บทที่ 2

ทฤษฎีที่เกี่ยวข้องในการวิจัย

2.1 Ridge Regression

2.1.1 เหตุผลและความสำคัญ

พิจารณาสมการถดถอยเชิงเส้น $Y = X\beta + U$ เมื่อ X คือ เมตริกซ์ขนาด $n \times (p + 1)$ $\underset{\sim}{\beta}$ คือ เวกเตอร์ของ Partial regression coefficient กล่าวคือ $\beta = (\beta_o , \beta_1, \ldots, \beta_p)$ และ $U = (u_1 , u_2,\ldots, u_n)$ คือเวกเตอร์ของ Stochastic term โดยเราควบคุมค่าของ u ไว้ด้วยข้อตกลงพื้นฐานว่า $E(U) = 0$ และ $E(U U') = \sigma^2 I_n$ นั้น โดยปกติเราสามารถหา BLUE (Best Linear Unbiased Estimator) ของ β_i , $i = 0 , 1, 2, \ldots , P$ หรือ เวกเตอร์ β ได้โดยวิธี Least Square หรือ Maximum Likelihood Method แต่ทั้งนี้ $\hat{\beta}_i$; $i = 0, 1, \ldots, P$ หรือเวกเตอร์ $\hat{\beta}$ จะเป็นตัวประมาณค่าที่ดีได้เฉพาะเมื่อเมตริกซ์ $X'X$ ซึ่งเสนอในรูปมาตรฐาน (Standardized form) มีลักษณะหรือธรรมชาติใกล้เคียง I_n ทั้งนี้เพราะถ้าตัวแปรอิสระ X_i ; $i = 1,2,\ldots,P$ เป็นอิสระเชิงเส้นต่อกัน (Linearly Independent) โดยที่ตัวแปรอิสระ X_i ได้รับการจัดรูปให้อยู่ในรูปมาตรฐานแล้ว เมตริกซ์ $X'X$ จะเป็น Unit matrix และเมื่อใดก็ตามที่เมตริกซ์ $X'X$ เบนห่างไปจาก Unit matrix มากขึ้น ปัญหาต่าง ๆ ก็จะติดตามมา ปัญหาที่สำคัญคือ ปัญหา Multicolinearity ซึ่งปัญหาดังกล่าวนี้จะอยู่ในระดับที่รุนแรงที่สุดเมื่อเมตริกซ์มีลักษณะของ Singular matrix ทั้งนี้เพราะปัญหา Multicolinearity มีผลให้ $\hat{\beta}_i$ ขาดความน่าเชื่อถือ เนื่องจาก $V(\hat{\beta}_i)$ มีค่าสูงมากทำให้นักวิจัยที่ไม่ระมัดระวังคัดเอา X_i ทิ้งไปทั้ง ๆ ที่ X_i อาจเป็นปัจจัยหลักของสมการถดถอยที่กำลังเกี่ยวข้องอยู่โดยนัยของการตรวจสอบนัยสำคัญทางสถิติของ β_i (โดยปกตินิยมใช้ t-test) และในอีกประการหนึ่งก็คือ ปัญหา Multicolinearity ทำให้ไม่อาจตัดสินใจได้ว่าปัจจัยใดควรเป็นปัจจัยที่ใช้ควบคุมความเคลื่อนไหวของตัวแปรตาม ทั้งนี้เพราะตัวแปรอิสระต่าง ๆ มีความสัมพันธ์กันในอัตราส่วนที่ค่อนข้างสูง และยิ่งตัวแปรอิสระหลายคู่ที่มีความสัมพันธ์ (สหสัมพันธ์) กันในเกณฑ์สูง ก็จะยิ่งมีปัญหาในการตัดสินใจและอภิปรายผล ทางออกที่นิยมประพฤติปฏิบัติก็คือ การคัดเอาตัวแปรบางส่วนโดยเฉพาะในคู่ที่มีสหสัมพันธ์สูง ๆ ทิ้งไปคู่ละหนึ่งตัวโดยอาศัย Priori information เป็นเกณฑ์

เพื่อทำลายความสัมพันธ์ระหว่างตัวแปรอิสระ X_i ผลที่ติดตามมาก็คงไม่ต่างไปจากการคัดตัวแปรอิสระ X_i ที่ไม่มีนัยสำคัญทางสถิติทิ้งไปดังที่กล่าวมาแล้ว การแก้ปัญหา Multicolinearity ตามแนวทางทั้ง 2 ประการข้างต้นจึงก่อให้เกิดปัญหาอย่างน้อย 2 ประการคือ

