中國統計學報
第35卷第4期
八十六年十二月
415-430頁

設限迴歸係數估計量之變異數估計

劉惠美　　　　　蔡典龍
國立中興大學統計系　國立台灣大學流病所

摘　要

近年來引用線性迴歸模型探討存活資料的各種迴歸係數估計量當中，以Buckley和James (B-J) (1979)所提之估計量評價較佳。本文主要是引用Liu和Tsai (1996)之觀點，針對B-J估計量之變異數提出一個估計量，此估計量與Schneider和Weissfeld (S-W) (1986)及Weissfeld和Schneider (W-S) (1987)所提之變異數估計量是同義的；但所得之估計量遠較S-W估計量與W-S估計量簡化且易於計算。本文另考慮將上述吾人所提之估計量結合拔靴法以及將上述之新估計量加以修正等三個方向。透過電腦模擬評估上述各種估計量，以結合拔靴法之估計量表現最好。

關鍵詞：Buckley-James估計量，設限迴歸。

美國數學會分類索引：主要62G05；次要62J10。

1. 緒論

考慮線性迴歸模式

$$T_i = \alpha + X_i^T \underline{\beta} + \varepsilon_i\,,\ i = 1, 2, \ldots, n, \tag{1. 1}$$

其中ε_i爲互相獨立且同態於連續分配函數F，且其期望值(expectation)爲零，變異數(variance)爲有限之未知母數σ^2。$X_i^T = \left(x_{i1}, \ldots, x_{i(p-1)}\right)$是由解釋變數所構成的向量，$\underline{\beta}$是$(p-1)\times 1$的未知參數向量，$T_i$代表第$i$個個體的實際存活時間。然而，在實際狀況中通常我們只能觀察到Y_i之存活時間。若令G_i爲互相獨立之設限時間(censoring time)，服從於分配函數C_i，$i = 1, 2, \ldots, n$，則觀察之存活時間與實際存活時間之關係可表示爲

$$Y_i = \min(T_i, C_i)\ 。 \tag{1. 2}$$

設$\delta_i = 1$，當$T_i \le C_i$；$\delta_i = 0$，當$T_i > C_i$。則在實際實驗中吾人僅能觀察到(Y_i, δ_i, X_i^T)。

利用線性迴歸模型來分析存活資料，其基本問題是如何估計迴歸係數$\underline{\beta}$及α。在近二、三十年來，有一些學者投入此研究中，提出各自主張的迴歸係數估計量。由於普遍認爲Buckley和James (B-J) (1979)迴歸係數估計量表現得相當不錯，因而有些學者對其迴歸係數估計量之變異數估計方法感到相當的興趣。Buckley和James在同一篇文章中亦提出其變異數之估計量，之後有Schneider和Weissfeld (S-W) (1986)，Simth (1986)，Weissfled和Schnieder (W-S) (1987)和Hillis (1993a, 1993b, 1994)等人的研究。在Schneider和Weissfeld (1986)之文章中，利用電腦模擬比較S-W與B-J的估計量，發現S-W的估計量較好。而Weissfeld和Schneider之研究中，亦利用電腦模擬比較W-S與B-J、S-W估計量，發現S-W估計量較好。Hillis (1993a)則證明S-W和W-S的估計量是同義的(equivalent)。而Hillis (1993b, 1994)則是分別利用電腦模擬比較B-J、S-W、Smith的估計量。在簡單線性迴歸模型與多元線性迴歸模型上，發現Smith的估計量較好。由於Smith的估計量較複雜，而B-J、S-W、W-S的估計量在估計$\hat{Cov}(\underline{\hat{\beta}})$時，$(X^T\Delta X)^{-1}$部分均相同，其中$\Delta = Diag(\delta_i)$，只差別於$\sigma^2$之估計。所以本文主要目的是提出一個型式較簡單且易於解釋了解之σ^2的估計量。此σ^2的估計量之構建乃是引用Liu和Tsai (1996)對設限

迴歸係數估計的想法。進而證明，所提之簡化變異數估計量是與S-W估計量及W-S估計量同義。

Efron (1979)提出拔靴法(bootstrap)之後，拔靴法在許多領域被廣泛地研究，有關拔靴法使用在存活資料上的方式，可分成Efron (1981)與Reid (1981)兩種方式。其主要區分是Efron (1981)方式在做重覆取樣時，是來自於全部資料即包含設限與未設限的資料；而Reid方式在做重覆取樣時，僅考慮部分資料一未設限的資料。本文也考慮使用拔靴法，但爲求合理吾人將所提之估計量結合Efron (1981)方式的拔靴法來計算出變異數估計量。

本文共分成四節。第二節敘述較簡化之變異數估計量並說明此一估計量與S-W估計量同義。第三節爲實例與電腦模擬結果，第四節是結論與建議。

2. 較簡化之變異數估計量

由上節所述之(1.1)與(1.2)式，Buckley和James (1979)提出$\underline{\beta}$之估計量並同時對此估計量之共變異矩陣提出建議。Schneider和Weissfeld (1986)引用B-J估計量之想法，利用條件期望值取代設限資料(censored data)之作法，構建出不同於B-J所提出之共變異矩陣估計量。Liu和Tsai (1996, L-T)利用以上累加之平均值來估計B-J所述之條件期望值。此法與B-J所採用的P-L (product-limit)估計量不同，但發現與B-J之迴歸係數估計量同義。而Liu和Tsai提供之估計量不論在計算上及解釋上都較B-J方法簡捷且易懂。由於S-W估計量是引用B-J之方法，故吾人臆測若採用L-T之方法，當可得到型式較簡易之估計量且會同義於S-W估計量。

