บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงการประมาณค่าสัมประสิทธิ์การถดถอยพหุของตัวแบบการถดถอยเชิงเส้นด้วยวิธีกำลังสองน้อยที่สุด วิธีริดจ์รีเกรสชันและการประมาณ ridge estimators ในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงแบบเบ้จะทำการแปลงข้อมูลโดยการแปลงที่ใช้การยกกำลัง(power transformation)ของ BoxและCox เพื่อแปลงข้อมูลให้มีการแจกแจงเข้าใกล้ภาวะปกติเสียก่อน ซึ่งรายละเอียดต่างๆมีดังนี้

2.1 ตัวแบบทั่วไป

ตัวแบบทั่วไปของการวิเคราะห์การถดถอยพหุเชิงเส้นสามารถเขียนได้ดังนี้

(2.1.1) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด nx1, n คือจำนวนค่าสังเกต

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด nx(p+1), p = จำนวนตัวแปรอิสระ - 1

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด (p+1)x1

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อน ขนาด nx1 ซึ่งมีการแจกแจงแบบปกติ

โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2 I_n$

จากตัวแบบ(2.1)ประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยที่สุดจะได้ค่าประมาณ

(2.1.2) $$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$$

ซึ่งการประมาณ $\hat{\beta}$ โดยวิธีนี้จะได้ค่าตัวประมาณ β ที่มีคุณสมบัติไม่เอนเอียง กล่าวคือ

$$
\begin{aligned}
E(\hat{\beta}) &= E[(X'X)^{-1}X'y] \\
&= E[(X'X)^{-1}X'(X\beta + \varepsilon)] \\
&= E[(X'X)^{-1}X'X(\beta) + (X'X)^{-1}X'\varepsilon] \\
&= E[\beta] + (X'X)^{-1}X'E[\varepsilon] \\
&= \beta
\end{aligned}
$$

ตัวประมาณ $\hat{\beta}$ จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดในบรรดาตัวประมาณที่ไม่เอนเอียง แต่ในการประมาณค่า β ด้วยวิธีกำลังสองน้อยที่สุดมีข้อสมมุติฐานที่จำเป็นข้อหนึ่งคือ ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันซึ่งในทางปฏิบัติเป็นไปได้น้อย ในกรณีที่ตัวแปรอิสระมีความสัมพันธ์กันคือมีสภาพไม่เหมาะสม(ill-condition) การประมาณค่า β ด้วยวิธีกำลังสองน้อยที่สุดอาจจะไม่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด ในการพิจารณาผลของตัวแปรอิสระมีพหุสัมพันธ์กันสามารถพิจารณาจากคุณสมบัติ 2 ข้อของตัวประมาณค่า β กล่าวคือ เมตริกซ์ความแปรปรวนของตัวประมาณ β และค่าเฉลี่ยกำลังสองของความแตกต่างระหว่าง $\hat{\beta}$ และ β ซึ่งสามารถเขียนให้อยู่ในรูปของ $X'X$ และ σ^2 * ได้ดังนี้

(2.1.3) $\quad \mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$

ให้ L_1 คือความแตกต่างของ $\hat{\beta}$ และ β ดังนั้น

---

*ในกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์กันจะมีผลทำให้ $|X'X|$ เข้าใกล้ศูนย์ ซึ่งมีผลทำให้ $\mathrm{Var}(\hat{\beta})$ มากเพราะ $\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$และค่าเฉลี่ยกำลังสองของความแตกต่างระหว่าง $\hat{\beta}$ และ $\beta[E(L_1^2)]$ มากเพราะ $E(L_1^2) = \sigma^2\mathrm{trace}(X'X)^{-1}$

(2.1.4) $$L_1^2 = (\hat{\beta} - \beta)'(\hat{\beta}' - \beta)$$

ค่าเฉลี่ยกำลังสองของความแตกต่างระหว่าง $\hat{\beta}$ และ β มีค่าเป็น

(2.1.5) $$E[L_1^2] = \sigma^2 \mathrm{trace}(X'X)^{-1}$$

$$E[L_1^2] = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)]$$

$$= E[\hat{\beta}'\hat{\beta}] - \beta'\beta$$

(2.1.6) $$E[\hat{\beta}'\hat{\beta}] = \beta'\beta + \sigma^2 \mathrm{trace}(X'X)^{-1}$$

