpha_k \partial \sigma} & -\frac{\partial^2 l}{\partial \alpha_k \partial \lambda} \\ -\frac{\partial^2 l}{\partial \mu \partial \sigma} & -\frac{\partial^2 l}{\partial \alpha_1 \partial \sigma} & \cdots & -\frac{\partial^2 l}{\partial \alpha_k \partial \sigma} & -\frac{\partial^2 l}{\partial \sigma^2} & -\frac{\partial^2 l}{\partial \sigma \partial \lambda} \\ -\frac{\partial^2 l}{\partial \mu \partial \lambda} & -\frac{\partial^2 l}{\partial \alpha_1 \partial \lambda} & \cdots & -\frac{\partial^2 l}{\partial \alpha_k \partial \lambda} & -\frac{\partial^2 l}{\partial \sigma \partial \lambda} & -\frac{\partial^2 l}{\partial \lambda^2} \end{bmatrix}_{\theta=\hat{\theta}}$$

藉此我們得知最大概似估計式 $\hat{\theta}$ 的變異數共變異數矩陣

(variance-covariance matrix)為

$$\text{Var}(\hat{\theta}) = i^{-1}(\hat{\theta}),$$

其中 $i^{-1}(\hat{\theta})$ 為 $i(\hat{\theta})$ 的反矩陣(inverse matrix)。在此一模型下，可以估

計其 100p 百分位數為

$$\hat{T}(p; \underline{x}) = \exp\{\hat{\mu} + \hat{\eta} + \hat{\sigma} \ln(T_{GG(0,1,\hat{\lambda})}(p))\}$$

為求得此一百分位數估計之變異數，令

$$g(\theta) = \log(T(p; \underline{x})) = \mu + \eta + \sigma \log(T_{GG(0,1,\lambda)}(p)) \circ$$

求得

$$d(\underline{\theta}) = \frac{\partial g(\underline{\theta})}{\partial \theta_j} = \begin{bmatrix} 1 \\ x_{1,j} \\ \vdots \\ x_{k,j} \\ \log(T_{GG(0,1,\hat{\lambda})}(p)) \\ \sigma \frac{\partial \log(T_{GG(0,1,\hat{\lambda})}(p))}{\partial \lambda} \end{bmatrix}$$

應用 Delta 方法得知 $\log(\hat{T}(p; \underline{x}))$ 的變異數估計為

$$\hat{Var}(g(\hat{\underline{\theta}})) = d(\hat{\underline{\theta}})^t \text{var}(\hat{\underline{\theta}}) d(\hat{\underline{\theta}}),$$

再度使用 Delta 方法，即可得到 $\hat{T}(p; \underline{x})$ 的變異數估計為

$$\hat{Var}(\hat{T}(p; \underline{x})) = [\hat{T}(p; \underline{x})]^2 \hat{Var}(\log(\hat{T}(p; \underline{x})))$$

1.2.4 區域概似估計

事實上，若引入共變數 $x_i, i=1, \dots, n$ 說明對數時間變化，其模型可為

$$\log T_i = \sum_{j=1}^k s_j(x_{ji}) + \sigma \epsilon$$

其中 $s_j(x_j), j=1, \dots, k$ 為任意平滑連續函數， k 為共變數個數。

$s_j(x_j), j=1, \dots, k$ 可應用由 Tibshirani & Hastie(1987) 提出區域概似加

以估計。在介紹區域概似方法之前首先介紹散佈點平滑化

(Scatterplot Smoothing) (參考 Cleveland(1979) 及 Friedman and

Stuetzle(1981))。在傳統迴歸模型中令 Y 為反應變數， x 為共變數，

再令 $s(x) = E[Y | X = x]$ 。假設 $s(x)$ 為非線性平滑函數，因此使用區域配

適(local fitting)概念，即利用 x_i 週遭的資料估計 $s(x_i)$ 為

$$\check{s}(x_i) = \text{Fit}(\{(x_j, y_j) : j \in N_i\}, x_i)$$

其中 $N_i = \{\max(i - \frac{[wn]-1}{2}, 1), \dots, i-1, i, i+1, \dots, \min(i + \frac{[wn]-1}{2}, n)\}$ 稱為 x_i 的對稱鄰近集 (symmetric nearest neighborhoods) 且 w 為其頻寬。利用區域配適線性模型可以得到

$$\tilde{s}(x_i) = \tilde{\beta}_{0i} + \tilde{\beta}_{1i}x_i$$

其中 $\tilde{\beta}_{0i}$ 與 $\tilde{\beta}_{1i}$ 為根據資料點為 $\{(x_j, y_j) : j \in N_i\}$ 以最小平方方法配適線性迴歸得到的係數。

推廣散佈點平滑化的概念到概似函數上。考慮資料 (t_i, x_i, δ_i) ， $i=1, \dots, n$ ，且互相獨立，其中 t 為存活時間， x 為共變數， δ 是設限指標。已知共變數為 x 時， T 的機率密度函數為

$$T | x \sim f(y, \theta)$$

其中 $\theta = s(x)$ 為 x 的平滑函數，可得到概似函數為

$$L(s(x_1), s(x_2), \dots, s(x_n)) = \prod_{i=1}^n [f(t_i, \theta_i)]^{\delta_i} [S(t_i, \theta_i)]^{1-\delta_i}$$

其中 $\theta_i = s(x_i)$ ， $i=1, \dots, n$ 。在線性迴歸模型之下，則區域概似函數為

$$L_i = \prod_{j \in N_i} [f(t_j, s(x_j))]^{\delta_j} [S(t_j, s(x_j))]^{1-\delta_j}, \quad i=1, \dots, n$$

其中 $s(x_j) = \beta_{0i} + \beta_{1i}x_j$ ， $s(x_i)$ 的區域概似估計函數 (Local likelihood estimate) 為

$$\hat{s}(x_i) = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_i$$

其中 $\hat{\beta}_{0i}$ 和 $\hat{\beta}_{1i}$ 為第 i 個人根據資料 $\{(x_j, t_j) : j \in N_i\}$ 所求 β_{0i} 與 β_{1i} 的最大區

域概似估計值。

考慮給定頻寬的散佈點平滑化，假設在此頻寬內為線性模型，則可以在區域內以最小平方法估計 y 得到 $\hat{y} = Py$ ，其中 P 是平滑矩陣 (smoother matrix)。進一步可求得殘差平方和的期望值為

$$\begin{aligned} E[RRS(y, \hat{y})] &= E[(y - \hat{y})'(y - \hat{y})] = E[((I - P)y)'((I - P)y)] \\ &= \sigma^2 tr(n - (2tr(P) - tr(P'P))) + E[y'](I - P)'(I - P)E[y] \end{aligned}$$

且定義 \hat{y} 的自由度 (degrees of freedom) $df(\hat{y}) = 2tr(P) - tr(P'P)$ ，且因為 P 的第 i 列為第 i 個病人對應之投影矩陣 (projection matrix) 的第 i 列，所以可得到 $tr(P'P) = tr(PP') = tr(P)$ ，因此 $df(\hat{y}) = 2tr(P) - tr(P) = tr(P)$ ，進一步使用 $tr(P)$ 來近似區域概似估計的自由度。當頻寬最大時，意即每位病人的對稱鄰近集皆包含全部資料時，此時 $tr(P) = 2$ ；若頻寬達到最小時，每位病人的對稱鄰近集只包含自己時，整體來說有 n 位病人，此時 $tr(P) = n$ 。

關於在區域概似估計下頻寬的選擇，考慮 AIC 準則如下：

$$AIC = -2\log L + 2df,$$

其中 df 為自由度。建議選擇使得 AIC 達到最小的頻寬。當平滑函數越平滑時，所選的頻寬越大反之平滑函數越不平滑時，選出的頻寬越小。

假如共變數為兩個或兩個以上，其參數假設為 $\theta = \sum_{j=1}^k s_j(x_j)$ ，其中 k 為共變數個數。可以採用向後配適法 (Backfitting)。假設 $k = 2$ 時，

第一步先選一個平滑函數 $\hat{s}_1(x_1)$ ，第二步選出平滑函數 $\hat{s}_2(x_2)$ ，在模型給定平滑函數 $\hat{s}_2(x_2)$ 下， $\hat{s}_1(x_1)$ 可能不會是最理想的，因此需要在有考慮 $\hat{s}_2(x_2)$ 的模型下，重新估計 $\hat{s}_1(x_1)$ ，得到修正過的 $\hat{s}_1^*(x_1)$ ，再使用 $\hat{s}_1^*(x_1)$ 得到修正過的 $\hat{s}_2^*(x_2)$ ，重複上面步驟直到 $\hat{s}_1^*(x_1)$ 與 $\hat{s}_2^*(x_2)$ 收斂到平穩的平滑函數。

將之推廣到模型已經有平滑函數 $\hat{s}_1(x_1), \dots, \hat{s}_r(x_r)$ 的情形，若加入新的平滑函數 $\hat{s}_{r+1}(x_{r+1})$ ，則向後配適過程是在 $\hat{s}_k(x_k)$ ， $k \neq j$ 為已知的模型

$$\theta = \sum_{k \neq j} \hat{s}_k(x_k) + s_j(x_j)$$

估計 $s_j(x_j)$ ， $j=1, \dots, r+1$ 。

Tibshirani et al. (1987) 運用區域概似估計法到 Cox 比例危險模型 (Cox proportional hazards model) 上，其 Cox 比例風險假設為

$$h(t | \underline{x}) = h_0(t) \exp(\beta' \underline{x})$$

使用 $\sum_{j=1}^k s_j(x_j)$ 替代 $\beta' \underline{x}$ ，利用 Cox 比例危險模型的部份概似函數

(partial likelihood function) 求得平滑函數 $\hat{s}_1(x_1), \dots, \hat{s}_k(x_k)$ 。因此本文運用區域概似估計法引入廣義伽瑪分布下的加速失敗時間模型，不需要假設對數時間與共變數間的關係。

第二章 統計分法

事實上，對應廣義伽瑪分布的加速失敗時間模型的對數時間模型中，對數時間與共變數的關係不必然為線性，當對數時間模型不定時，則所得到模型將為廣義伽瑪半參數加速失敗時間模型：

$$\log T_i = \sum_{j=1}^k s_j(x_{ji}) + \sigma \varepsilon_\lambda,$$

其中 k 為共變數個數， $s_j(x_{ji})$, $j=1, \dots, k$, $i=1, \dots, n$ ，是非線性平滑連續函數， ε_λ 之分布為 $\log GG(0,1,\lambda)$ 。

針對此種廣義伽瑪半參數模型，本文引用區域概似方法估計 $s_j(x_{ji})$, $j=1, \dots, k$ ，但是 Tibshirani et al. (1987) 運用在 Cox 比例風險模型，只需針對部分概似函數裡的 $\sum_{j=1}^k s_j(x_j)$ ，而廣義伽瑪半參數 AFT 模型的概似函數裡有干擾參數 σ 與 λ ，因此需要使用類似向後配適法之過程。

其演算法如下：

步驟一：在對數時間與共變數為線性關係下，估計出 $\hat{\sigma}, \hat{\lambda}$

步驟二：在 $\sigma = \hat{\sigma}$ 及 $\lambda = \hat{\lambda}$ 下，使用區域概似估計，對共變數個別配適

選出其適當的頻寬

步驟三：估計出個別共變數的 $\hat{s}_1(x_1), \dots, \hat{s}_k(x_k)$

步驟四：使用向後配適(backfitting)估計出修正的平滑函數

$$\hat{s}_1^*(x_1), \dots, \hat{s}_k^*(x_k)$$

步驟五：令 $\hat{s}_1^*(x_1), \dots, \hat{s}_k^*(x_k)$ 為已知，重新估計修正的 $\hat{\sigma}^*, \hat{\lambda}^*$