本節內容分爲兩部份。在第一部份中提出一較簡化之變異數估計量。第二部份嘗試將第一部份所述之估計量做適當的修改。

2.1 同義的變異數估計量($\hat{\sigma}^2_{L-T}$)

吾人針對設限迴歸係數估計量之變異數以Liu和Tsai (1996)之觀點提出下列之估計量。令$z_i = y_i - \alpha - x_i^T\underline{\beta}$，$i = 1, 2, \ldots, n$，則

$$\hat{Var}(\hat{\underline{\beta}}) = \hat{\sigma}^2_{L-T}(X^T \Delta X)^{-1},$$

其中

$$\hat{\sigma}^2_{L-T} = \frac{1}{n-p}\sum_{i=1}^{n}\left\{\delta_i\hat{z}_i^2 + (1-\delta_i)\hat{E}_{L-T}(\varepsilon_i^2|\varepsilon_i > z_i)\right\} \text{。} \tag{2.1}$$

這裡p爲參數的個數，$\hat{z}_i = y_i - \hat{\alpha}_{L-T} - x_i^T\underline{\hat{\beta}}_{L-T}$，$\hat{\alpha}_{L-T}$及$\underline{\hat{\beta}}_{L-T}$爲利用L-T所求得之$\alpha$及$\beta$的估計量，以及

$$\hat{E}_{L-T}(\varepsilon_i^2|\varepsilon_i > z_i) = \frac{\sum\limits_{\hat{z}_j>\hat{z}_i}\delta_j\hat{z}_j^2 + \sum\limits_{\hat{z}_j>\hat{z}_i}(1-\delta_j)\hat{z}^2_{c,j}}{\sum\limits_j I(\hat{z}_j > \hat{z}_i)}, \tag{2.2}$$

其中

$$\hat{z}^2_{c,j} = \frac{\sum\limits_{\hat{z}_t>\hat{z}_j}\delta_t\hat{z}_t^2 + \sum\limits_{\hat{z}_t>\hat{z}_j}(1-\delta_t)\hat{z}^2_{c,t}}{\sum\limits_t I(\hat{z}_t > \hat{z}_j)} \text{。} \tag{2.3}$$

從(2.1)式，吾人乃是利用Buckley和James之想法，以條件期望值取代設限資料。不同於S-W估計量的是(2.2)式是以上累加之平均值來估計條件期望值，非引用P-L估計量。在不失一般性之下，假設殘差項(residuals)已被排序(sorted)。令$\hat{z}_{(i)}$代表第i個順序之殘差，吾人可以證明下述定理。其相關的證明請參考Liu和Tsai(1996)：

<u>定理:</u>假設$\hat{z}_{(i)}$是設限資料的殘差。則

$$\hat{E}_{L-T}(\varepsilon^2_{(i)}|\varepsilon_{(i)} > z_{(i)}) = \hat{E}_{S-W}(\varepsilon^2_{(i)}|\varepsilon_{(i)} > z_{(i)}) \text{。}$$

由此一定理可知$\hat{\sigma}^2_{L-T} = k\hat{\sigma}^2_{S-W}$，其中$k$爲一常數值。換句話說，L-T估計量等價於S-W估計量。此一想法James (1986)曾經提及，只是他並未作進一步明確的探討。另外，在Hillis (1993a)之文章中，亦證明$\hat{\sigma}^2_{S-W}$與$\hat{\sigma}^2_{W-S}$是同義的。因此，$\hat{\sigma}^2_{L-T}$也同義於$\hat{\sigma}^2_{W-S}$。

2.2 修正$\hat{\sigma}^2_{L-T}(\hat{\sigma}^2_{L-Tadj})$

由於存活時間是連續變數，從(1.1)知ε_i也是連續變數。唯當以(2.2)式估計$E(\varepsilon_i^2|\varepsilon_i > z_i)$時是取所有比$\hat{z}_i$大的$\hat{z}_j$之平方做平均，但其中最小的$\hat{z}_j$仍與$\hat{z}_i$有一段差距。所以吾人嘗試作連續修正(countinuity correction)並修改(2.2)式，而產生如下之估計量：

$$\hat{E}_{L-Tadj}(\varepsilon_i^2|\varepsilon_i > z_i) = \frac{\sum\limits_{\hat{z}_j^* > \hat{z}_i^*} \delta_j \hat{z}_j^{*2} + \sum\limits_{\hat{z}_j^* > \hat{z}_i^*} (1-\delta_j)\hat{z}_{c,j}^{*\;2} + q\hat{z}_i^{*2}}{\sum\limits_j I(\hat{z}_j^* > \hat{z}_i^*) + q}, \tag{2.4}$$