เมื่อ ε มีการกระจายแบบปกติ จะได้ว่า

(2.1.7) $$\mathrm{Var}(L_1^2) = 2\sigma^4 \mathrm{trace}(X'X)^{-2}$$

ให้ λ_i เป็น eigenvalue ของ $X'X$ เมื่อ $i = 1,2,..,p$ เมื่อ p คือ จำนวนตัวแปรอิสระ

$$\sum_{i=1}^{p} \lambda_i = \mathrm{Trace}\ (X'X)$$

โดยที่ $(\lambda_{max} = \lambda_1) \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \ldots \geqslant (\lambda_p = \lambda_{min}) \geqslant 0$

จากสมการ(2.1.5)และ(2.1.7)จะได้ว่า

$$E[L_1^2] = \sigma^2 \sum_{i=1}^{p} (1/\lambda_i)$$

$$\mathrm{Var}[L_1^2] = 2\sigma^4 \sum_{i=1}^{p} (1/\lambda_i)^2$$

ในกรณีที่ตัวแปรอิสระมีสภาพไม่เหมาะสมกล่าวคือ eigenvalue บางค่าของเมตริกซ์

$X'X$ จะมีค่าน้อยมากๆ ซึ่งมีผลทำให้ความแตกต่างระหว่าง $\hat{\beta}$ และ β มีมาก

2.2 การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีริดจ์รีเกรสชัน

Hoerl และ Kennard(1970:55-67) ได้เสนอวิธีริดจ์รีเกรสชันเพื่อแก้ปัญหา multicollinearity ซึ่งวิธีนี้จะให้ค่าประมาณสัมประสิทธิ์การถดถอยพหุที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำลง หลักการทั่วไปของวิธีนี้คือเมื่อพบว่า $|X'X|$ มีค่าเข้าใกล้ศูนย์ซึ่งมีผลทำให้ $(X'X)^{-1}$ หาค่าไม่ได้ ทำให้ผลบวกกำลังสองของ $\hat{\beta}$ มีค่ามากผิดจากความเป็นจริง ดังนั้นจึงพยายามทำให้ $|X'X|$ เพิ่มขึ้นโดยการบวกค่าคงที่ k ที่มากกว่าศูนย์เข้ากับสมาชิกทุกตัวบนเส้นทแยงมุม ซึ่งจะช่วยทำให้ characteristic root มีค่ามากขึ้นและผลบวกกำลังสองของ $\hat{\beta}$ ลดลง

กำหนดให้ค่าประมาณ β ด้วยวิธีกำลังสองน้อยที่สุดและวิธีริดจ์รีเกรสชันมีค่าเป็น $\hat{\beta}$ และ $\tilde{\beta}$ ตามลำดับ โดยที่

(2.2.1) $$\tilde{\beta}(k) = [X'X + kI]^{-1} X'y \quad ; k>0$$

และ $[X'X + kI]^{-1} = W$ ดังนั้นจากสมการ (2.2.1) จะได้ว่า

(2.2.2) $$\tilde{\beta}(k) = WX'y$$

เราสามารถจัดตัวประมาณค่า $\tilde{\beta}$ (k)ให้อยู่ในรูปของ $\hat{\beta}$ ได้ดังนี้

(2.2.3) $$\tilde{\beta}(k) = [I + k(X'X)^{-1}]^{-1}\hat{\beta}$$

(2.2.4) $$= Z\hat{\beta}$$

โดยที่ $Z = [I + k(X'X)^{-1}]^{-1}$

ให้ $E_i(W)$ และ $E_i(Z)$ เป็นค่า eigenvalue ของ W และZ ตามลำดับ ซึ่งได้จากการแก้สมการ characteristic equation

$$|W - E_i I| = 0$$

และ $$|Z - E_i I| = 0$$

(2.2.5) $$E_i(W) = 1/(\lambda_i + k)$$

(2.2.6) $$E_i(Z) = \lambda_i/(\lambda_i + k)$$

เราอาจเขียน Z ให้อยู่ในรูปฟังก์ชันของ W ได้ดังนี้

$$Z = I - k(X'X + kI)^{-1}$$

(2.2.7) $$= I - kW$$

จากสมการ (2.2.3) ผลบวกกำลังสองของ $\underset{\sim}{\beta}(k)$ จะมีค่าน้อยกว่าผลบวกกำลังสองของ $\hat{\underset{\sim}{\beta}}$ เมื่อ $k > 0$ กล่าวคือ