重複步驟四及步驟五直至收斂到穩定的平滑函數 $\hat{s}_1^*(x_1), \dots, \hat{s}_k^*(x_k)$ 與

$\hat{\sigma}^*, \hat{\lambda}^*$ 。此時收斂的平滑函數為

$$\hat{s}_j(x) = \frac{(x_{j,i} - x)(\hat{\beta}_{0,j,i-1} + \hat{\beta}_{1,j,i-1}x) + (x - x_{j,i-1})(\hat{\beta}_{0,j,i} + \hat{\beta}_{1,j,i}x)}{x_{j,i} - x_{j,i-1}},$$

$x_{j,i-1} < x < x_{j,i}$ ， $j=1, \dots, k$ 。

在此如果考慮在全部共變數下選取其適當的頻寬，則需要的電腦計算量太大，因此本文在一開始就對共變數個別配適後選出其適當的頻寬，以及使用類似向後配適法之過程重新估計 $\hat{\sigma}, \hat{\lambda}$ ，得到修正後的 $\hat{\sigma}^*, \hat{\lambda}^*$ 。得到區域概似估計後，進一步可由對數時間模型得到共變數為 x 的 100p 百分位存活時間為

$$\hat{T}(p; x) = \exp\left\{\sum_{j=1}^k \hat{s}_j(x_j) + \hat{\sigma} \ln(T_{GG(0,1,\hat{\lambda})}(p))\right\},$$

此外，我們可以藉由自助法(bootstrap)隨機重複抽取樣本的資料，每次抽取的樣本數為真實資料的樣本數，重複抽取 200 次後估計 $\hat{T}(p; x)$ 的變異數。

事實上，在配適 $s_j(x_j)$, $j=1, \dots, k$ 之後，或可發現這些 $s_j(x_j)$ 為 x_1, \dots, x_k 的某一種函數對應。若可以合理的將 x_1, \dots, x_k 轉換成 $h_1(x_1), \dots, h_k(x_k)$ 後，使得對數時間與 $h_1(x_1), \dots, h_k(x_k)$ 皆接近線性關係，則可應用第一章第二節裡的廣義伽瑪加速失敗時間模型分析資料，若無法合理找到適當的

共變數轉換得到對數線性模型，則可直接採用廣義伽瑪半參數加速失敗時間模型加以描述。

第三章 模擬研究

3.1 模擬方法

為瞭解應用各種可能的 AFT 模型估計中位存活時間的表現，本文執行一個模擬研究。此一模擬中，考慮在不同存活分布，設限機率及樣本數之下，研究不同模型型態及分布時，傳統的韋伯 AFT 模型，韋伯半參數 AFT 模型與廣義伽瑪 AFT 模型及本文所提及的廣義伽瑪半參數 AFT 模型，估計中位數存活時間的偏誤(bias)及均方誤(mean squares error)。

在此一模擬研究中，考慮存活時間服從廣義伽瑪分布，且考慮 $\sigma = 0.6, \lambda = 0.3, 1$ 及 2 的三種情況，其對應之風險函數為弧形型態、單調遞增與浴缸型態。AFT 模型中僅具一個連續共變數，但是對數時間與共變數間係可能為線性或非線性。因此，對數模型為

$$\log T = s(x) + \sigma \cdot \ln(T_{GG(0,1,\lambda)})$$

其中共變數由均勻分布 $U(0, 2)$ 中生成。此處考慮三種模型分別為

$$\text{模型 I: } s(x) = 3 - 0.7x$$

$$\text{模型 II: } s(x) = 1.5 + 3.5x - 2x^2$$

$$\text{模型 III: } s(x) = 2 - \sin(2x)$$

及模型圖形參見圖二。

此外考慮均勻分布 $U(0, R)$ 的設限時間變數，其中 R 的決定是在給

定的對數時間模型誤差分布下使得設限機率分別為 0.3 和 0.5。

藉由標準廣義伽瑪分布與伽瑪分布的關係，生成標準廣義伽瑪隨機變數 T_0 ，即 $T_0 = T_G^{1/\lambda}$ ，其中 T_0 為服從 $GG(0,1,\lambda)$ 之隨機變數， T_G 為服從尺度參數為 λ^2 及型態參數為 λ^{-2} 的伽瑪分布之隨機變數。然後依模型生成廣義伽瑪分布的隨機變數

$$T_i = \exp\{s(x_i) + \sigma \log(T_0)\}$$

文中隨機變數皆利用 R 程式中的 `runif`、`survival package` 與 `rgammag` 生成。每一種情形下的模擬重複 1000 次，每一次皆生成樣本數 $n=100$ 或 300 的資料。在模擬中研究當共變數 $x=0.5, 1$ 及 1.5 時，四種模型配適資料，估計中位存活時間的偏誤及均方誤。此時的估計偏誤為中位存活時間估計值的平均數與真正中位存活時間的差異，至於均方誤則為中位存活時間估計之變異數與上述估計偏誤平方的總和。結果列於表一到表三。

3.2 模擬結果

首先由表二得知在韋伯分布的模型 I 中，以線性模型估計中位存活時間的偏誤及均方誤皆比以區域概似方法配模的結果佳。但是在模型 II 中當共變數 $x=1$ 時，韋伯或廣義伽瑪分布下的線性模型皆嚴重的低估中位存活時間，因此，對應的均方誤也明顯的比區域概似估計的

高。而在模型Ⅲ中當共變數 $x=1$ 時，韋伯或廣義伽瑪分布下的線性模型則嚴重的高估中位存活時間，對應的均方誤也明顯比概似估計高。這是因為在 $x=1$ 時，模型Ⅱ及Ⅲ皆明顯的異於線性模型。整體來說在模型Ⅱ或模型Ⅲ之下，區域概似估計在共變數 $x=1$ 時結果比以 $x=0.5, 1.5$ 時佳，因此當 x 值越靠近中心附近時，其區域概似估計結果表現越好，反之越差。但是一般而言，當模型不是線性時，區域概似估計的配模結果比配適線性模式較佳。

由表一與表三中得知，在廣義伽瑪分布且非韋伯分布下的韋伯線性 AFT 與韋伯半參數 AFT，其偏誤及均方誤整體上比以廣義伽瑪 AFT 模型及半參數 AFT 模型的結果差，因為韋伯分布未包含真實分布。在模型Ⅰ中，廣義伽瑪 AFT 模型估計中位存活時間的偏誤及均方誤皆比以其他三種模型的結果佳。但在模型Ⅱ中當共變數 $x=1$ 時，廣義伽瑪 AFT 模型也嚴重低估中位存活時間導致均方誤明顯比廣義伽瑪半參數 AFT 模型的高。在模型Ⅲ中當共變數 $x=1$ 時，則是明顯高估。因此當對數時間與共變數間的關係未知時，採用廣義伽瑪半參數 AFT 模型比其他三種模型來的穩健，也說明廣義伽瑪半參數 AFT 模型是很廣泛的分布族且不需要限制共變數與對數時間的關係。

就其設限機率而言，區域概似方法在設限機率越大時，其偏誤及均方誤上升幅度較線性模型的大。就其樣本數而言，樣本數越大則區

區域概似方法的偏誤及均方誤比以線性模型估計的改善較大。因此當樣本數過小或設限機率過大時，區域概似估計方法的偏誤及均方誤皆可能過大。

第四章 實例分析

在此分析一組研究原發性肝硬化(primary biliary cirrhosis ,PBC)與其危險因子關係的臨床資料(Fleming and Harrington, 1991)。此一肝硬化臨床資料共有312位罹患原發性肝硬化之病人且病人從登記患有原發性肝硬化到死亡的時間，即為存活時間；若病人中途離開或結束長期追蹤時仍然存活，則此一時間列為右設限時間。此臨床資料主要研究原發性長期膽汁滯留造成慢性肝炎，最後導致肝硬化，即肝臟組織纖維化和結節化的症狀，與共變數為年齡(age)、血清白蛋白(alb)、膽紅素(bili)、凝血酶原時間(protime)與水腫(edema)的相關性。其中年齡是從出生至登記患有原發性硬化的時間。而血清白蛋白由肝臟製造，肝硬化病人的血清白蛋白會降至正常值以下或更低。膽紅素則是血紅素破壞後的代謝產生，正常血中結合膽紅素含量很低，濾過量極少，因此尿中檢驗不出膽紅素，肝硬化病人其膽紅素偏高。凝血酶原時間在肝硬化病患上其值會延長，而水腫(edema)為順序型變數(ordinal data)，當沒有水腫也沒有利尿紀錄記作0，沒有水腫但有利尿紀錄與有水腫但沒有利尿紀錄記作0.5，有水腫也有利尿紀錄記作1，而此資料中的age、alb、bili與protime皆為連續型變數(continuous data)。

年齡、血清白蛋白、膽紅素與凝血酶原時間分別使用其中位數將

共變數個別區分成兩群後，其中低於中位數之共變數為第一群，高於中位數之共變數為第二群，使用Kaplan-Meier(1958)方法估計個別共變數之第一群與第二群的存活函數與百分位存活時間，而水腫共變數則是沒有水腫也沒有利尿記錄與其他兩群分別使用Kaplan-Meier方法估計，由圖七發現存活時間會隨age、bili、protime增加而縮短，隨alb增加而增長，圖八則顯示圖形約為通過原點的直線，因此並無顯著證據說明此資料不適合配適加速失敗時間模型。

因此引入五個共變數分別配適廣義伽瑪AFT模型與廣義伽瑪半參數AFT模型，其中廣義伽瑪半參數AFT模型中的四個連續型變數使用區域概似方法配適，而edema則使用線性模型配適，得到兩種模型的AIC分別為745.431及729.485。所以在廣義伽瑪半參數AFT模型改善較多，且可經由廣義伽瑪半參數AFT模型探討共變數引入對數線性模式的可能轉換，其中廣義伽瑪的位置參數與各連續型共變數的關係圖參見圖九。由圖九可知age、alb與對數時間為線性相關，而bili、protime與對數時間為非線性相關，且隨著其值越大參數下降趨勢減緩，考慮bili與protime取對數變數轉換後，再使用兩種模型配適，參見表四。

由表四得知在廣義伽瑪AFT模型下線性模型的AIC較小，且由圖十可知 $\log(\text{bili})$ 、 $\log(\text{protime})$ 與對數時間的關係接近線性，因此我們選出的模型為共變數變換後的廣義伽瑪AFT模型。

接著在共變數變換後的廣義伽瑪AFT模型下使用逐步選模法 (stepwise procedure) 選擇出顯著的共變數(表五)，得知在廣義伽瑪加速失敗時間模型下，五個共變數都顯著影響研究對象的存活時間。

且由表六可知，age、log(bili)、log(protime)與edema增加則存活時間皆縮短，alb下降則存活時間縮短，也可由對數模型得到研究對象的100p百分位存活時間估計 $\hat{T}(p; x)$ 與其變異數。且由 $\hat{\sigma}, \hat{\lambda}$ 得知此模型的風險函數為弧形型態，在此針對幾位病人繪出其風險函數(圖十一)，可知第18位病人死亡風險最高。