其中

$$\hat{z}_{c,j}^{*\;2} = \frac{\sum\limits_{\hat{z}_t^* > \hat{z}_j^*} \delta_t \hat{z}_t^{*2} + \sum\limits_{\hat{z}_t^* > \hat{z}_j^*} (1-\delta_t)\hat{z}_{c,t}^{*\;2} + q\hat{z}_j^{*2}}{\sum\limits_t I(\hat{z}_t^* > \hat{z}_j^*) + q},$$

$0 < q < 1$，$\hat{z}_i^* = y_i - \hat{\alpha}_{L-Tadj} - x_i\underline{\hat{\beta}}_{L-Tadj}$，$\hat{\alpha}_{L-Tadj}$及$\underline{\hat{\beta}}_{L-Tadj}$爲利用修正Liu和Tsai法所求得$\alpha$及$\underline{\beta}$之估計量。因而

$$\hat{\sigma}_{L-Tadj}^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left\{\delta_i\hat{z}_i^{*2} + (1-\delta_i)\hat{E}_{L-Tadj}(\varepsilon_i^2|\varepsilon_i > z_i)\right\}。 \tag{2.5}$$

附帶一提的是(2.1)與(2.5)在沒有設限資料情況下，就是一般常見的$\hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-p} = MSE$。綜觀(2.1)與(2.5)式，在計算上比$\hat{\sigma}_{S-W}^2$簡單且易於明瞭。

3. 實例與電腦模擬

在第二節中吾人已提出二個估計量，在本節中吾人首先將此二個估計量以及Schmee和Hahn (1979)的估計量和拔靴法估計量運用在史丹佛心臟移植資料上。其次則將此四個估計量，透過電腦模擬的方式，從均方誤的觀點來比較它們的優劣。

3.1 實例

本節所採用的資料爲史丹佛心臟移植資料(Miller和Halpern, 1982)。史丹佛心臟移植計畫開始於1967年10月直到1980年2月爲止，共有184人接受心臟

移植，其中有些人做了多次心臟移植。資料內容包含這184人到1980年2月時，他們的存活時間、設限的或未設限的狀況、第一次心臟移植時的年齡和病人的T5互斥分數(mismatch score)。其中T5互斥分數是由史丹佛大學Charles Bieber博士所發展出來，它用來評估捐贈者的心臟與接受者免疫系統的互斥程度，以HLA抗原來判斷，當T5小於1時，表示相互接受程度良好，當T5大於1時，表示相互接受程度不太好。在這184人之中有27人的T5互斥分數是被遺漏掉了，因此只有157個病人記錄完整，可供迴歸分析之用。而在這157個病人中，於1980年2月時，有55個人依然活著，102個人已經死亡，所以55筆資料是設限的，102筆資料是未設限的。

採用線性迴歸模型爲

$$log_{10}T_i = \beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \varepsilon_i \ (i = 1, 2, \ldots, 152)$$

來說明，其中$\varepsilon_i \overset{iid}{\sim} F$，$i = 1, 2, \ldots, 152$，$E(\varepsilon_i) = 0$，$Var(\varepsilon_i)$爲某一個固定值。$C_i$來自於一機率分配函數$G_i$且彼此獨立，而$T_i$與$C_i$互相獨立，且

$$Y_i = \min(log_{10}T_i, C_i),\ \delta_i = \begin{cases} 1 & if \quad log_{10}T_i \le C_i \\ 0 & if \quad log_{10}T_i > C_i \end{cases},$$

其中$T_i (i = 1, 2, \ldots, 152)$是眞正的存活時間，$C_i (i = 1, 2, \ldots, 152)$是設限值。

在上述的線性迴歸模型之下，使用第二節所提出的兩個估計量(1)$\hat{\sigma}^2_{L-T}$、(2)$\hat{\sigma}^2_{L-Tadj}$，另外吾人考慮(3)修改Schmee和Hahn (1979)估計量運用在F爲常態分配上及(4)利用Liu和Tsai (L-T) (1996)之估計量結合拔靴法，使用Efron的方式；在表1及表2中分別以$\hat{\sigma}^2_{L-T}$、$\hat{\sigma}^2_{L-Tadj}$、$\hat{\sigma}^2_{S-Hadj}$、$\hat{\sigma}^2_{Bootstrap}$表示。並包含B-J、S-H等人的估計量，運用在此實際例子上。程式使用SAS/IML完成，計算結果列於表1。

在此實例的計算上，如果需要使用到遞迴(iteration)處理者，其收斂標準定爲

$$\max_{i=0,1,2} \left\{ \left| \frac{(\hat{\beta}_{i,j} - \hat{\beta}_{i,j+1})}{\hat{\beta}_{i,j+1}} \right| \right\} < 0.001\,,\ j = 0, 1, 2, \ldots,$$

其遞迴次數(j)均不超過50次。$\hat{\sigma}^2_{S-W}$、$\hat{\sigma}^2_{W-S}$與$\hat{\sigma}^2_{L-T}$是同義的，在此例子上也可以得到印證。表1中$\hat{\sigma}^2_{S-W}$與$\hat{\sigma}^2_{L-T}$僅差別$(\frac{n_u - p}{n_u} \times \frac{n}{n-p} = \frac{94}{97} \times \frac{152}{149})$的倍數