(2.2.8) $$\underset{\sim}{\beta}'(k)\underset{\sim}{\beta}(k) < \hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}$$

เราสามารถพิสูจน์ข้อความข้างบนได้ดังนี้

จากนิยาม : $\underset{\sim}{\beta}(k) = Z\hat{\underset{\sim}{\beta}}$ โดยที่ $X'X$ และ Z มีคุณสมบัติเป็นเมตริกซ์สมมาตรที่เป็นบวกแน่นอน (symmetric positive definite) จะได้ว่า

$$\underset{\sim}{\beta}'(k)\underset{\sim}{\beta}(k) = (Z\hat{\underset{\sim}{\beta}})'(Z\hat{\underset{\sim}{\beta}})$$

$$= \sum_{i=1}^{p} E_i^2(Z)\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}$$

(2.2.9) $$\leqslant E^2_{i\,(max)}(Z)\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}$$

เมื่อ $E_{max}(Z) = \lambda_1/(\lambda_1+k)$; λ_1 เป็น eigenvalue ที่มีค่ามากที่สุดของเมตริกซ์ $X'X$ ดังนั้น

$$\underset{\sim}{\beta}'(k)\underset{\sim}{\beta}(k) \quad < \quad \hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}$$

โดยที่ $\underset{\sim}{\beta}(k)$ มีคุณสมบัติดังนี้

ให้ $\underset{\sim}{\beta}(k)$ คือ ridge regression estimators ของ $\underset{\sim}{\beta}$ จากสมการนอร์มอล

$$(X'X + kI)\underset{\sim}{\beta} \quad = \quad X'\underset{\sim}{y}$$

โดยที่ k เป็นตัวคงที่ใดๆที่ไม่ติดลบคือ $0 \leqslant k \leqslant 1^{**}$ และ $\underset{\sim}{\beta}(k)$ จะมีค่าเท่ากับ $\hat{\underset{\sim}{\beta}}$ เมื่อ $k = 0$ และเมื่อ k มีค่าสูงขึ้น $\underset{\sim}{\beta}(k)$จะมีค่าลดลงเมื่อ $k \longrightarrow \infty$ แล้ว $\underset{\sim}{\beta}'(k)\underset{\sim}{\beta}(k) \longrightarrow 0$

$\underset{\sim}{\beta}(k)$เป็นการแปลงเชิงเส้น(linear transformation)ของตัวประมาณกำลังสองน้อยที่สุด กล่าวคือ

$$\begin{aligned}
\underset{\sim}{\beta}(k) &= (X'X + kI)^{-1}X\underset{\sim}{y} \\
&= (X'X + kI)^{-1}(X'X)(X'X)^{-1}X'\underset{\sim}{y} \\
&= (X'X + kI)^{-1}X'X[(X'X)^{-1}X'\underset{\sim}{y}] \\
&= [I + k(X'X)^{-1}]^{-1}\hat{\underset{\sim}{\beta}}
\end{aligned}$$

โดยที่ $\underset{\sim}{\beta}(k)$ สัมพันธ์กับ $\hat{\underset{\sim}{\beta}}$ ในลักษณะที่ว่า $\underset{\sim}{\beta}(k) = [I + k(X'X)^{-1}]^{-1}\hat{\underset{\sim}{\beta}}$ และเมื่อ $k=0$ จะได้ว่า $\underset{\sim}{\beta}(k) = \hat{\underset{\sim}{\beta}}$ กล่าวคือ ตัวประมาณกำลังสองน้อยที่สุดเป็นกรณีเฉพาะของริดจ์รีเกรสชัน เมื่อ

**ในกรณีที่ standardized variable จะได้ว่าเมตริกซ์ $X'X$ ก็คือ correlation matrix ซึ่งสมาชิกในแนวทแยงมุมจะมีค่าอยู่ระหว่าง 0 ถึง 1 ดังนั้นจึงทำการปรับค่าเมตริกซ์ $X'X$ ในกรณีเกิดปัญหา multicollinearity ด้วยการบวกค่าคงที่ k; $0 \leqslant k \leqslant 1$

$k = 0$ แต่ $\tilde{\underline{\beta}}(k)$ เป็นตัวประมาณที่เอนเอียงของ $\underline{\beta}$ โดยที่

$$E(\tilde{\underline{\beta}}(k)) = E(Z\hat{\underline{\beta}}) = Z\underline{\beta} = \underline{\beta}$$