1. การคัดตัวแปรอิสระ X_i ทิ้งไปบางส่วนอาจทำให้ตัวแปรที่สำคัญถูกกำจัดโดยความรู้เท่าไม่ถึงการณ์

2. ทำให้สมการประมาณค่า $Y = X\hat{\beta}$ หย่อนคุณภาพในเชิงพยากรณ์เมื่อพิจารณาในรูปของ R^2 ที่ลดปริมาณลงกว่าเดิมก่อนคัดตัวแปรอิสระทิ้งไปบางส่วนตามข้อ 1

ปัญหาทั้งหลายดังที่กล่าวมาสามารถพิสูจน์ให้เห็นได้ดังนี้

จากสมการถดถอย $Y = X\beta + U$ โดยที่ $E(U) = 0$ และ $E(UU') = \sigma^2 I_n$ เราจะพบว่าค่าประมาณของ β คือ

$$\hat{\beta} = (X'X)^{-1} X'Y \qquad \text{.............(1)}$$

และเมื่อพิจารณาคุณสมบัติของ $\hat{\beta}$ จะพบว่า

(1) $E(\hat{\beta}) = \beta$ ....(2)

(2) $Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ ....(3)

(3) $\hat{\beta}$ เป็น BLUE ....(4)

(4) Expected Square Length ของระยะห่างระหว่างเวกเตอร์ $\hat{\beta}$ และ β มีค่าเท่ากับ $\sigma^2 \sum_{i=1}^{p} (1/\lambda_i)$ เมื่อ λ_i คือEigen value ของเมตริกซ์ $X'X$ หรือ Expected Square Length ของเฉพาะเวกเตอร์ $\hat{\beta}$ มีค่าเท่ากับ $\beta'\beta + \sigma^2 \text{ trace } (X'X)^{-1}$

<u>หมายเหตุ</u> การศึกษาและวิเคราะห์ในลำดับต่อจากนี้เป็นต้นไปจะศึกษาเฉพาะกรณีของ Standardized variable เท่านั้น

สำหรับคุณสมบัติข้อที่ 4 สามารถพิสูจน์ได้ดังนี้

กำหนดให้ $L_1 = (\hat{\beta} - \beta)$ คือ Error vector หรือระยะห่างระหว่างเวกเตอร์ของค่าประมาณ (Estimated vector) คือ $\hat{\beta}$ กับเวกเตอร์ของค่าจริงของพารามิเตอร์ (True parametor vector) คือβ

ดังนั้น Square Length ของ Error vector คือ $L_1^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta)$

ดังนั้น
$$
\begin{aligned}
E(L_1^2) &= E\left[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)\right] \\
&= E\left[\{(X'X)^{-1} X'U\}'\{(X'X)^{-1} X'U\}\right] \\
&= E\left[\{U'X(X'X)^{-1}\}\{(X'X)^{-1} X'U\}\right] \\
&= \sigma^2 \text{ trace } \left[X(X'X)^{-1}(X'X)^{-1} X'\right] \\
&= \sigma^2 \text{ trace } (X'X)^{-1} \quad \ldots\ldots\ldots\ldots\ldots(5)
\end{aligned}
$$

แต่ $(\hat{\beta} - \beta)'(\hat{\beta} - \beta) = \hat{\beta}'\hat{\beta} - 2\beta'\hat{\beta} + \beta'\beta$

ดังนั้น
$$
\begin{aligned}
E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) &= E(\hat{\beta}'\hat{\beta}) - 2\beta' E(\hat{\beta}) + \beta'\beta \\
&= E(\hat{\beta}'\hat{\beta}) - \beta'\beta
\end{aligned}
$$

นั่นคือ $E(L_1^2) = \sigma^2 \text{ trace } (X'X)^{-1} \quad \ldots.(6)$

หรือ $E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \text{ trace } (X'X)^{-1} \quad \ldots(7)$