關係，$\hat{\sigma}^2_{S-W}$與$\hat{\sigma}^2_{W-S}$僅差別$(\frac{n_u-p}{n_u}=\frac{94}{97})$的倍數關係，$\hat{\sigma}^2_{W-S}$與$\hat{\sigma}^2_{L-T}$僅差別$(\frac{n}{n-p}=\frac{152}{149})$的倍數關係，所以的確此三種估計量是同義的。在表1中$\hat{\sigma}^2_{S-Hadj}$之值均是最大的；$\hat{\sigma}^2_{B-J}$之值則是最小的。就整體而言，不同的估計量有明顯的差異。可是由於目前這方面並未有嚴格的理論基礎，因此吾人無法斷定那種估計量較好，所以利用電腦模擬的方式，以便獲得更多的訊息，從均方誤的觀點來比較它們的優劣。

表2陳列若干不同再抽樣次數之拔靴法計算結果，這些結果產生方式乃是一次再抽樣10000次，然後分別看其再抽樣次數爲前25次，前50次，前75次，...，前1000次等的結果。

表1. 變異數估計量

估計量	Intercept	*Age*	*Age*^{2}
$\hat{\sigma}^2_{B-J}$	0.506768	0.001418	2.321E-7
$\hat{\sigma}^2_{S-W}$	0.687542	0.001924	3.149E-7
$\hat{\sigma}^2_{W-S}$	0.666277	0.001864	3.052E-7
$\hat{\sigma}^2_{S-H}$	0.728174	0.002037	3.336E-7
$\hat{\sigma}^2_{L-T}$	0.679692	0.001902	3.113E-7
$\hat{\sigma}^2_{Bootstrap}$ (B^*=10000)	0.631976	0.001910	3.330E-7
$\hat{\sigma}^2_{S-H\,\mathrm{adj}}$	0.891164	0.002493	4.082E-7
$\hat{\sigma}^2_{L-T\,\mathrm{adj}}$	0.670561	0.001876	3.072E-7

B*為拔靴法的再抽樣次數。

表2. 拔靴法變異數估計量

再抽樣次數(B)	Intercept	*Age*	*Age*^{2}
25	0.6602359	0.0021461	4.0954E-7
50	0.6103284	0.0018822	3.3861E-7
75	0.6310883	0.0020606	3.7771E-7
100	0.5790354	0.0019262	3.6012E-7
300	0.5667772	0.0017958	3.2305E-7
500	0.5592763	0.0017574	3.1491E-7
1000	0.5897165	0.0018254	3.2303E-7
10000	0.631976	0.0019099	3.3307E-7

3.2 電腦模擬

本節利用電腦模擬的方式，從均方誤的觀點來比較它們的優劣。本節採用線性迴歸模型爲：

$$T_i = 1.4 + 0.1X_i - 0.001X_i^2 + \varepsilon_i, (i = 1, 2, \ldots, 50)$$

來說明，其中X_i不是隨機變數，它由11～60，每一個間隔1，共有50個。因此$\varepsilon_i \stackrel{iid}{\sim} F$，$i = 1, 2, \ldots, 50$，$E(\varepsilon_i) = 0$，$Var(\varepsilon_i)$爲某一個固定值。$C_i$來自於一機率分配函數$G_i$且彼此獨立，而$T_i$與$C_i$互相獨立，且

$$Y_i = \min(T_i, C_i), \delta_i = \begin{cases} 1 & if \quad T_i \le C_i \\ 0 & if \quad T_i > C_i \end{cases},$$

其中$T_i(i = 1, 2, \ldots, 50)$是眞正的存活時間，$C_i(i = 1, 2, \ldots, 50)$是設限值。

在上述的線性迴歸模型之下，比較前述之四個估計量，其結果列於表3到表5。使用分配生成隨機變數，誤差項(error term)的分配F包含有(1)$\varepsilon_{i1} \sim N(0, 0.5^2)$、(2)$\varepsilon_{i2} \sim N(0, 0.25^2)$以及(3)$\varepsilon_{i3}$爲參合常態分配，其爲$0.1N(0, 0.5^2)$ $+0.9N(0, 0\ 25^2)$等三種。考慮參合常態分配被包含是爲了去了解在誤差項的分配是對稱的但有比常態分配$N(0, 0.25^2)$較重的尾部(heavier tails)時，估計量的表現情況是如何。在設限分配方面，採用平移的(shift)分配，即$G_i(c_i) = G(c_i - x_i\beta), \forall_i = 1, 2, \ldots, 50$，是爲了在每一個$X_i = x_i$之下發生設限的機率是相同的。此處分配$G$有(1)U(-2,4)、(2)U(0,2)、(3)U(1,1)、(4)Exp(3)-2。注意此處的指數分配乃是將Exp(3)往左平移二，以便取得與其它3個分配相同的期望值。採用上述四種分配，以便產生不同的設限值比率(設限值占全部樣本個數的比率)。因此在隨機變數的組合上，共有12種。

在此電腦模擬的計算上，如果需要使用到遞迴處理的部分，其收斂標準定爲

$$\max_{i=0,1,2} \left\{ \left| \frac{(\hat{\beta}_{i,j} - \hat{\beta}_{i,j+1})}{\hat{\beta}_{i,j+1}} \right| \right\} < 0.001, j = 0, 1, 2, \ldots,$$