และ

$$\begin{aligned}
Cov(\tilde{\underline{\beta}}(k)) &= E[\tilde{\underline{\beta}}(k) - E(\tilde{\underline{\beta}}(k))][\tilde{\underline{\beta}}(k) - E(\tilde{\underline{\beta}}(k))]' \\
&= E[\tilde{\underline{\beta}}(k) - Z\underline{\beta}][\tilde{\underline{\beta}}(k) - Z\underline{\beta}]' \\
&= E[Z(\hat{\underline{\beta}} - \underline{\beta})][Z(\hat{\underline{\beta}} - \underline{\beta})'] \\
&= E[Z(\hat{\underline{\beta}} - \underline{\beta})][(\hat{\underline{\beta}} - \underline{\beta})'(Z)'] \\
\hat{\underline{\beta}} - \underline{\beta} &= [(X'X)^{-1}X'(X\underline{\beta} + \underline{\varepsilon}) - \underline{\beta}] \\
&= \underline{\beta} + (X'X)^{-1}X'\underline{\varepsilon} - \underline{\beta} = (X'X)^{-1}X'\underline{\varepsilon} \\
Cov(\tilde{\underline{\beta}}(k)) &= E[Z(X'X)^{-1}X'\underline{\varepsilon}(\underline{\varepsilon}'X(X'X)^{-1}Z'] \\
&= Z(X'X)^{-1}X'E(\underline{\varepsilon}\underline{\varepsilon}')X(X'X)^{-1}Z' \\
&= \sigma^2 Z(X'X)^{-1}Z'
\end{aligned}$$

จากสมการ(2.2.3)และ(2.2.4)จะได้ว่า

$$\begin{aligned}
Z &= [I+k(X'X)^{-1}]^{-1} \\
&= [I+k(X'X)^{-1}]^{-1}(X'X)^{-1}(X'X) \\
&= \{X'X[I+k(X'X)^{-1}]\}^{-1}(X'X) \\
&= (X'X+kI)^{-1}(X'X)
\end{aligned}$$

ดังนั้น

$$\begin{aligned}
Cov(\tilde{\underline{\beta}}(k)) &= \sigma^2 Z(X'X)^{-1}Z' \\
&= \sigma^2(X'X+kI)^{-1}X'X(X'X)^{-1}(X'X)(X'X+kI)^{-1} \\
&= \sigma^2(X'X+kI)^{-1}(X'X)(X'X+kI)^{-1} \qquad (2.2.10)
\end{aligned}$$

จากการประมาณค่าสัมประสิทธิ์การถดถอย $\underline{\beta}$ ด้วยวิธีกำลังสองน้อยที่สุดจะได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองดังนี้

$$MSE(\hat{\beta}) = Var(\hat{\beta}) + (bias)^2$$

$$= \sigma^2(X'X)^{-1} + 0$$

เนื่องจาก $E(\hat{\beta}) = \beta$ ดังนั้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณสัมประสิทธิ์การถดถอยพหุด้วยวิธีริดจ์รีเกรสชัน จะมีค่าดังนี้

$$MSE(\hat{\beta}(k)) = \text{trace } Cov(\hat{\beta}(k)) + \beta'(Z-I)'(Z-I)\beta$$

$$= \text{Varience} + (\text{bias})^2$$

(2.2.11) $$= \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2} + k^2\beta'[X'X+kI]^{-2}\beta$$

เราสามารถพิสูจน์ (2.2.11) ได้ดังนี้

$$MSE(\hat{\beta}(k)) = E(\hat{\beta}(k)-\beta)'(\hat{\beta}(k)-\beta)$$

$$\hat{\beta}(k) = [I+k(X'X)^{-1}]^{-1}\hat{\beta} = Z\hat{\beta}$$

$$MSE(\hat{\beta}(k)) = E[Z\hat{\beta}-\beta]'[(Z\hat{\beta}-\beta)]$$

พิจารณาจาก $Z\hat{\beta}-\beta$ จะได้ว่า $Z\hat{\beta}-\beta = Z\hat{\beta}-Z\beta+Z\beta-\beta$

$$= Z(\hat{\beta}-\beta)+(Z-I)\beta$$

นั่นคือ $(Z\hat{\beta}-\beta)' = (\hat{\beta}-\beta)'Z'+\beta'(Z-I)'$

ดังนั้น $MSE(\hat{\beta}(k)) = E[(\hat{\beta}-\beta)'Z'+\beta'(Z-I)'][Z(\hat{\beta}-\beta)+(Z-I)\beta]$

$$= E[(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta)+(\hat{\beta}-\beta)'Z'(Z-I)\beta$$
$$+\beta'(Z-I)'Z(\hat{\beta}-\beta)+\beta'(Z-I)'(Z-I)\beta]$$