แสดงว่า $E(L_i^2) = \sigma^2 \text{ trace } (X'X)^{-1}$ หรือ $E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \text{trace}(X'X)^{-1}$ ซึ่งชี้ให้เห็นว่าโดยปกติแล้วเวกเตอร์ $\hat{\beta}$ มีความยาวถัวเฉลี่ยที่โน้มเอียงไปในทางที่จะมีค่าสูงกว่าเวกเตอร์ β ซึ่งเป็นเวกเตอร์ของค่าจริงของพารามิเตอร์อยู่แล้ว

อย่างไรก็ตาม ในกรณีของ Orthogonality ซึ่งเป็นกรณีที่ตัวแปรอิสระเป็นอิสระเชิงเส้นต่อกันนั้น จะพบว่า

$$E(L_1^2) = n\sigma^2$$

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + n\sigma^2$$

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

แต่เราสามารถคำนวณหาเมตริกซ์ P ที่มีผลให้ $P'(X'X)\ P = \Lambda = diag\ (\lambda_1, \lambda_2, ..., \lambda_p)$ โดยที่ λ_i คือ Eigen value ของเมตริกซ์ $X'X$ ที่ $\lambda_i > 0$ ได้เสมอ

โดยอาศัยความรู้ดังกล่าวนี้ทำให้สมการที่ 6 และ 9 เปลี่ยนรูปไปเป็นฟังก์ชันของ λ_i ได้ดังนี้

$$E(L_i^2) = \sigma^2\ trace\ diag\ (1/\lambda_1, 1/\lambda_2, ..., 1/\lambda_p) = \sigma^2 \sum_{i=1}^{p} 1/\lambda_i \quad ...(10)$$

$$Var(L_1^2) = 2\sigma^4\ trace\ diag\ (1/\lambda_1^2, 1/\lambda_2^2, ..., 1/\lambda_p^2 = 2\sigma^4 \sum_{i=1}^{p} 1/\lambda_i^2 \quad ...(11)$$

และจากความรู้เกี่ยวกับ Rank of matrix และ Eigen value problem เราทราบว่าถ้าเมตริกซ์ $X'X$ เป็น Singular matrix หรือ Near - singular matrix แล้ว λ_{min} จะมีค่าเท่ากับ 0 หรือ ใกล้ 0 ขณะที่ปัญหา Multicolinearity คือปัญหาที่ $X'X$ มีธรรมชาติของ Near - singularity หรือ Singularity ผลที่ติดตามมาก็คือ $\lambda_{min} \to 0$

จากสมการ (10) และ (11) ถ้า $\lambda_{min} \to 0$ หรือมี Eigen value หลายตัวที่มีค่าใกล้ 0 เราจะพบลักษณะของ Unboundedness ใน MSE หรือ $E(L_1^2)$ กล่าวคือ

$$E(L_1^2) \to \infty \quad \text{และ} \quad Var\ (L_1^2) \to \infty$$

หรืออาจกล่าวได้ว่า เมื่อใดก็ตามที่ $X'X$ เป็น Ill - conditioning design matrix หรือ Near-singularity แล้วจะมีผลให้

(1) $\hat{\beta}$ ประมาณค่า β ได้ไม่ดี เพราะเวกเตอร์ $\hat{\beta}$ ต่างจากเวกเตอร์ β มาก

(2) ความผันแปรของ $\hat{\beta}$ มีปริมาณสูงมากจนทำให้ไม่มั่นใจในความถูกต้องและน่าเชื่อถือของค่าประมาณดังกล่าว (หรือ MSE ขาดลักษณะของ Boudedness)

(3) ทำให้ฟังก์ชัน $W'\beta = w_1\beta_1 + w_2\beta_2 + \ldots w_p\beta_p$ เป็นตัวประมาณค่าที่เอนเอียง

อนึ่ง เมื่อพิจารณาสมการ (7) จะพบว่าโดยปกติ $\hat{\beta}$ จะต่างจากเวกเตอร์จริงอยู่แล้ว และเมื่อมีสถานการณ์ของ Ill-conditioning คือ $X'X$ มีธรรมชาติของ Near - singularity ประกอบเพิ่มเติมเข้ามาอีก จะพบว่า

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \sum_{i=1}^{p} 1/\lambda_i \to \infty \text{ ถ้า } \lambda_{min} \to 0 \qquad \ldots(12)$$