其遞迴次數(j)以不超過30次爲原則，如果遞迴次數已達到30次，但仍然未達到收斂標準時，那就採取最後兩次估計量的平均數來做爲估計量。在每一個組合之下，樣本大小爲50，總共模擬1000次。

在使用拔靴法時，再抽樣次數爲100次。本模擬中在生成分配F、G與使用拔靴法時均有使用隨機變數，它們的種子數(seed)均不同。程式使用SAS/IML完成，隨機變數由Uniform、Normal、Ranexp等函數生成。所有的輸出結果請看表3到表5，迴歸係數的眞實值均爲$(\beta_0,\beta_1,\beta_2)=(1.4, 0.1, -0.001)$，迴歸係數估計量之變異數的眞實值分別是$(Var(\hat{\beta}_0),Var(\hat{\beta}_1),Var(\hat{\beta}_2))=$ (0.1949 , 0.000751 , $1.44E-7$)、(0.04874 , 0.000188 , $3.61E-8$)、(0.06336 , 0.000244 , $4.69E-8$)，每一個表是在同一個ε_i的分配之下，包含四種不同分配G之計算結果。

表3到表5中，估計值是指模擬所計算出1000次$\hat{Var}(\hat{\beta}_i)$之平均值；變異數是指模擬所計算出1000次$\hat{Var}(\hat{\beta}_i)$之變異數；偏誤是模擬所計算出1000次$\hat{Var}(\hat{\beta}_i)$之平均值減掉$Var(\hat{\beta}_i)$之眞實值；均方誤是偏誤之平方與變異數之和。

由表3到表5的輸出結果，吾人提出以下的結論:

(1) 在迴歸係數估計量的變異數估計方面，由均方誤的觀點來看，設限值比率從0.00006到0.0503，是$\hat{\sigma}^2_{L-Tadj}$表現最好；設限值比率從0.05302到0.48702，則是$\hat{\sigma}^2_{Bootstrap}$表現最好。

(2) 雖然吾人的誤差項ε_i的分配爲常態分配，但是$\hat{\sigma}^2_{S-Hadj}$並未表現得最好，也許會令有些人懷疑，可是必需留意的是資料的產生是由ε_i的分配F與分配G所決定出的。同樣地情形也在Heller和Simonoff (1990)的文獻中發生，雖然ε_i的分配爲常態分配，但是最好的估計量是B-J估計量，而不是S-H估計量。

(3) 不同的估計量在均方誤與估計值均有明顯的差異。

(4) 再從偏誤與變異數來看，設限值比率從0.00006到0.0503，$\hat{\sigma}^2_{L-Tadj}$之均方誤是最小，但其偏誤未必一定是最小(從絕對值來看)，有時是最小，有時不是，但是其變異數一定是最小；設限值比率從0.05302到0.48702，$\hat{\underline{\beta}}_{L-T}$結合拔靴法的均方誤是最小，但其偏誤未必一定是最小(從絕對值來看)，有時是最小，有時不是，但是其變異數一定是最小。所以這兩個估計量分別在這兩個設限值比率區間最穩定。

表3. 迴歸係數估計量之變異數估計($\varepsilon_{i1} \sim N(0,0.5^2)$)

		G: U(-2, 4), 設限值比率：0.3331			G: U(1, 1), 設限值比率：0.0224		
估計量	運算	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$
$\hat{\sigma}^2_{L-T}$	估計值	0.304822	0.001168	2.24E-07	0.192769	0.000744	1.43E-07
	變異數	0.012417	1.56E-07	5.63E-15	0.001699	2.50E-08	9.19E-16
	偏誤	0.109880	0.000416	7.98E-08	-0.002173	-7.85E-06	-1.48E-09
	均方誤	0.024491	3.29E-07	1.20E-14	0.001704	2.51E-08	9.21E-16
$\hat{\sigma}^2_{Bootstrap}$	估計值	0.257722	0.000986	1.89E-07	0.185029	0.000714	1.37E-07
	變異數	0.005835	8.58E-08	3.20E-15	0.002289	3.41E-08	1.26E-15
	偏誤	0.062780	0.000234	4.48E-08	-0.009914	-0.000038	-7.46E-09
	均方誤	0.009777	1.41E-07	5.21E-15	0.002387	3.55E-08	1.31E-15
$\hat{\sigma}^2_{S-H\,adj}$	估計值	0.315160	0.001207	2.32E-07	0.200214	0.000772	1.48E-07
	變異數	0.013363	1.66E-07	5.95E-15	0.002174	3.21E-08	1.18E-15
	偏誤	0.120217	0.000456	8.73E-08	0.005272	0.000021	4.04E-09
	均方誤	0.027815	3.74E-07	1.36E-14	0.002201	3.25E-08	1.20E-15
$\hat{\sigma}^2_{L-T\,adj}$	估計值	0.309030	0.001184	2.27E-07	0.192284	0.000742	1.42E-07
	變異數	0.012399	1.55E-07	5.57E-15	0.001662	2.45E-08	8.99E-16
	偏誤	0.114087	0.000433	8.29E-08	-0.002659	-9.72E-06	-1.84E-09
	均方誤	0.025415	3.42E-07	1.24E-14	0.001669	2.46E-08	9.02E-16