此外我們針對廣義伽瑪AFT模型下的Cox-snell殘差圖(圖十二)我們發現圖形趨近45度之直線，且離差殘差圖(圖十三)均勻分布沒有明顯的趨勢，也驗證研究對象間互相獨立的假設，因此變數轉換後的線性AFT模型配適此筆資料適當。

最後使用Cox比例風險模型配適此筆資料，由圖十四可知兩條曲線約平行，因此沒有明顯證據說明違反比例風險假設，且可由Martingale殘差圖(圖十五)也說明共變數bili取對數變數轉換後更接近直線關係，再由Cox-snell殘差圖(圖十六)發現圖形接近45度之直線。因此，上述資料也可以針對轉換後共變數應用Cox線性模式加以分析。本文針對廣義伽瑪半參數AFT模型研究此筆資料，主要目的除了瞭解共變數與存活時間關係外，進一步能繪出病人的風險函數。

第五章 結論及未來研究

事實上，本文模型考慮在加速失敗時間模型下，假設存活時間為一個較廣泛的分布族且對數時間與共變數關係未知時，利用區域概似方法研究對數時間與共變數之間的關係，藉由平滑函數 $s_j(x_j)$, $j=1, \dots, k$ 或可發現這些 $s_j(x_j)$ 為 x_1, \dots, x_k 的某一種函數對應，若可以合理的將 x_1, \dots, x_k 轉換成 $h_1(x_1), \dots, h_k(x_k)$ 後，得到對數線性模型，則可運用較簡單之模型，廣義伽瑪加速失敗時間模型分析資料。

本文假設存活分布為廣義伽瑪分布，但若此分布還不足以包含真實分配時，在應用上將會有誤差，因此另可參考一個廣泛的分布族——廣義韋伯分布(Bagdonavicius et al. 2001)，此一分布包含韋伯分布，且有多元化的風險函數，其存活函數為

$$S(t, \theta, \nu, \gamma) = \exp\{1 - (1 + (\frac{t}{\theta})^\nu)^{1/\gamma}\}, \quad (\theta, \nu, \gamma > 0); \quad t \geq 0$$

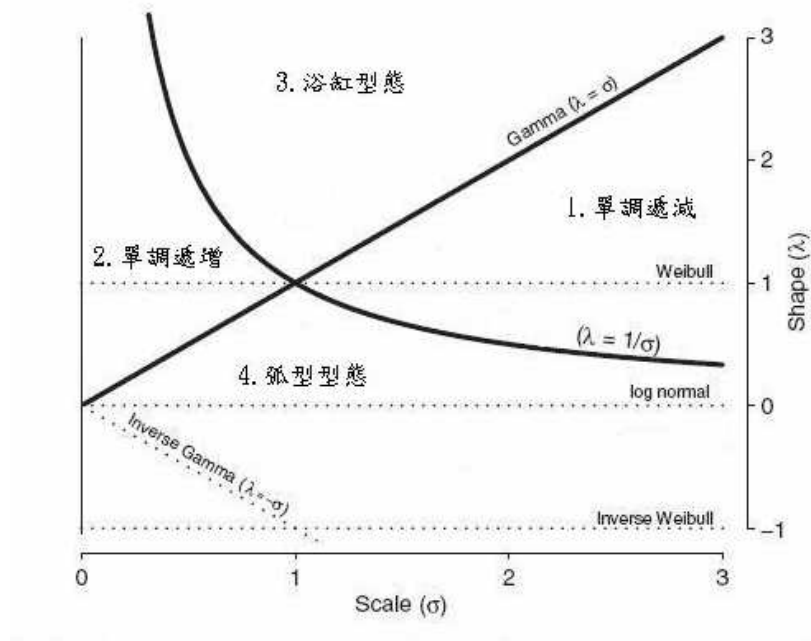
所以在廣義韋伯分布下也可獲得 AFT 模型。因此假若廣義韋伯分布包含真實分布且對數時間與共變數關係未知時，建議參考廣義韋伯分布。

此外，本文所提及之模型當資料包含左設限(left censored)及區間設限(interval censored)時，亦可經由概似函數求得其參數估計值。

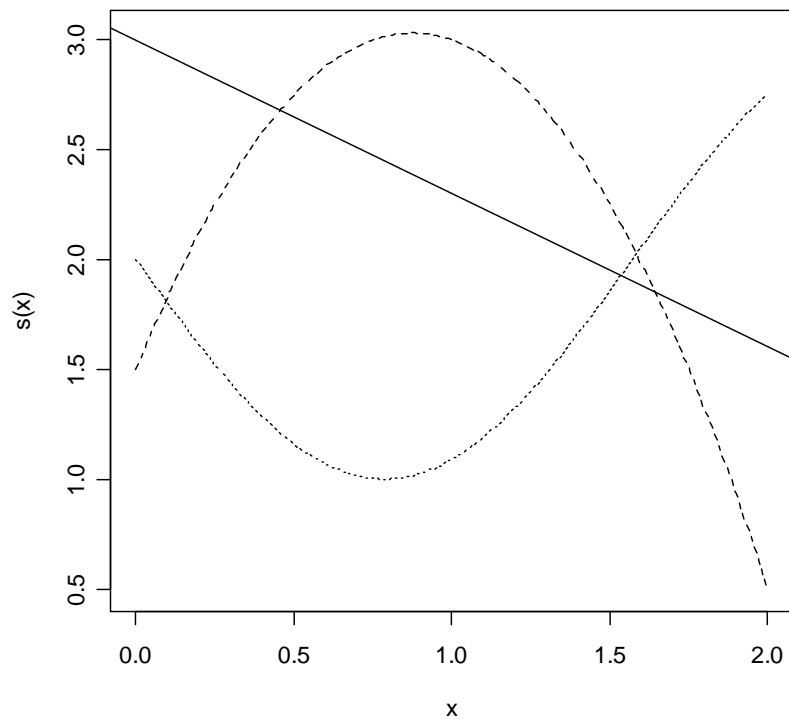
參考文獻

- [1] Bagdonavicius, V. and Nikulin, M.(2001). *Accelerated Life models Modeling and Statistical Analysis*. Springer: New York.
- [2] Cleveland, W. S.(1979). Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association*, 74, 828-836.
- [3] Cox ,C.,Chu, H., Schneider, M. F. and Munoz, A.(2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution, *Statistics in Medicine*, 26, 4352-4374.
- [4] Fleming, T. R. and Harrington, D. P.(1991). *Counting Processes and Survival Analysis*. Springer: New York.
- [5] Friedman, J. H. and Stuetzle, W.(1981). Projection Pursuit Regression, *Journal of the American Statistical Association*, 76, 817-823.
- [6] Friedman, J. H. and Stuetzle, W.(1981). Smoothing of Scatterplots, technical report(Orion 003), Stanford University, Dept. of Statistics.
- [7] Mudholkar, G. S., Srivastava, D. K. and Kollia, G. D.(1996). A generalization of the Weibull distribution with application to the analysis of survival data, *Journal of the American Statistical Association*, 91, 1575-1583.
- [8] Tibshirani, R. and Hastie, T.(1987). Local Likelihood Estimation, *Journal of the American Statistical Association*, 82, 398., 559-567
- [9] Sparling, Y. H., Younes, N. and Lachin, J. M.(2006). Parametric survival models for interval-censored data with time-dependent covariates, *Biostatistics*, 7, 4, 599-614.
- [10] Stacy, E. W.(1962). A generalization of the gamma distribution, *Annals of Mathematical Statistics*, 33, 1187-1192.

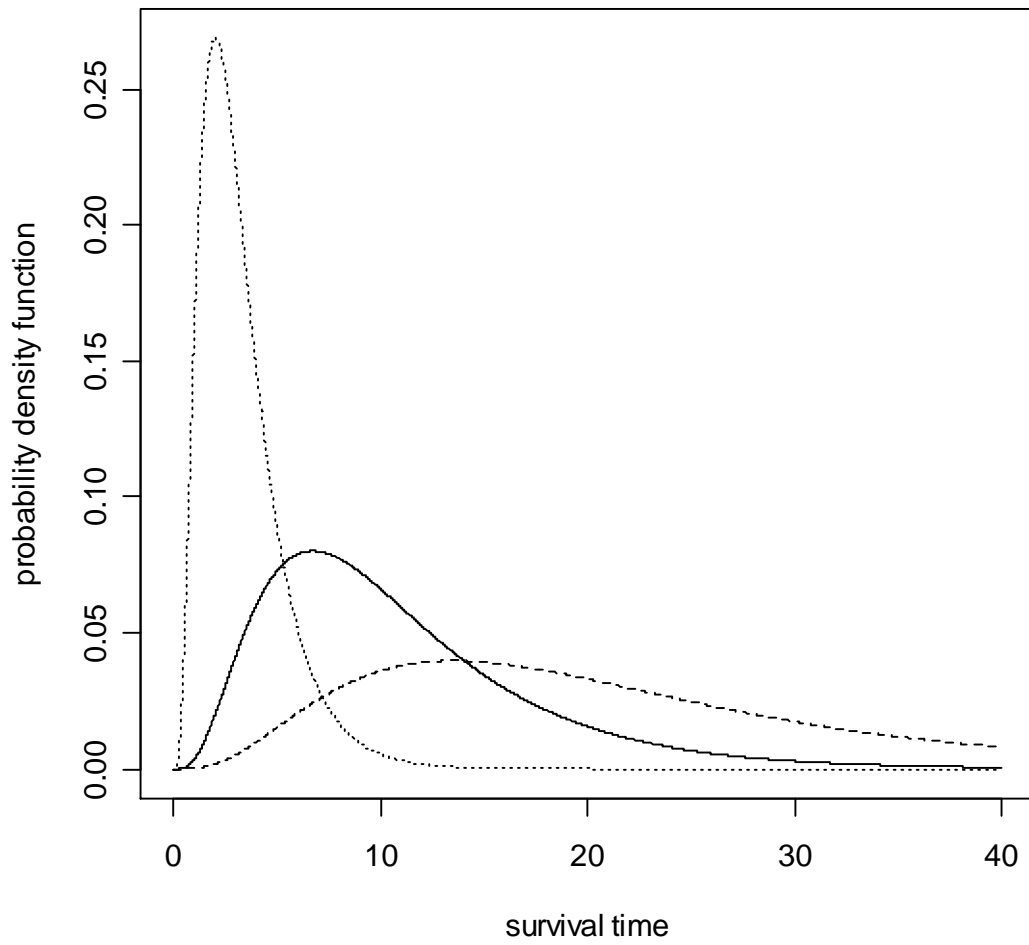
- [11] Stacy, E. W. and Mihram G. A.(1965). Parameter estimation for a generalized gamma distribution, *Technometrics*, 7, 349-358.
- [12] Wei, L. J.(1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis, *Statistics in Medicine*, 11, 1871-1879.