จากเหตุผลปัญหา และสิ่งที่เป็นไปได้ภายใต้สภาวะของ Ill-conditioning หรือ Near-Singularity ดังกล่าว ทำให้จำเป็นต้องคำนวณหาตัวประมาณค่าของพารามิเตอร์ β ที่เหมาะสมกว่า $\hat{\beta}$ ภายใต้สถานการณ์เดียวกัน โดยคาดว่าตัวประมาณค่าดังกล่าวมีค่าใกล้เคียงกับพารามิเตอร์ β

2.1.2 <u>Ridge Regression</u>

ให้ B คือ ตัวประมาณค่าใหม่ของเวกเตอร์ β นอกเหนือไปจาก $\hat{\beta}$ จะพบว่าเวกเตอร์ B ให้ค่า Residual Sum of Square

$$\Phi = (Y - XB)'(Y - XB)$$

$$= Y'Y - 2B'X'Y + B'X'X B \qquad \ldots\ldots\ldots\ldots.(13)$$

นำ $2\hat{\beta}'X'Y$ บวกเข้าและลบออกทางด้านขวาของเครื่องหมายเท่ากับในสมการ (13) จะพบว่า

$$\Phi = Y'Y - 2B'X'Y + B'X'XB - 2\hat{\beta}'X'Y + 2\hat{\beta}'X'Y \qquad \ldots\ldots\ldots(14)$$

$$= (Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'Y) + (B'X'X B - 2B'X'Y + \hat{\beta}'X'Y)\ldots(15)$$

แต่จาก (Normal Equation) ของ OLS เราทราบว่า $X'Y = (X'X)\hat{\beta}$ ดังนั้นสมการ (15) จึงเปลี่ยนรูปเป็น

$$\Phi = (Y'Y - 2\beta'X'Y + \hat{\beta}'X'X\hat{\beta}) + (B'X'XB - 2B'X'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta})$$

นั่นคือ $\Phi = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (B - \hat{\beta})'X'X(B - \hat{\beta})$ ....(16)

หรือ $\Phi = \Phi_{min} + \Phi(B)$ ....(17)

โดยที่ $\Phi_{min} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$ คือ Residual Sum of Square ที่มีค่าต่ำที่สุด และ $\Phi(B)$ คือ Quadratic form ของ $(B - \hat{\beta})$

ผลที่ได้รับจากสมการ (17) ทำให้ทราบว่าเวกเตอร์ B มิได้ให้ค่า Φ ต่ำไปกว่า $\hat{\beta}$ ได้ วิธีที่จะให้ได้ค่าประมาณของ β จึงควรเป็นวิธีที่มีผลให้เวกเตอร์ของค่าประมาณมีค่าใกล้เคียงกับพารามิเตอร์ β ภายใต้ข้อจำกัดว่า $\Phi(B)$ มีค่าคงที่ หรือนัยหนึ่ง ค่าประมาณของ β คือ B ควรคำนวณได้มาจากการ minimize $B'B$ ภายใต้ข้อจำกัดว่า $(B - \hat{\beta})'X'X(B - \hat{\beta})$ มีค่าคงที่เท่ากับ Φ_o

ดังนั้น โดยอาศัย Lagrangian Multiplier Method โดยกำหนดให้ 1/k คือ Lagrange Multiplier ; k > 0 จะพบว่า

$$F = B'B + (1/k)\left[(B - \hat{\beta})'X'X(B - \hat{\beta}) - \Phi_o\right] \quad .....(18)$$

$$\frac{\partial F}{\partial B} = 0 \qquad 2B + (1/k)\,2X'X(B - \hat{\beta}) = 0 \quad .....(19)$$

$$kB + X'XB - X'X\hat{\beta} = 0$$

$$(kI_p + X'X)B = X'X\hat{\beta} \quad .....(20)$$

แต่จาก OLS เราทราบว่า $\hat{\beta} = (X'X)^{-1}X'Y$ ดังนั้นสมการ (20) จึงเปลี่ยนรูปไปดังนี้

$$(X'X + kI)B = X'Y$$

นั่นคือ $B = \hat{\beta}^* = (X'X + kI)^{-1}X'Y$ .....(21)

$\hat{\beta}^*$ ในสมการ (21) คือ Ridge Estimator

ตัวประมาณค่า $\hat{\beta}^*$ หรือ B ในสมการ (21) เป็นตัวประมาณค่าที่มีคุณสมบัติหลายประการที่ดีกว่า OLS ซึ่งจะได้กล่าวถึงต่อไป และมีความเกี่ยวข้องกับ OLS ดังนี้