		G: U(0, 2), 設限值比率：0.1007			G: Exp(3)-2, 設限值比率：0.4830		
估計量	運算	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$
$\hat{\sigma}^2_{L-T}$	估計值	0.217405	0.000838	1.61E-07	0.405130	0.001551	2.99E-07
	變異數	0.002897	4.15E-08	1.53E-15	0.042774	4.95E-07	1.87E-14
	偏誤	0.022462	0.000087	1.67E-08	0.210188	0.000800	1.55E-07
	均方誤	0.003401	4.89E-08	1.81E-15	0.086953	1.13E-06	4.26E-14
$\hat{\sigma}^2_{Bootstrap}$	估計值	0.194181	0.000749	1.44E-07	0.326178	0.001242	2.38E-07
	變異數	0.002634	3.91E-08	1.44E-15	0.012380	1.78E-07	6.53E-15
	偏誤	-0.000762	-2.89E-06	-3.49E-10	0.131235	0.000491	9.37E-08
	均方誤	0.002635	3.91E-08	1.44E-15	0.029603	4.19E-07	1.53E-14
$\hat{\sigma}^2_{S-H\,adj}$	估計值	0.223498	0.000861	1.66E-07	0.431047	0.001650	3.18E-07
	變異數	0.003086	4.41E-08	1.62E-15	0.050635	5.98E-07	2.30E-14
	偏誤	0.028555	0.000110	2.12E-08	0.236104	0.000899	1.74E-07
	均方誤	0.003902	5.62E-08	2.08E-15	0.106381	1.41E-06	5.32E-14
$\hat{\sigma}^2_{L-T\,adj}$	估計值	0.215029	0.000829	1.59E-07	0.417888	0.001600	3.08E-07
	變異數	0.002734	3.91E-08	1.44E-15	0.044665	5.11E-07	1.93E-14
	偏誤	0.020087	0.000077	1.50E-08	0.222945	0.000848	1.64E-07
	均方誤	0.003137	4.51E-08	1.67E-15	0.094370	1.23E-06	4.62E-14

表4. 迴歸係數估計量之變異數估計($\varepsilon_{i2} \sim N(0,0.25^2)$)

估計量	運算	G：U(-2,4)，設限值比率：0.3327			G：U(1,1)，設限值比率：0.0001		
		$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$
$\hat{\sigma}^2_{L-T}$	估計值	0.075591	0.000290	5.55E-08	0.049019	0.000189	3.63E-08
	變異數	0.000729	8.97E-09	3.24E-16	0.000103	1.53E-09	5.65E-17
	偏誤	0.026856	0.000102	1.95E-08	0.000283	1.10E-06	2.11E-10
	均方誤	0.001450	1.93E-08	7.03E-16	0.000103	1.53E-09	5.65E-17
$\hat{\sigma}^2_{Bootstrap}$	估計值	0.066893	0.000258	4.95E-08	0.045844	0.000177	3.39E-08
	變異數	0.000428	6.17E-09	2.26E-16	0.000131	1.95E-09	7.19E-17
	偏誤	0.018157	0.000070	1.35E-08	-0.002891	-0.000011	-2.13E-09
	均方誤	0.000758	1.10E-08	4.07E-16	0.000139	2.07E-09	7.64E-17
$\hat{\sigma}^2_{S-H\ adj}$	估計值	0.078117	0.000299	5.74E-08	0.049026	0.000189	3.63E-08
	變異數	0.000791	9.80E-09	3.55E-16	0.000103	1.53E-09	5.66E-17
	偏誤	0.029381	0.000111	2.13E-08	0.000290	1.12E-06	2.16E-10
	均方誤	0.001654	2.22E-08	8.10E-16	0.000103	1.53E-09	5.66E-17
$\hat{\sigma}^2_{L-T\ adj}$	估計值	0.082415	0.000316	6.06E-08	0.049019	0.000189	3.63E-08
	變異數	0.000832	1.02E-08	3.67E-16	0.000103	1.53E-09	5.65E-17
	偏誤	0.033680	0.000128	2.45E-08	0.000283	1.10E-06	2.11E-10
	均方誤	0.001966	2.65E-08	9.67E-16	0.000103	1.53E-09	5.65E-17