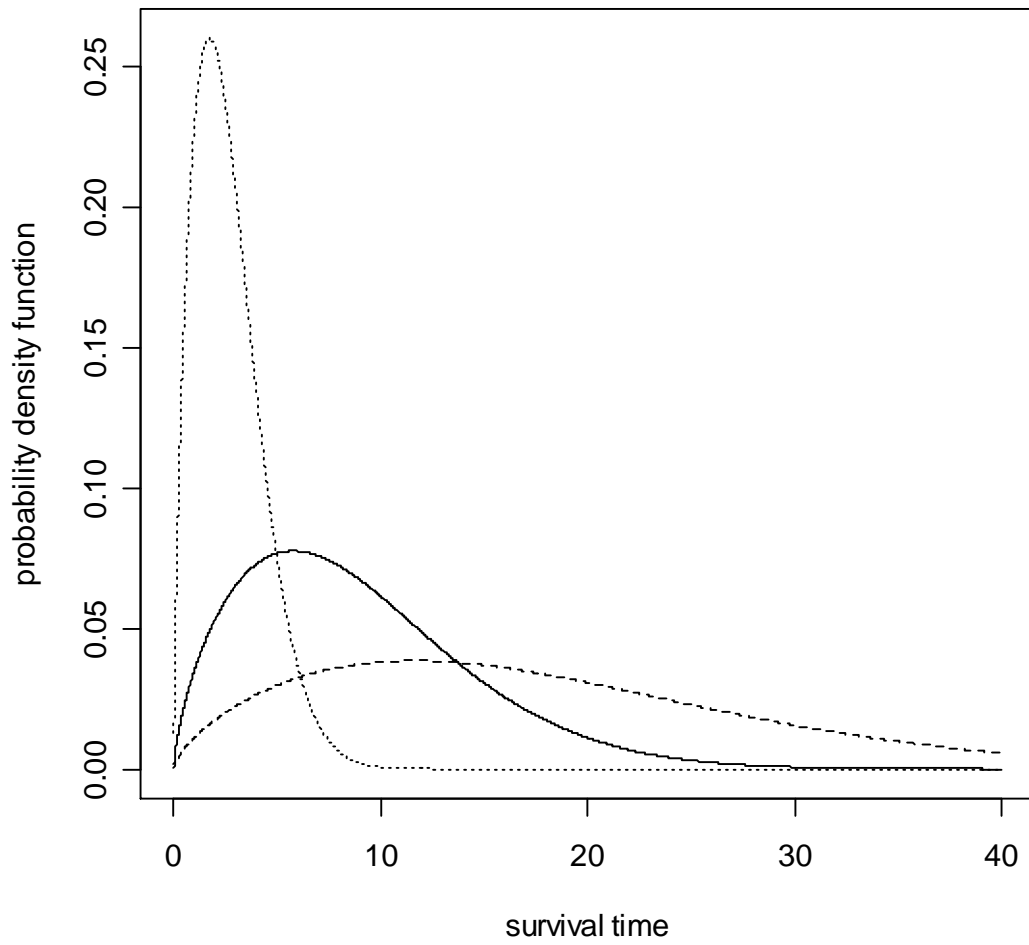
圖一 四種風險函數型態對應之參數圖



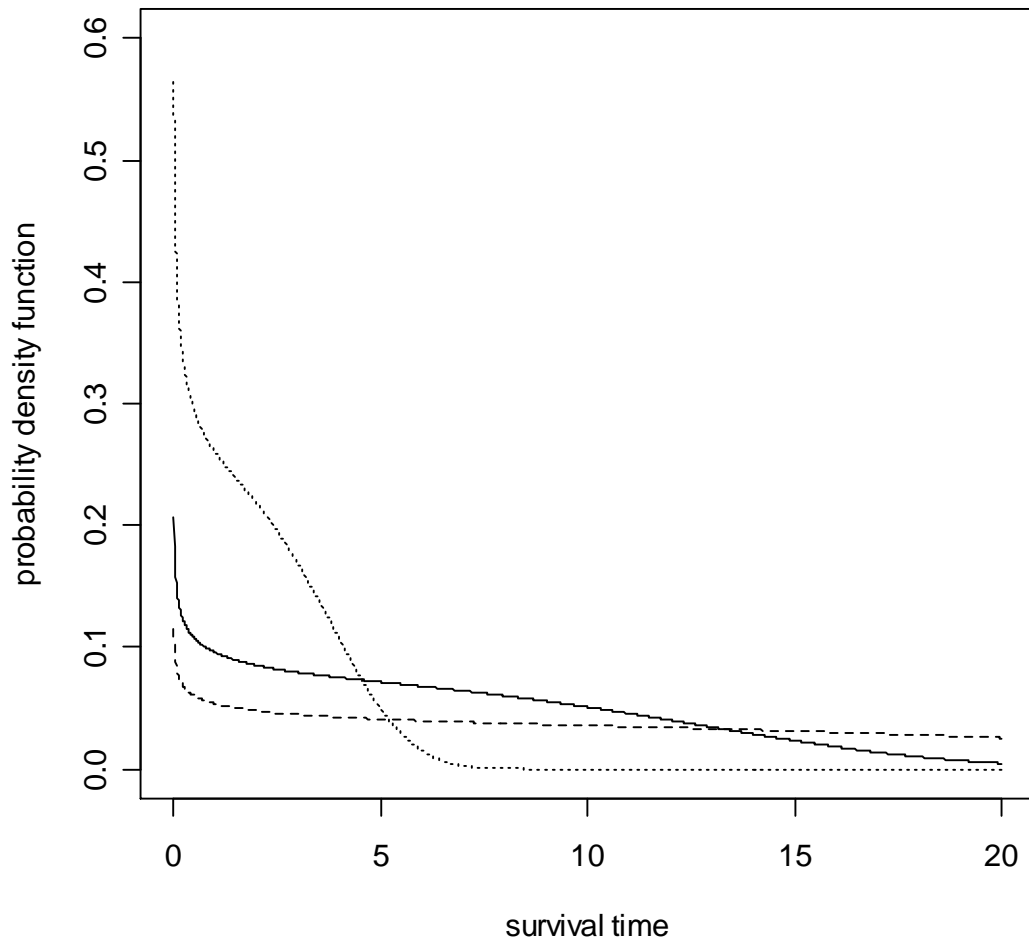
圖二 模擬中三種對數時間模型型態，模型 I (—) 為線性，模型 II (---) 為二次式，模型 III (···) 為週期函數。



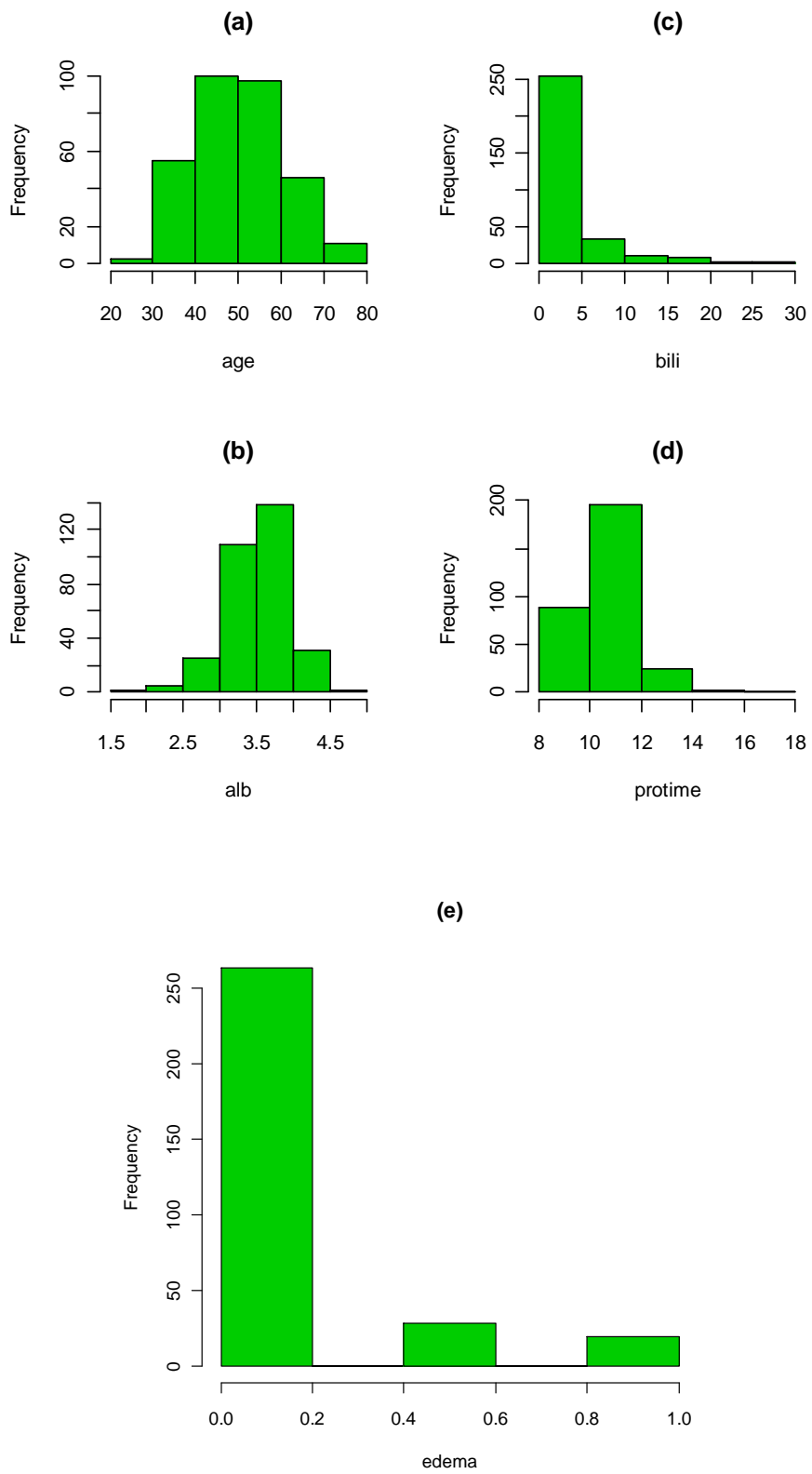
圖三 $\sigma=0.6, \lambda=0.3, x=1$ 時模擬中三種對數時間模型的機率密度函數，模型 I (—)為線性，模型 II (---)為二次式，模型 III (···)為週期函數。