เนื่องจาก $E(\hat{\beta}-\beta) = 0$

ดังนั้น $MSE(\hat{\beta}(k)) = E[(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta)]+\beta'(Z-I)'(Z-I)\beta$

$$= a + b$$

โดยที่ $a = E(\hat{\beta}-\beta)Z'Z(\hat{\beta}-\beta)$

จะได้ว่า $\hat{\beta}-\beta = \beta+(X'X)^{-1}X'\varepsilon-\beta$

เมื่อ $a = E(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta) = E[\varepsilon'X(X'X)^{-1}Z'Z(X'X)^{-1}X'\varepsilon]$

$$= \sigma^2 \text{trace}[X(X'X)^{-1}Z'Z(X'X)^{-1}X']$$

(2.2.22) $$= \sigma^2 \text{trace}[(X'X)^{-1}Z'Z]^{***}$$

$Z = [I+k(X'X)^{-1}]^{-1}$ เป็นเมตริกซ์สมมาตร

$$(X'X)^{-1}Z'Z = (X'X)^{-1}ZZ = (X'X)^{-1}[I+k(X'X)^{-1}]^{-1}[I+k(X'X)^{-1}]^{-1}$$

$$= [\{I+k(X'X)^{-1}\}(X'X)]^{-1}[I+k(X'X)^{-1}]^{-1}$$

$$= (X'X+kI)^{-1}(X'X)(X'X)^{-1}\{I+k(X'X)^{-1}\}^{-1}$$

$$= (X'X+kI)^{-1}(X'X)[\{I+k(X'X)^{-1}\}(X'X)]^{-1}$$

(2.2.23) $$= (X'X+kI)^{-1}(X'X)(X'X+kI)^{-1}$$

กล่าวคือ จาก(2.2.22)และ(2.2.23)จะได้ $a = \sigma^2 \text{trace}[(X'X+kI)^{-1}(X'X)(X'X+kI)^{-1}]$

$$\text{MSE}(\hat{\beta}(k)) = \sigma^2 \text{trace}[(X'X+kI)^{-1}(X'X)(X'X+kI)^{-1}] + \beta'(Z-I)'(Z-I)\beta$$

$$= \text{trace}[\text{Cov}(\hat{\beta}(k))] + \beta'(Z-I)'(Z-I)\beta$$

$$= \text{trace}[\text{Cov}(\hat{\beta}(k))] + (\text{bias})^2$$

$$\text{MSE}(\hat{\beta}(k)) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2} + k^2\beta'(X'X + kI)^{-2}\beta$$

***trace(ABCDE)=trace(BCDEA) ดังนั้น $\text{trace}[X'(X'X)^{-1}Z'Z(X'X)^{-1}X]$

$= \text{trace}[(X'X)^{-1}Z'Z(X'X)^{-1}X'X] = \text{trace}[(X'X)^{-1}Z'Z]$

2.3 ตัวประมาณริดจ์ (Ridge Estimators)

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธีริดจ์รีเกรสชันนั้นจะต้องกำหนดค่าคงที่ k ซึ่งค่า k นี้จะอยู่ในช่วง 0 ถึง 1 การกำหนดค่า k นี้มีได้หลายวิธีด้วยกัน แต่จากการศึกษาของ Wichern และ Churchill, Gibbon, McDonald และ Galarneau, TZE-SAN-LEE พบว่าวิธีที่ดีในการประมาณ k มีอยู่ 3 วิธีคือ

1. วิธี HKB(Hoerl-Kennard-Baldwin method)

วิธีนี้เสนอโดย Hoerl Kennard และ Baldwin(1975) โดยกำหนดให้

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

เมื่อ p = จำนวนตัวแปรอิสระ

$\hat{\beta} = (X'X)^{-1}X'y$ เป็นค่าประมาณกำลังสองน้อยที่สุด

และ $\hat{\sigma}^2 = \frac{1}{n-p}(y'y - \hat{\beta}'X'y)$

2. วิธี TZE-SAN-LEE(TZE-SAN-LEE method)