估計量	運算	G：U(0,2)，設限值比率：0.0503			G：Exp(3)-2，設限值比率：0.4855		
		$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$
$\hat{\sigma}^2_{L-T}$	估計值	0.051842	0.000200	3.84E-08	0.101480	0.000388	7.45E-08
	變異數	0.000134	1.92E-09	7.03E-17	0.002345	2.61E-08	9.38E-16
	偏誤	0.003106	0.000012	2.34E-09	0.052744	0.000200	3.84E-08
	均方誤	0.000144	2.06E-09	7.57E-17	0.005127	6.60E-08	2.41E-15
$\hat{\sigma}^2_{Bootstrap}$	估計值	0.047140	0.000182	3.49E-08	0.088831	0.000339	6.51E-08
	變異數	0.000139	2.08E-09	7.71E-17	0.000946	1.41E-08	5.36E-16
	偏誤	-0.001595	-5.98E-06	-1.16E-09	0.040095	0.000151	2.90E-08
	均方誤	0.000142	2.12E-09	7.85E-17	0.002553	3.70E-08	1.38E-15
$\hat{\sigma}^2_{S-H\ adj}$	估計值	0.052350	0.000202	3.88E-08	0.108215	0.000413	7.95E-08
	變異數	0.000136	1.95E-09	7.16E-17	0.002750	3.09E-08	1.10E-15
	偏誤	0.003615	0.000014	2.72E-09	0.059480	0.000226	4.34E-08
	均方誤	0.000149	2.15E-09	7.90E-17	0.006288	8.17E-08	2.99E-15
$\hat{\sigma}^2_{L-T\ adj}$	估計值	0.051600	0.000199	3.82E-08	0.120801	0.000461	8.87E-08
	變異數	0.000131	1.87E-09	6.86E-17	0.003368	3.63E-08	1.27E-15
	偏誤	0.002865	0.000011	2.17E-09	0.072066	0.000273	5.26E-08
	均方誤	0.000139	1.99E-09	7.33E-17	0.008561	1.11E-07	4.04E-15

表5. 迴歸係數估計量之變異數估計(參合常態分配)

估計量	運算	G：U(-2,4)，設限值比率：0.3330			G：U(1,1)，設限值比率：0.0024		
		$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$
$\hat{\sigma}^2_{L-T}$	估計值	0.096924	0.000372	7.13E-08	0.063003	0.000243	4.67E-08
	變異數	0.001316	1.69E-08	5.99E-16	0.000202	3.00E-09	1.11E-16
	偏誤	0.033567	0.000127	2.44E-08	-0.000354	-1.29E-06	-2.23E-10
	均方誤	0.002443	3.31E-08	1.19E-15	0.000202	3.00E-09	1.11E-16
$\hat{\sigma}^2_{Bootstrap}$	估計值	0.086283	0.000332	6.39E-08	0.058922	0.000227	4.36E-08
	變異數	0.000887	1.30E-08	4.79E-16	0.000258	3.80E-09	1.41E-16
	偏誤	0.022926	0.000088	1.70E-08	-0.004435	-0.000017	-3.31E-09
	均方誤	0.001412	2.08E-08	7.68E-16	0.000277	4.10E-09	1.52E-16
$\hat{\sigma}^2_{S-H\,adj}$	估計值	0.100842	0.000387	7.42E-08	0.063352	0.000244	4.69E-08
	變異數	0.001463	1.89E-08	6.72E-16	0.000213	3.16E-09	1.17E-16
	偏誤	0.037485	0.000142	2.73E-08	-0.000004	5.34E-08	3.55E-11
	均方誤	0.002868	3.92E-08	1.42E-15	0.000213	3.16E-09	1.17E-16
$\hat{\sigma}^2_{L-T\,adj}$	估計值	0.103563	0.000397	7.62E-08	0.062995	0.000243	4.67E-08
	變異數	0.001407	1.79E-08	6.35E-16	0.000202	3.00E-09	1.11E-16
	偏誤	0.040207	0.000153	2.93E-08	-0.000361	-1.32E-06	-2.29E-10
	均方誤	0.003024	4.12E-08	1.49E-15	0.000202	3.00E-09	1.11E-16

估計量	運算	G：U(0,2)，設限值比率：0.0530			G：Exp(3)-2，設限值比率：0.4870		
		$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$	$Var(\hat{\beta}_0)$	$Var(\hat{\beta}_1)$	$Var(\hat{\beta}_2)$
$\hat{\sigma}^2_{L-T}$	估計值	0.066338	0.000256	4.92E-08	0.132305	0.000507	9.77E-08
	變異數	0.000354	5.18E-09	1.92E-16	0.004269	5.16E-08	1.83E-15
	偏誤	0.002982	0.000012	2.31E-09	0.068948	0.000263	5.08E-08
	均方誤	0.000363	5.32E-09	1.97E-16	0.009023	1.21E-07	4.42E-15
$\hat{\sigma}^2_{Bootstrap}$	估計值	0.060164	0.000232	4.46E-08	0.116074	0.000444	8.52E-08
	變異數	0.000324	4.68E-09	1.73E-16	0.002316	3.43E-08	1.28E-15
	偏誤	-0.003192	-0.000012	-2.27E-09	0.052717	0.000200	3.83E-08
	均方誤	0.000335	4.83E-09	1.78E-16	0.005096	7.41E-08	2.75E-15
$\hat{\sigma}^2_{S-H\,adj}$	估計值	0.066747	0.000257	4.95E-08	0.142202	0.000545	1.05E-07
	變異數	0.000331	4.84E-09	1.80E-16	0.005086	6.20E-08	2.24E-15
	偏誤	0.003391	0.000013	2.61E-09	0.078846	0.000301	5.81E-08
	均方誤	0.000342	5.01E-09	1.86E-16	0.011303	1.52E-07	5.62E-15
$\hat{\sigma}^2_{L-T\,adj}$	估計值	0.065899	0.000254	4.89E-08	0.150793	0.000577	1.11E-07
	變異數	0.000332	4.87E-09	1.80E-16	0.005114	6.01E-08	2.14E-15
	偏誤	0.002543	9.94E-06	1.98E-09	0.087437	0.000333	6.43E-08
	均方誤	0.000339	4.96E-09	1.84E-16	0.012759	1.71E-07	6.28E-15