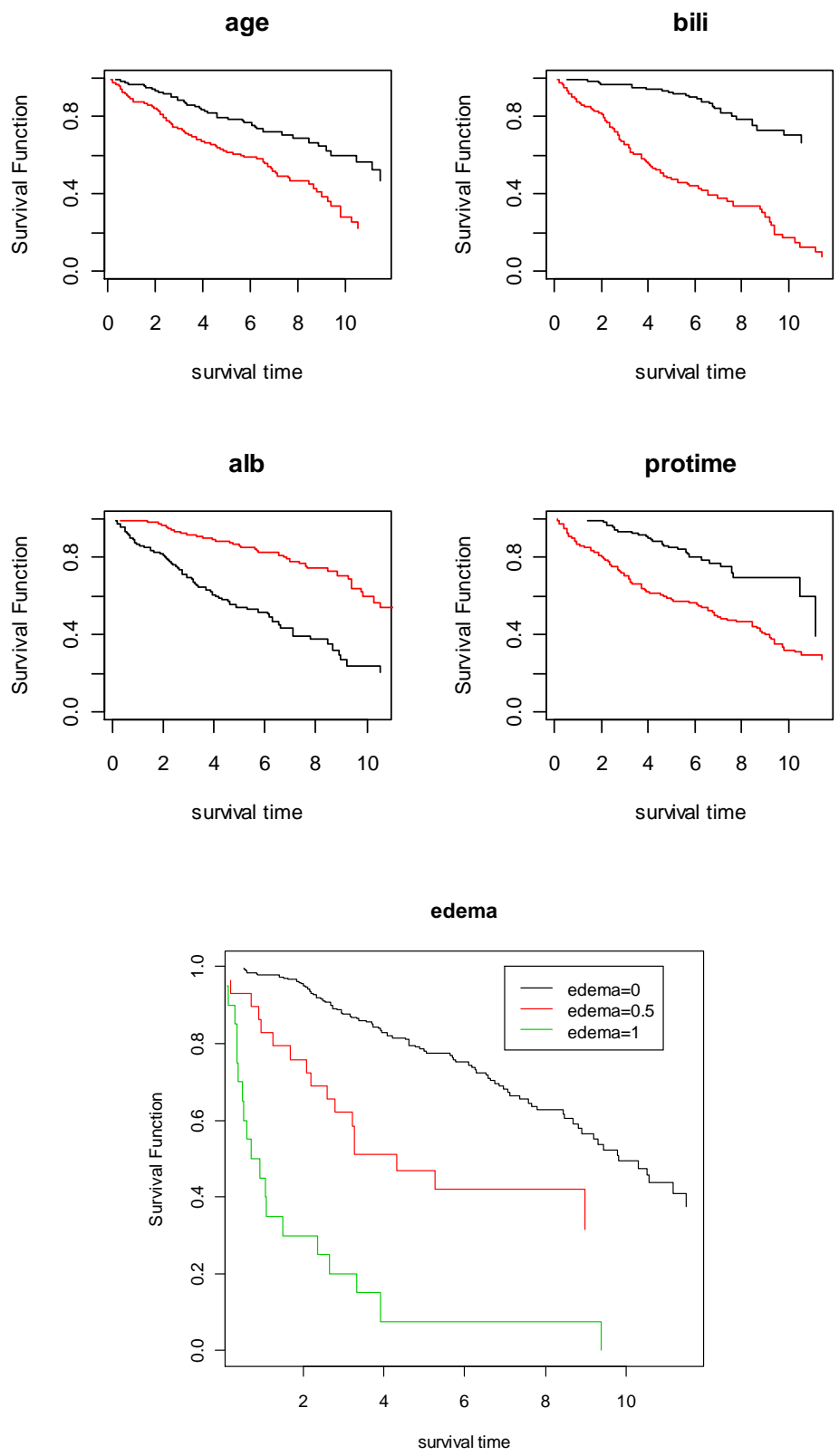
圖四 $\sigma=0.6, \lambda=1, x=1$ 時模擬中三種對數時間模型的機率密度函數，
 模型 I (—)為線性，模型 II (---)為二次式，模型 III (···)為
 週期函數。



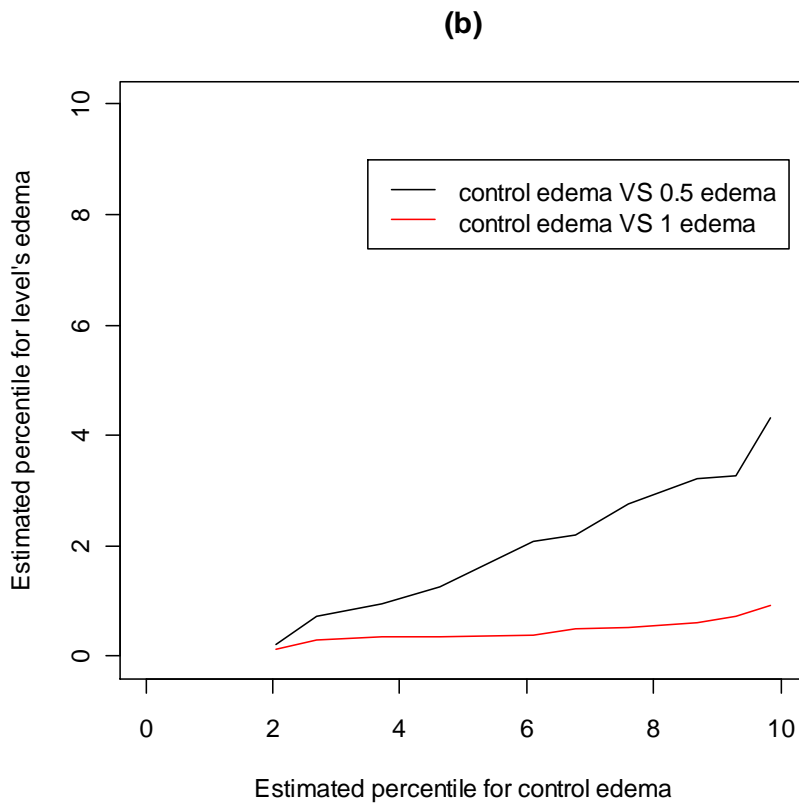
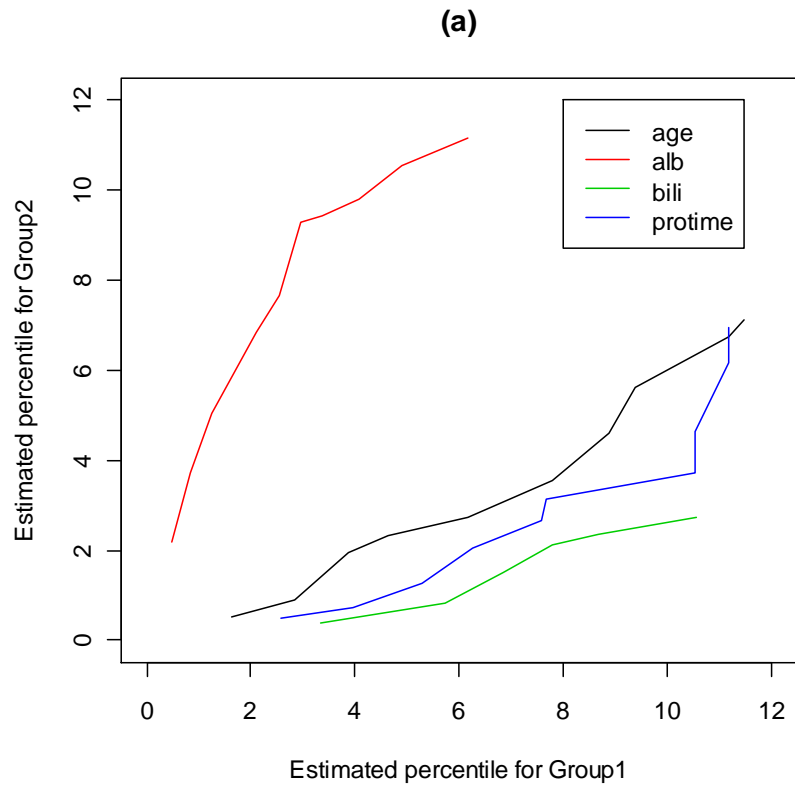
圖五 $\sigma=0.6, \lambda=2, x=1$ 時模擬中三種對數時間模型的機率密度函數，
 模型 I (—)為線性，模型 II (---)為二次式，模型 III (...)為
 週期函數。