วิธีนี้เสนอโดย TZE-SAN-LEE(1986) โดยกำหนดค่า k ดังนี้

$$k = \lambda_p$$

เมื่อ λ_p เป็นค่า eigenvalue ที่มีค่าน้อยที่สุดของ$(X'X)$

$$MSE(\hat{\beta}(k)) = \text{Varience} + (\text{bias})^2$$

$$= \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+k)^2 + k^2\beta'(X'X+kI)^{-2}\beta$$

จะได้ว่าค่า bias เพิ่มขึ้นเมื่อ k มีค่าเพิ่มขึ้น ดังนั้นจึงควรเลือก k ที่มีค่าน้อยๆ เพื่อทำให้ bias มีค่าน้อย กล่าวคือ

$$k = \lambda_p$$
$$= \lambda_{min}$$

3. <u>วิธี McD&G(McDonald-Galarneau method)</u>

วิธีนี้เสนอโดย McDonald และ Galarneau โดยพิจารณาจากข้อเท็จจริงที่ว่า $\hat{\beta}'\hat{\beta}$ มากกว่า $\beta'\beta$ อยู่ $\sigma^2 trace(X'X)^{-1}$ กล่าวคือ

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 trace(X'X)^{-1}$$

หรือ $E(\hat{\beta}'\hat{\beta})-\sigma^2 trace(X'X)^{-1} = \beta'\beta$

ดังนั้นจึงควรเลือกค่า k ที่มีผลทำให้ $\tilde{\beta}(k)$ มีค่าใกล้เคียงกับ β โดยการหาจากสมการ

$$\tilde{\beta}'(k)\tilde{\beta}(k) = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 trace(X'X)^{-1}$$

เมื่อ $\hat{\sigma}^2 = \dfrac{1}{n-p}(y'y - \hat{\beta}'X'y)$

2.4 <u>การแปลงข้อมูลโดยการแปลงที่ใช้การยกกำลัง(Power Transformation)</u>

ในกรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบเบ้จะทำการแปลงข้อมูลโดยการแปลงที่ใช้การยกกำลังเพื่อที่จะทำให้ข้อสมมุติของความเป็นปกติใช้ได้

BoxและCox(1964:211-243)ได้เสนอการแปลงข้อมูลของตัวแปรตามดังนี้

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda}-1}{\lambda} & ; \lambda \neq 0 \\ \log_e y_i & ; \lambda = 0 \end{cases}$$

ซึ่งการแปลงรูปตัวแปรตามในลักษณะนี้ทำให้ $Y_i^{(\lambda)}$ มีการแจกแจงแบบปกติซึ่งขั้นตอนการแปลงรูปตัวแปรตามมีดังนี้

1) จากสมการ $y_i = \beta_1+\beta_2X_{2i} + \ldots + \beta_pX_{pi}+\varepsilon_i$; $i=1,2,\ldots,n$ เมื่อ ε_i มิได้มีการแจกแจงแบบปกติ เรานำ geometric mean ของ $y_1,y_2,\ldots,y_n$ คือ $G = \sqrt[n]{y_1y_2\ldots y_n}$ มาหาร y_i ; $i=1,2,\ldots,n$ ทำให้ได้ $y_{i*} = y_i/G$; $i=1,2,\ldots,n$

จะพบว่า $\sum_{i=1}^{n} \ln y_{i*} = 0$

2) จากสมการ $\dfrac{y_{i*}^{\lambda}-1}{\lambda} = \beta_1+\beta_2X_{2i}+\beta_3X_{3i} + \ldots + \beta_pX_{pi}+\varepsilon_i$; $i=1,2,\ldots,n$

ซึ่ง Boxและcox ถือว่ากรณีเช่นนี้ $\varepsilon_i \sim N(0,\sigma^2)$ ทำให้

$$y_{i*}^{\lambda} \sim N(\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\ldots+\beta_pX_{pi},\sigma^2)$$

จะได้ว่าฟังก์ชันภาวะน่าจะเป็น(likelihood function)ของ $y_{i*}^{(\lambda)}$ คือ

$$L = \prod_{i=1}^{n} f(y_{i*}^{\lambda})$$

$$L = (1/2\pi\sigma^2)^{n/2}\exp\{-(1/2\sigma^2)\sum_{i=1}^{n}(y_{i*}^{(\lambda)}-\beta_1-\beta_2X_{2i}-\ldots-\beta_pX_{pi})^2\}.J$$