(5) 從表3到表5來看，均方誤隨著設限值比率的增加而增加。

(6) 在迴歸係數估計量之變異數估計方面，這四種估計量只有在設限值比率低時(約0.05)，均方誤才絕大部分來自於估計量的變異數部分，偏誤相對於眞實值而言很小，表示這些估計量都估計得不錯，只是修改後的$\hat{\sigma}^2_{L-T}$估計量表現得最好；但是只要設限值比率稍大時，偏誤就相當嚴重。

4. 結論

在電腦運算上吾人發現$\hat{\sigma}^2_{S-W}$與$\hat{\sigma}^2_{W-S}$較$\hat{\sigma}^2_{L-T}$估計量費時，基本上這是合理的，因爲$\hat{\sigma}^2_{S-W}$與$\hat{\sigma}^2_{W-S}$在求取權數時，必需先計算P-L估計量，而$\hat{\sigma}^2_{L-T}$卻不需如此麻煩。事實上，$\hat{\sigma}^2_{L-T}$是以二階樣本動差來表示，此對於一般非統計學者，應是較易明瞭其估計之動機與意義。其次，除常數項倍數，$\hat{\sigma}^2_{L-T}$、$\hat{\sigma}^2_{S-W}$及$\hat{\sigma}^2_{W-S}$皆爲同義。

在電腦模擬結果來看，吾人觀察到當設限值比率較低時，以$\hat{\sigma}^2_{L-Tadj}$表現最好；當設限值比率不低時，$\hat{\sigma}^2_{Bootstrap}$表現最好。

自由度之大小，會影響σ^2估計值之大小，當自由度爲$n-p$, $\frac{SSE}{n-p}$，是σ^2的不偏估計量，此時資料是完全可觀察的，亦即沒有設限資料。由於$\hat{\sigma}^2_{S-W}$與$\hat{\sigma}^2_{L-T}$之間關係式爲$\hat{\sigma}^2_{L-T} = \frac{n_u-p}{n_u} \times \frac{n}{n-p}\hat{\sigma}^2_{S-W}$，所以在設限值比率很低時，自由度以$n-p$或是$n_u-p$來處理時，差別不大。此時$\hat{\sigma}^2_{L-Tadj}$表現得較佳是可以理解的；但是設限比率不是很低時，這時自由度會產生較大之處理的區別，在此情形下，吾人建議採用拔靴法來估計。

參考文獻

Buckley, J. and James, I. (1979). Linear regression with censored data. Biometrika, **66**, 429-36.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. The Annals of Statistics, **7**, 1-26.

Efron, B. (1981). Censored data and the bootstrap. Journal of the American Statistical Association, **76**, 312-19.

Heller, G. and Simonoff, J. S. (1990). A comparison of estimators for regression with a censored response variable. Biometrika, **77**, 515-20.

Hillis, S. L. (1993a). The equivalence of two Buckley-James variance estimators. Communications in Statistics, A, **22**, 479-84.

Hillis, S. L.(1993b). A comparison of three Buckley-James variance estimators. Communications in Statistics, B, **22**, 955-74.

Hillis, S. L. (1994). A heuristic generalization of Smith's Buckley-James variance estimators. Communications in Statistics, B, **23**, 813-31.

James, I. R. (1986). On estimating equations with censored data. Biometrika, **73**, 35-42.

Liu, H. and Tsai, D. L. (1996). A simpler estimate which is equivalent to the Buckley - James estimator. Proceedings of the Biometrics Section of the American Statistical Association, 212-216.

Miller, R. and Halpern, J. (1982). Regression with censored data. Biometrika, **69**, 521-31.

Reid, N. (1981). Estimating the median survival time. Biometrika, **68**, 601-8.

Schmee, J. and Hahn, G. J. (1979). A simple method for regression analysis with censored data. Technometrics, **21**, 417-32.

Schneider, H. and Weissfeld, L. (1986). Estimation in linear models with censored data. Biometrika, **73**, 741-5.

Smith, P. J. (1986). Estimation in linear regression with censored response. In Proceedings of the Pacific Statistical Congress (1986), Eds. I. F. Francis, B. F. J. Manly and F. C. Lam, pp. 261-265. Amsterdam: North-Holland.

Weissfeld, L. A. and Schneider, H.(1987). Inferences based on the Buckley-James procedure. Communications in Statistics, A, **16**, 1773-87.

[民國86年8月29日收稿，86年11月7日修訂，86年12月3日接受]

Estimation Based on Buckley-James Procedure

Huimei Liu
Department of Statistics
National Chung–Hsing University
D. L. Tsai
Institute of Epidemiology
National Taiwan University

ABSTRACT

The estimator proposed by Buckley and James (B-J) (1979) for estimating the linear regression coefficients with censored data has performed well in several studies. This article presents an estimator for the variance of the B-J estimator. The proposed estimator is equivalent to those given in Schneider and Weissfeld (1986) and Weissfeld and Schneider (1987), in some sense. But our estimator is much easier to be understood and needs less computation time.

Key words and phrases: Buckley-James estimator, censored regression.

AMS 1991 subject classifications: Primary 62G05; secondary 62J10.