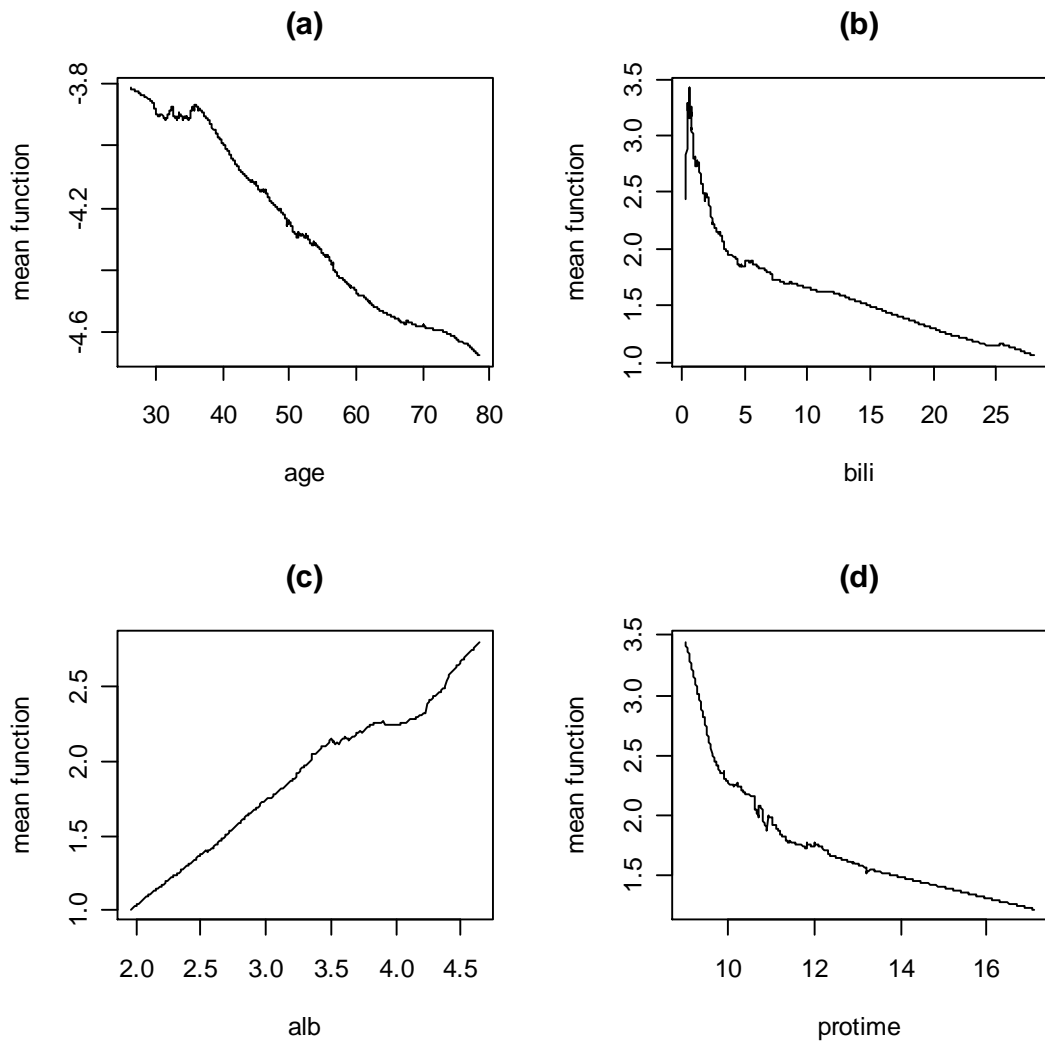
圖六 五個共變數分布之直方圖



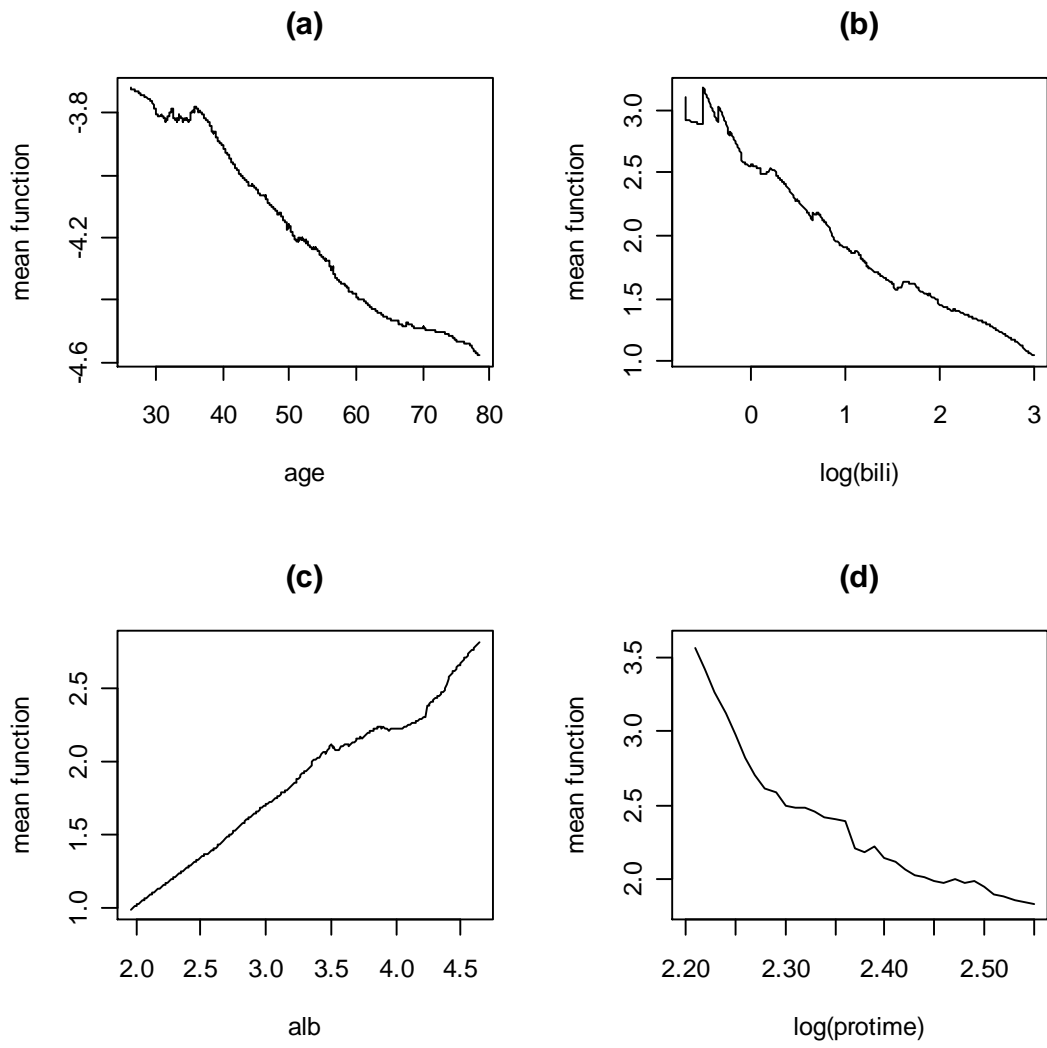
圖七 肝硬化資料之Kaplan-Meier存活函數，黑色與紅色分別為第一群與第二群