โดยที่ J คือ จาโคเบียนที่แปลงรูปจาก y_{i*} เป็น $y_{i*}^{(\lambda)}$ ดังนั้น

$$J = \prod_{i=1}^{n} \left| \frac{dy_{i*}^{(\lambda)}}{dy_{i*}} \right|$$

$$= \prod_{i=1}^{n} y_{i*}^{\lambda-1}$$

นั่นคือ $L = (1/2\pi\sigma)^{n/2} \exp\{-(1/2\sigma^2)\sum_{i=1}^{n}(y_{i*}^{(\lambda)}-\beta_1-\beta_2X_{2i}-\ldots-\beta_pX_{pi})^2\}\prod_{i=1}^{n} y_{i*}^{\lambda-1}$

$$\ln L = -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma-(1/2\sigma^2)\sum_{i=1}^{n}(y_{i*}^{(\lambda)}-\beta_1-\beta_2X_{2i}-\ldots-\beta_pX_{pi})^2+(\lambda-1)\sum_{i=1}^{n}\ln y_{i*}$$

แต่ $\sum_{i=1}^{n} \ln y_{i*}=0$ ดังนั้น

$$\ln L = \frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma-(1/2\sigma^2)\sum_{i=1}^{n}(y_{i*}^{(\lambda)}-\beta_1-\beta_2X_{2i}-\ldots-\beta_pX_{pi})^2$$

และ MLE ของ $\beta(\lambda)$ และ $\sigma^2(\lambda)$ ได้จาก $\frac{\ln L}{\beta} = 0$ และ $\frac{\ln L}{\sigma} = 0$ จะได้ว่า

$$\hat{\beta}(\lambda) = (X'X)^{-1}X'y_*^{(\lambda)}$$

$$\hat{\sigma}^2(\lambda) = \frac{1}{n}(y_*^{(\lambda)} - X\hat{\beta}(\lambda))'(y_*^{(\lambda)}-X\hat{\beta}(\lambda))$$

ดังนั้น ขั้นตอนต่างๆ มีดังนี้

1) หารค่า y แต่ละค่าด้วยค่ากลางเรขาคณิตของ y
2) สำหรับค่า λ ทำการถดถอย $y_{i*}^{(\lambda)}$ บน X_i และคำนวณ $\hat{\sigma}^2(\lambda)$
3) เลือกค่า λ ที่ให้ค่า $\hat{\sigma}^2(\lambda)$ ที่น้อยที่สุด จะได้ค่า λ ที่เป็น MLE ของ λ

2.5 ผลงานวิจัยที่เกี่ยวข้อง

ผลงานที่วิจัยที่เกี่ยวข้องกับตัวประมาณริดจ์ บางผลงานที่น่าสนใจมีดังนี้

Wichern และ Churchill(1978: 301-311)ได้ศึกษาเปรียบเทียบตัวประมาณริดจ์บางตัว ได้ผลสรุปว่าไม่มีวิธีใดที่จะดีที่สุดในทุกกรณี แต่วิธี McG&D โดยทั่วไปจะให้ค่า MSE ต่ำในกรณีที่มี multicollinearity ระดับปานกลางและสูง

Gibbons(1981:131-139)ได้ศึกษาตัวประมาณริดจ์บางตัวได้ผลสรุปว่า โดยทั่วๆไปแล้ววิธี HKB จะให้ผลดี แต่ไม่มีวิธีใดที่จะเด่นกว่าวิธีอื่นในทุกระดับความสัมพันธ์

Hoerl Kennard และ Baldwin(1975:105-123)ได้ศึกษาตัวประมาณริดจ์แบบ HKBได้ผลสรุปว่า

(1) การใช้ ridge estimator โดยวิธี HKB นี้จะให้ MSE น้อยกว่าวิธีกำลังสองน้อยที่สุดมากกว่า 50% ขึ้นไป

(2) ถ้าจำนวนตัวแปรอิสระเพิ่มขึ้น(p)จะทำให้ MSE มีค่าน้อยลง

(3) หากตัวแปรอิสระ X มีความสัมพันธ์กันสูง จะทำให้ MSE มีค่าน้อยลง

(4) MSE มีค่าน้อยลงขณะที่ σ^2 มีค่าเพิ่มขึ้น