(1) $\hat{\beta}^*$ คือ ฟังก์ชันของ $\hat{\beta}$

<u>พิสูจน์</u>

จาก $\hat{\beta}^* = (X'X + kI)^{-1} X'Y = W X'Y$ จะพบว่า ....(22)

$$\hat{\beta}^* = (X'X + kI)^{-1} (X'X)(X'X)^{-1} X'Y$$

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

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

$$= Z\hat{\beta} \quad \text{เมื่อ} \quad Z = \left[I + k(X'X)^{-1}\right]^{-1} \qquad ......(23)$$

(2) Eigen value ของเมตริกซ์ W และ Z คือ $\S_i(W) = 1/(\lambda_i + k)$ และ $\S_i(Z) = 1/(1 + k/\lambda_i)$ ตามลำดับ เมื่อ λ_i คือ Eigen value ค่าที่ i ของเมตริกซ์ $X'X$

<u>พิสูจน์</u>

จากความรู้เรื่อง Characteristic value problems และ Congruence เราสามารถคำนวณหาเมตริกซ์ P ที่มีผลให้ $P'(X'X)P = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$

ดังนั้น $W = (X'X + kI)^{-1} \sim (\Lambda + kI)^{-1}$

$$\sim \text{diag}\left[1/(\lambda_1 + k)\ ,\ 1/(\lambda_2 + k), \ldots, 1/(\lambda_p + k)\right] \quad \ldots\ldots\ldots\ldots (24)$$

$$= \text{diag}\left[\ \S_1(W),\ \S_2(W), \ldots, \S_p(W)\right] \ldots (25)$$

และ $Z = \left[I + k(X'X)^{-1}\right]^{-1} \sim \left[I + k\Lambda^{-1}\right]^{-1}$

$$\sim \text{diag}\left[1/(1+k/\lambda_1), 1/(1+k/\lambda_2), \ldots, 1/1+k/\lambda_p)\right]$$

$$= \text{diag}\left[\lambda_1/(\lambda_1+k),\ \lambda_2/(\lambda_2+k), \ldots, \lambda_p/(\lambda_p+k)\ \right] \quad \ldots\ldots\ldots\ldots (26)$$

$$= \text{diag}\left[\S_1(Z),\ \S_2(Z), \ldots, \S_p(Z)\right] \ldots (27)$$

(3) $Z = \left[I+k(X'X)^{-1}\right]^{-1}$ สามารถเสนอได้อีกรูปหนึ่งดังนี้คือ

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

<u>พิสูจน์</u>

จาก $W = (X'X + kI)^{-1}$ จะพบว่า

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

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

นั่นคือ $(X'X + kI)^{-1} = Z(\ X'X)^{-1}$

จาก (28) แสดงว่า $(X'X + kI)\ Z = X'X$

หรือ $X'X\ Z + kZ = X'X$

นั่นคือ $X'XZ = X'X - kz$

ดังนั้น $Z = (X'X)^{-1}\ (X'X - kZ)$

$$= I - k(X'X)^{-1}\ Z \quad \ldots\ldots\ldots\ldots (29)$$

แต่จาก (23) เราทราบว่า $Z = \left[I + k(X'X)^{-1}\right]^{-1}$ เมื่อแทนค่า Z ลงในสมการ (29) จึงมีผลให้สมการ (29) เปลี่ยนรูปไปสู่รูปใหม่ดังนี้

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

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

$$= I - k(X'X + kI)^{-1} = I - kW \qquad \text{..........(30)}$$

(4) เวกเตอร์ $\hat{\beta}^*$ จะมีความยาวต่ำกว่าเวกเตอร์ $\hat{\beta}$ ในทุกค่าของ k

<u>พิสูจน์</u> จาก $\hat{\beta}^* = Z\hat{\beta}$ จะพบเวกเตอร์ $\hat{\beta}^*$ มีความยาวดังนี้

$$L(\hat{\beta}^*) = \hat{\beta}^{*\prime}\hat{\beta}^* = \hat{\beta}'Z'Z\hat{\beta}^* \qquad \text{...........(31)}$$

แต่จากความรู้เรื่อง Orthogonalization สามารถคำนวณหาเมตริกซ์ P ที่มีผลให้

$$P'ZP = \Lambda = \text{diag}\left[\S_1(Z), \S_2(Z), \ldots, \S_p(Z)\right]$$

$$Z'Z = \text{diag}\left[\S_1^2(Z), \S_2^2(Z), \ldots \S_p^2(Z)\right] = \Lambda'\Lambda$$