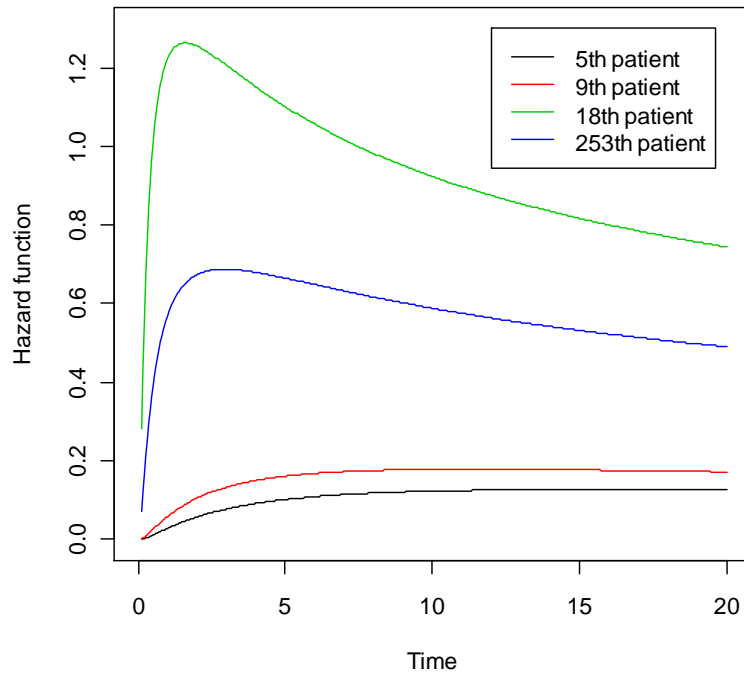
圖八 五個共變數下加速失敗時間模型之百分位點圖



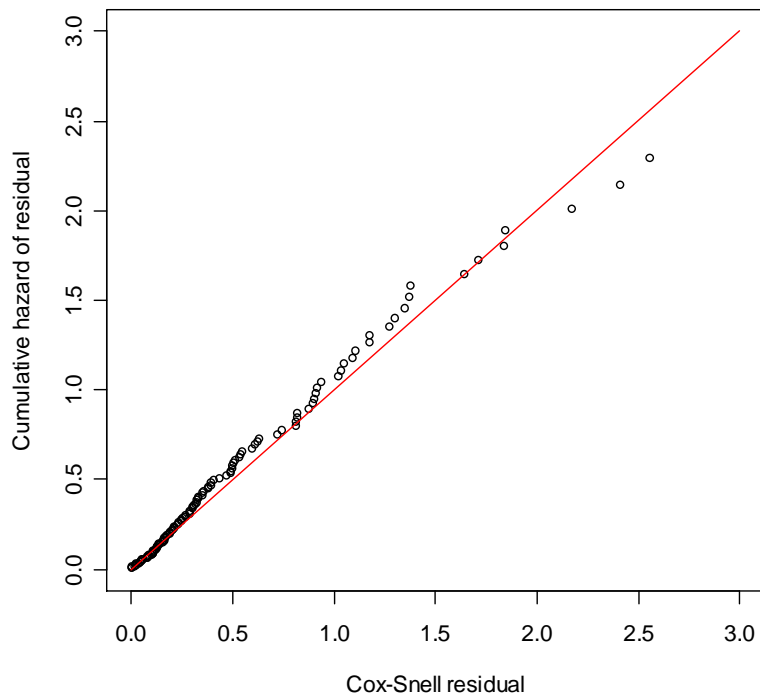
圖九 估計廣義伽瑪半參數AFT模型與共變數的關係



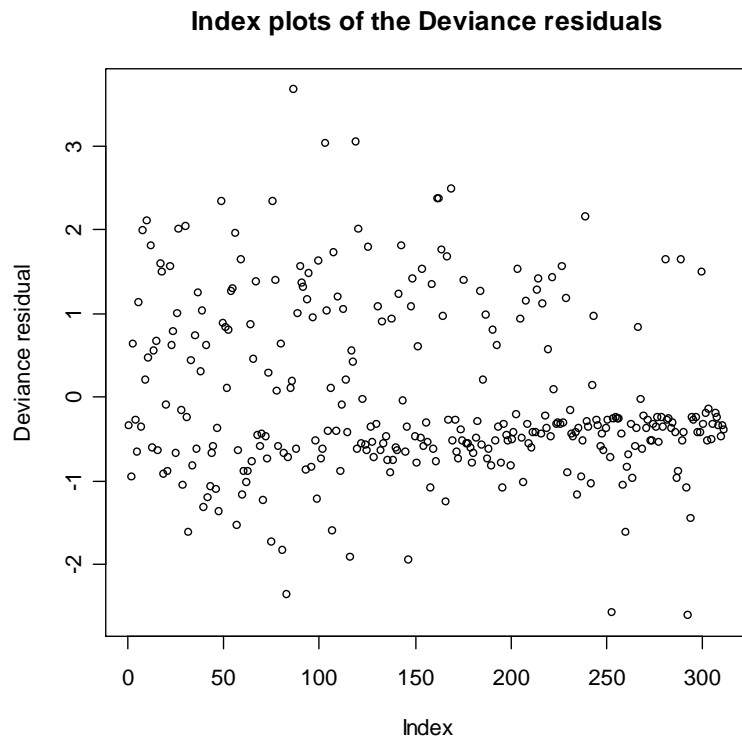
圖十 廣義伽瑪半參數AFT模型轉換後共變數的關係



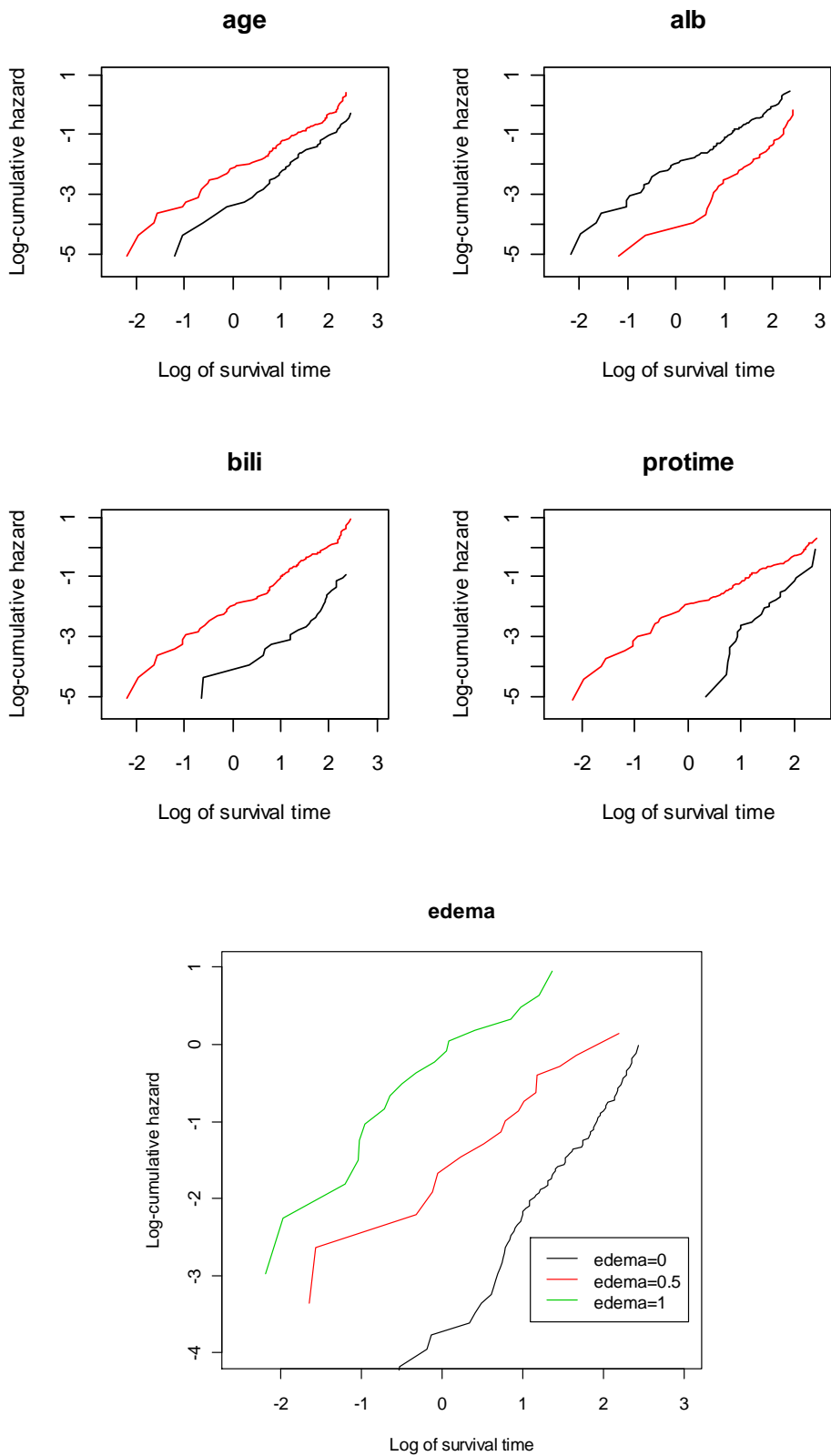
圖十一 廣義伽瑪 AFT 模型下病人的風險函數



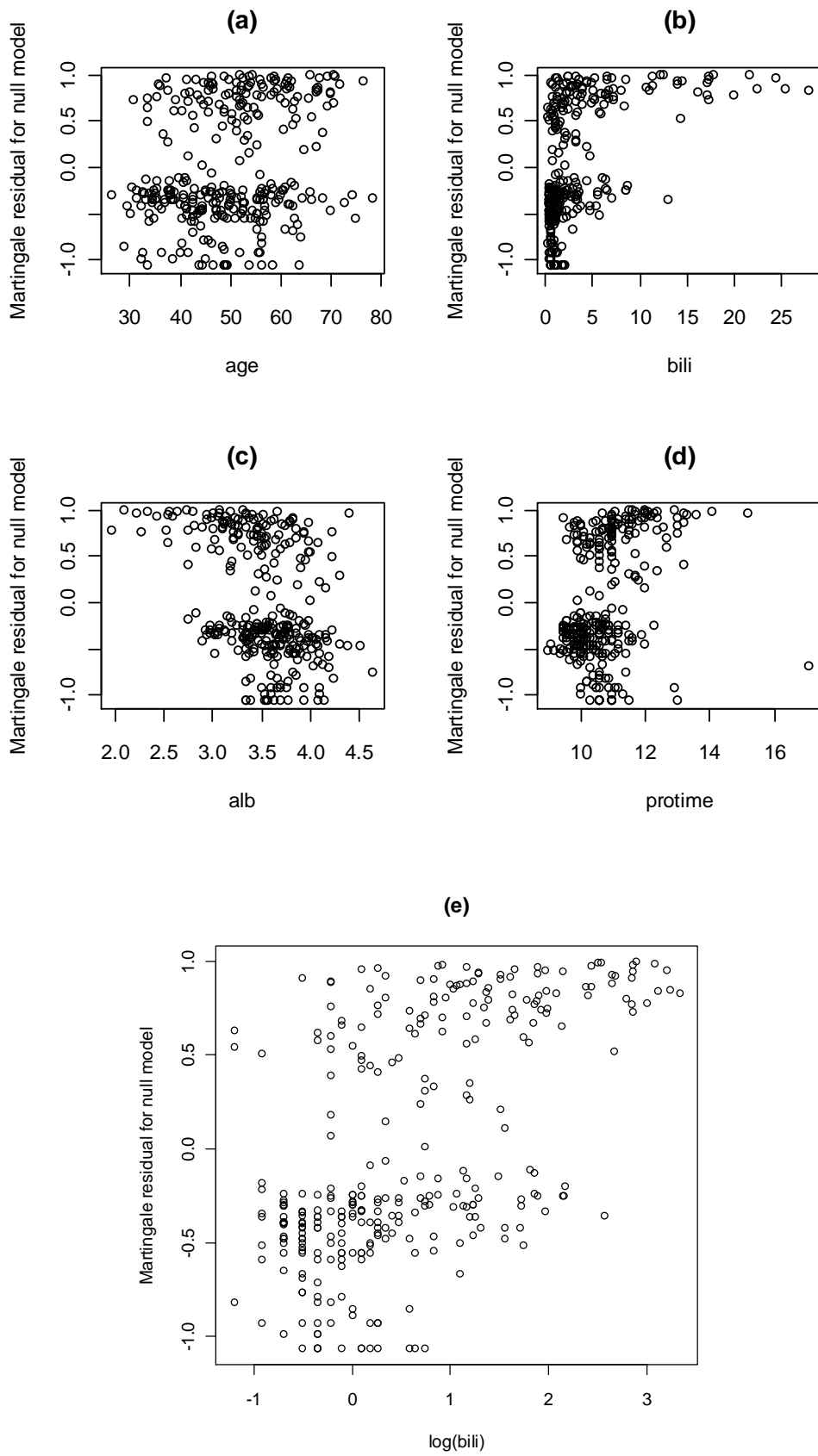
圖十二 共變數轉換後的廣義伽瑪 AFT 模型的 Cox-snell 殘差圖



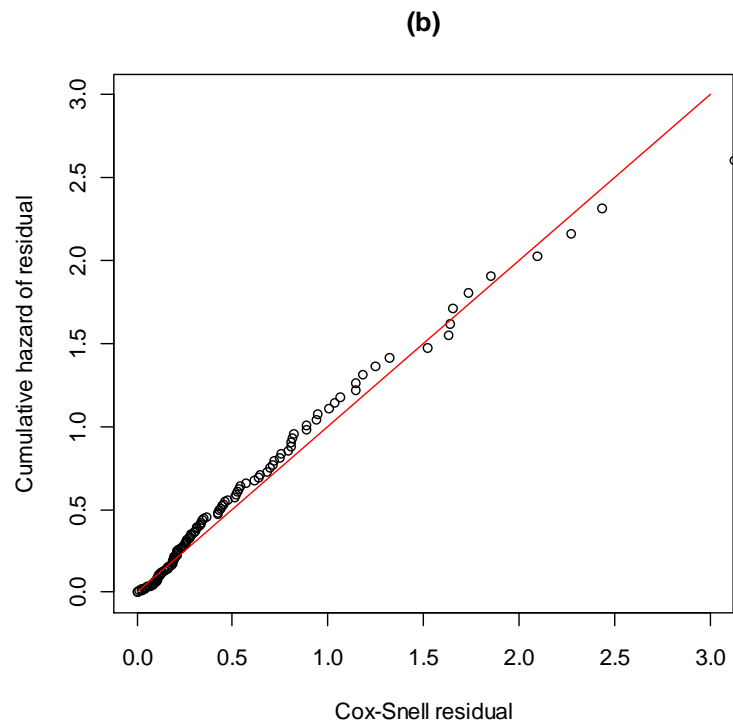
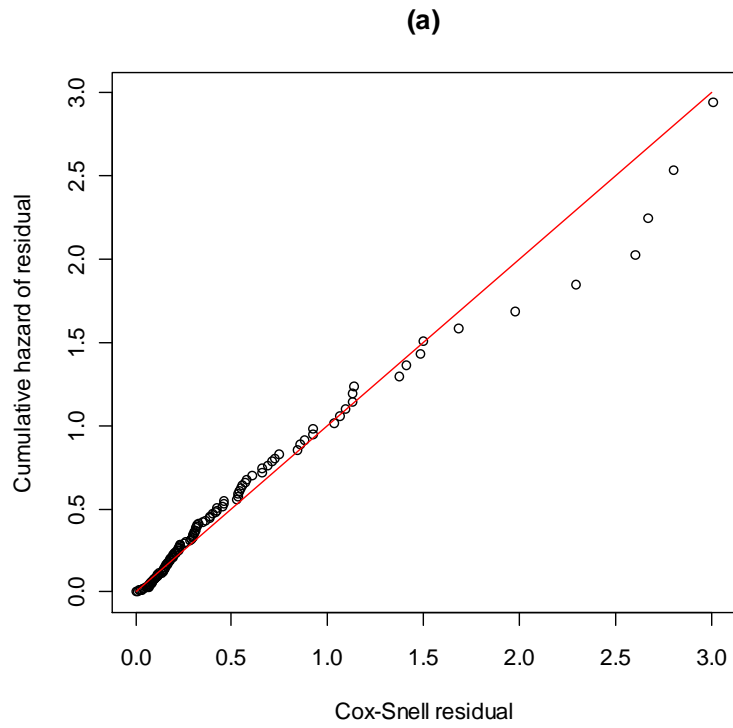
圖十三 共變數轉換後的廣義伽瑪 AFT 模型之離差殘差圖



圖十四 肝硬化資料之對數累積風險與對數時間關係圖，黑色與紅色分別為第一群與第二群



圖十五 未引入共變數之Cox比例危險模型的Martingale殘差圖



圖十六 Cox 比例風險模型的 Cox-snell 殘差圖，(a)與(b)分別為變數轉換前與變數轉換後

表一 廣義伽瑪加速失敗模型 $\sigma=0.6 \lambda=0.3$ 之下，估計中位存活時間的
偏誤(均方誤)

(a) 設限機率 30%

Model	n	x	線性韋伯	半參數韋伯	線性廣義伽瑪	半參數廣義伽瑪
I	100	0.5	1.108(3.048)	1.083(3.267)	0.504(2.104)	0.512(2.219)
		1	0.638(0.867)	0.631(0.907)	0.134(0.501)	0.126(0.535)
		1.5	0.281(0.432)	0.266(0.472)	-0.130(0.387)	-0.130(0.400)
	300	0.5	0.714(1.059)	0.695(1.042)	0.045(0.543)	0.043(0.511)
		1	0.595(0.510)	0.595(0.524)	0.027(0.156)	0.026(0.172)
		1.5	0.497(0.368)	0.494(0.377)	0.023(0.121)	0.024(0.126)
II	100	0.5	-0.123(4.626)	-0.654(4.072)	-1.398(5.821)	-1.202(4.787)
		1	-7.882(63.134)	-0.144(7.177)	-8.705(76.859)	-0.707(7.230)
		1.5	-1.590(3.667)	-0.944(2.017)	-2.127(5.498)	-1.370(2.906)
	300	0.5	0.077(1.522)	0.003(1.324)	-1.197(2.665)	-0.684(1.684)
		1	-7.721(59.975)	-0.223(2.281)	-8.564(73.719)	-0.968(2.995)
		1.5	-1.398(2.323)	0.075(0.529)	-1.962(4.165)	-0.537(0.730)
III	100	0.5	0.905(0.957)	0.476(0.392)	0.476(0.375)	0.308(0.239)
		1	2.052(4.366)	0.363(0.282)	1.669(2.937)	0.197(0.170)
		1.5	0.526(0.685)	0.915(1.590)	0.256(0.534)	0.682(1.166)
	300	0.5	0.835(0.741)	0.345(0.173)	0.390(0.197)	0.145(0.066)
		1	1.963(3.904)	0.368(0.188)	1.563(2.490)	0.164(0.070)
		1.5	0.337(0.235)	0.396(0.378)	0.040(0.135)	0.100(0.204)

(b)設限機率 50%

Model	n	x	線性韋伯	半參數韋伯	線性廣義伽瑪	半參數廣義伽瑪
I	100	0.5	0.874(3.457)	0.935(4.246)	0.531(3.018)	0.582(3.522)
		1	0.516(0.807)	0.502(0.900)	0.149(0.630)	0.144(0.695)
		1.5	0.235(0.408)	0.231(0.485)	-0.111(0.429)	-0.108(0.449)
	300	0.5	0.455(0.967)	0.457(0.950)	0.056(0.769)	0.060(0.719)
		1	0.478(0.406)	0.470(0.420)	0.0343(0.195)	0.0341(0.207)
		1.5	0.464(0.336)	0.463(0.360)	0.030(0.137)	0.032(0.146)
II	100	0.5	0.185(6.640)	-0.720(6.324)	-0.944(6.936)	-1.077(6.649)
		1	-0.805(65.997)	0.196(19.547)	-8.648(76.207)	-0.245(17.649)
		1.5	-1.974(4.939)	-1.022(2.427)	-2.262(6.191)	-1.330(3.191)
	300	0.5	0.644(2.649)	-0.120(2.045)	-0.631(2.267)	-0.561(2.223)
		1	-7.788(61.077)	-0.357(3.804)	-8.440(71.727)	-0.843(3.946)
		1.5	-1.774(3.485)	0.001(0.619)	-2.071(4.642)	-0.499(0.817)
III	100	0.5	0.871(0.896)	0.455(0.394)	0.510(0.440)	0.313(0.273)
		1	1.894(3.746)	0.358(0.292)	1.642(2.885)	0.209(0.192)
		1.5	0.180(0.517)	1.003(2.475)	0.132(0.657)	0.821(1.992)
	300	0.5	0.815(0.707)	0.331(0.166)	0.436(0.242)	0.148(0.074)
		1	1.805(3.308)	0.360(0.182)	1.532(2.404)	0.169(0.076)
		1.5	-0.019(0.135)	0.342(0.421)	-0.110(0.183)	0.122(0.290)

表二 廣義伽瑪加速失敗模型 $\sigma=0.6 \lambda=1$ 之下，估計中位存活時間的

偏誤(均方誤)

(a)設限機率 30%

Model	n	x	線性韋伯	半參數韋伯	線性廣義伽瑪	半參數廣義伽瑪
I	100	0.5	0.052(1.525)	-0.034(2.221)	0.082(1.662)	-0.004(2.317)
		1	0.019(0.378)	0.036(0.841)	0.047(0.470)	0.055(0.876)
		1.5	0.025(0.289)	-0.010(0.403)	0.051(0.350)	0.011(0.451)
	300	0.5	0.031 (0.463)	0.021 (0.649)	0.048 (0.497)	0.036 (0.678)
		1	0.006 (0.126)	-0.005 (0.276)	0.019 (0.151)	0.007 (0.298)
		1.5	0.0003 (0.096)	-0.019 (0.138)	0.010 (0.113)	-0.010 (0.150)
II	100	0.5	-0.962(4.415)	-1.703(4.960)	-1.622(6.023)	-1.635(5.060)
		1	-7.342(54.690)	-1.338(5.536)	-7.772(61.291)	-1.275(5.756)
		1.5	-1.800(4.109)	-1.508(2.925)	-2.081(5.168)	-1.441(2.880)
	300	0.5	-0.710(1.663)	-1.055(1.819)	-1.319(2.847)	-1.027(1.819)
		1	-7.224(52.459)	-1.502(3.498)	-7.620(58.369)	-1.471(3.493)
		1.5	-1.716(3.220)	-0.804(0.902)	-1.975(4.168)	-0.778(0.911)
III	100	0.5	0.456 (0.314)	0.210 (0.146)	0.280(0.203)	0.216(0.157)
		1	1.487 (2.333)	0.225 (0.130)	1.336 (1.918)	0.232 (0.143)
		1.5	-0.291 (0.418)	0.123 (0.536)	-0.384 (0.523)	0.135 (0.561)
	300	0.5	0.412 (0.203)	0.115 (0.043)	0.243 (0.096)	0.117 (0.047)
		1	1.414 (2.038)	0.215 (0.074)	1.265 (1.641)	0.218 (0.080)
		1.5	-0.472 (0.320)	-0.353 (0.238)	-0.576 (0.438)	-0.349 (0.240)

(b)設限機率50%

Model	n	x	線性韋伯	半參數韋伯	線性廣義伽瑪	半參數廣義伽瑪
I	100	0.5	0.109(2.784)	0.183(4.472)	0.100(2.922)	0.163(4.536)
		1	0.025(0.535)	0.096(1.359)	0.012(0.629)	0.082(1.386)
		1.5	0.022(0.341)	0.022(0.521)	0.012(0.419)	0.013(0.559)
	300	0.5	-0.182(0.807)	-0.140(1.109)	-0.172(0.818)	-0.130(1.109)
		1	-0.003(0.171)	0.020(0.417)	0.004(0.194)	0.030(0.429)
		1.5	0.178(0.143)	0.176(0.203)	0.185(0.168)	0.182(0.222)
II	100	0.5	-0.006(5.522)	-1.039(4.743)	-0.570(6.203)	-1.056(4.927)
		1	-7.378(55.427)	-1.089(10.183)	-7.701(60.559)	-1.078(11.148)
		1.5	-1.150(2.207)	-0.532(1.198)	-1.326(2.711)	-0.505(1.387)
	300	0.5	0.427(2.042)	-0.468(1.451)	-0.149(1.820)	-0.426(1.477)
		1	-7.223(52.520)	-1.442(4.522)	-7.510(56.812)	-1.405(4.471)
		1.5	-1.111(1.503)	0.147(0.401)	-1.236(1.815)	0.218(0.556)
III	100	0.5	0.455(0.326)	0.207(0.172)	0.317(0.251)	0.202(0.178)
		1	1.403(2.107)	0.231(0.154)	1.316(1.894)	0.228(0.163)
		1.5	-0.503(0.708)	0.298(1.255)	-0.490(0.789)	0.286(1.251)
	300	0.5	0.419(0.212)	0.113(0.048)	0.296(0.132)	0.113(0.052)
		1	1.321(1.789)	0.214(0.080)	1.240(1.587)	0.213(0.085)
		1.5	-0.708(0.622)	-0.316(0.316)	-0.719(0.655)	-0.318(0.323)

表三 廣義伽瑪加速失敗模型 $\sigma=0.6 \lambda=2$ 之下，估計中位存活時間的

偏誤(均方誤)

(a)設限機率30%

Model	n	x	線性韋伯	半參數韋伯	線性廣義伽瑪	半參數廣義伽瑪
I	100	0.5	-0.378(1.734)	-0.367(1.930)	0.204(1.664)	0.173(1.899)
		1	-0.552(0.661)	-0.040(0.842)	-0.052(0.466)	-0.611(0.585)
		1.5	-0.611(0.632)	-0.618(0.637)	-0.188(0.352)	-0.211(0.353)
	300	0.5	-0.623(0.850)	-0.614(0.824)	0.021(0.440)	0.007(0.455)
		1	-0.566(0.440)	-0.581(0.493)	0.012(0.149)	0.004(0.181)
		1.5	-0.483(0.321)	-0.487(0.323)	0.014(0.104)	0.005(0.110)
II	100	0.5	-1.678(5.215)	-1.974(5.600)	-1.535(5.229)	-1.393(3.739)
		1	-6.174(38.641)	-1.706(8.133)	-6.071(37.567)	1.076(7.682)
		1.5	-1.896(4.184)	-1.669(3.276)	-1.821(3.996)	-1.181(1.987)
	300	0.5	-1.410(2.828)	-1.442(2.778)	-1.134(2.327)	-0.660(1.207)
		1	-6.049(36.802)	-1.793(4.467)	-5.878(34.815)	-1.067(2.362)
		1.5	-1.841(3.582)	-1.182(1.628)	-1.734(3.226)	-0.503(0.546)
III	100	0.5	0.119(0.096)	-0.072(0.103)	0.229(0.157)	0.105(0.119)
		1	0.793(0.724)	-0.119(0.094)	0.895(0.916)	0.039(0.090)
		1.5	-0.191(0.333)	0.131(0.622)	-0.113(0.299)	0.339(0.705)
	300	0.5	0.096(0.035)	-0.137(0.048)	0.210(0.077)	0.084(0.039)
		1	0.748(0.590)	-0.123(0.042)	0.849(0.757)	0.088(0.037)
		1.5	-0.327(0.191)	-0.256(0.211)	-0.255(0.145)	0.027(0.140)

(b)設限機率50%

Model	n	x	線性韋伯	半參數韋伯	線性廣義伽瑪	半參數廣義伽瑪
I	100	0.5	0.229(4.124)	0.295(4.572)	0.340(4.185)	0.407(4.777)
		1	-0.300(0.759)	-0.257(1.037)	-0.106(0.711)	-0.090(0.965)
		1.5	-0.514(0.652)	-0.507(0.642)	-0.304(0.514)	-0.301(0.519)
	300	0.5	-0.155(1.028)	-0.150(1.061)	0.005(0.866)	0.030(1.003)
		1	-0.338(0.321)	0.310(0.383)	0.027(0.208)	0.025(0.267)
		1.5	-0.379(0.270)	-0.374(0.262)	-0.026(0.139)	-0.026(0.135)
II	100	0.5	-0.949(5.909)	-1.305(6.894)	-0.861(7.002)	-1.117(6.628)
		1	-6.026(37.167)	-0.991(13.397)	-6.011(37.143)	-0.753(14.418)
		1.5	-2.018(4.757)	-1.496(3.129)	-2.018(4.823)	-1.233(2.571)
	300	0.5	-0.719(2.106)	-0.993(2.562)	-0.416(2.246)	-0.690(2.081)
		1	-5.904(35.162)	-1.194(4.426)	-5.772(33.682)	-0.966(3.858)
		1.5	-1.957(4.045)	-0.993(1.366)	-1.908(3.859)	-0.514(0.710)
III	100	0.5	0.185(0.156)	0.004(0.173)	0.241(0.196)	0.097(0.186)
		1	0.833(0.838)	-0.077(0.132)	0.867(0.899)	0.008(0.126)
		1.5	-0.180(0.568)	0.524(2.404)	-0.183(0.553)	0.610(2.723)
	300	0.5	0.148(0.058)	-0.093(0.051)	0.256(0.107)	0.064(0.046)
		1	0.780(0.651)	-0.087(0.048)	0.847(0.761)	0.068(0.046)
		1.5	-0.326(0.241)	-0.034(0.342)	-0.325(0.220)	0.097(0.349)

表四 共變數轉換後兩種模型的 $-2\ln L$ 與 AIC

模型	$-2\ln L$	AIC
廣義伽瑪AFT模型	698.1646	714.164
廣義伽瑪半參數AFT模型	690.1788	728.674

表五 共變數轉換後配適廣義伽瑪AFT模型之 $-2\ln L$ 值

Vaiables in model	$-2\ln L$
None	901.2774
age	882.4816
alb	838.717
log(bili)	770.807
log(protime)	836.208
edema	841.2248
log(bili)+log(protime)	747.716
log(bili)+log(protime)+alb	722.287
log(protime)+alb	790.8812
log(bili)+alb	742.4396
log(bili)+log(protime)+alb+edema	712.5486
log(protime)+alb+edema	773.0108
log(bili)+alb+edema	729.005
log(bili)+log(protime)+edema	726.742
log(bili)+log(protime)+alb+edema+age	698.1646
log(protime)+alb+edema+age	768.086
log(bili)+alb+edema+age	711.428
log(bili)+log(protime)+edema+age	709.782
log(bili)+log(protime)+alb+age	706.1368

表六 共變數轉換後之廣義伽瑪 AFT 模型，其共變數分別為 age、alb、
 log(bili)、log(protime)、edema 的係數與 σ, λ 之估計值與標
 準差

	α_0	α_1	α_2	α_3	α_4	α_5	σ	λ
估計值	7.98	-0.02	0.56	-0.56	-2.57	-0.50	0.76	0.44
標準差	1.76	0.006	0.157	0.064	0.692	0.181	0.068	0.191