ดังนั้น $\hat{\beta}^{*\prime}\hat{\beta}^* = \S_1^2(Z)\hat{\beta}_1^2 + \S_2^2(Z)\hat{\beta}_2^2 + \ldots + \S_p^2(Z)\hat{\beta}_p^2$ ....(32)

จาก (32) ถ้าแทน $\S_1^2(Z)$ ด้วย $\S_{max}^2(Z)$ เมื่อ $\S_{max}(Z) = \S_{(1)}(Z) > \S_{(2)}(Z) > \ldots > \S_{(p)}(Z) = \S_{min}(Z)$, $i = 1,2,\ldots,p$ เมื่อ $\S_{max} = \lambda_1/\lambda_1 + k$ สมการ (32) จะเปลี่ยนรูปเป็นอสมการดังนี้

$$\hat{\beta}^{*\prime}\hat{\beta}^* \leq \S_{max.}^2(Z)\,\hat{\beta}'\hat{\beta} \qquad \text{............(33)}$$

แต่ $\S_{max.} = \lambda_{(1)}/(\lambda_{(1)} + k)$ เมื่อ $\lambda_{(1)} = \max(\lambda_1, \lambda_2, \ldots, \lambda_p)$ จะพบว่า $\S_{max.}^2$ มีค่าต่ำกว่า $\S_i^2(Z)$; $i = 1,2,\ldots p$ ซึ่งแสดงว่าแม้แต่ $\S_{max.}^2\,\hat{\beta}'\hat{\beta}$ ยังมีค่าไม่น้อยกว่า $\hat{\beta}^{*\prime}\hat{\beta}^*$ หากเปลี่ยนเป็น $\S_{min}^2(Z)$ หรือ $\S_{(i)}^2(Z)$ ใด ๆ $\S_{min.}^2(Z)\,\hat{\beta}'\hat{\beta}$ และ $\S_{(i)}^2(Z)\hat{\beta}^{*\prime}\hat{\beta}^*$ ย่อมต้องมีค่าสูงกว่า $\hat{\beta}^{*\prime}\hat{\beta}^*$

นั่นคือ $\hat{\beta}^{*\prime}\hat{\beta}^* < \hat{\beta}'\hat{\beta}$ ....(34)

หรือเวกเตอร์ $\hat{\beta}^*$ จะเป็นเวกเตอร์ที่มีระยะความยาวต่ำกว่า $\hat{\beta}$ ซึ่งเป็น OLS-Estimator ของเวกเตอร์ β เสมอ

2.1.3 MSE ของ Ridge Regression

พิจารณา $L_1^2(k) = (\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta)$ จะพบว่า Expected Square Length ของ Error vector $(\hat{\beta}^* - \beta)$ ก็คือ MSE ของ $\hat{\beta}^*$ ซึ่งแสดงให้เห็นได้ดังนี้

$$E\left[L_1^2(k)\right] = E\left[(\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta)\right] \quad ....(35)$$

แต่จาก (23) $\hat{\beta}^* = \left[I_p + k(X'X)^{-1}\right]^{-1}\hat{\beta}$

$= Z\hat{\beta}$

ดังนั้น $E\left[L_1^2(k)\right] = E\left[(Z\hat{\beta} - \beta)'(Z\hat{\beta} - \beta)\right]$

$= E\left[\hat{\beta}'Z'Z\hat{\beta} - 2\beta'Z\hat{\beta} + \beta'\beta\right]$

$= E\left[(\hat{\beta}'Z'Z\hat{\beta})\right] - 2\beta'Z\beta + \beta'\beta$ ......(36)

เมื่อนำ $2\beta'Z'Z\beta$ บวกเข้าและลบออกจาก (36) จะพบว่า

$$E\left[L_i^2(k)\right] = E(\hat{\beta}'Z'Z\hat{\beta}) - 2\beta'Z'Z\beta + 2\beta'Z'Z\beta - 2\beta'Z\beta + \beta'\beta$$

แต่ $2\beta'Z'Z\beta = \beta'Z'Z\beta + \beta'Z'Z\beta$ และจาก (2) $E(\hat{\beta}) = \beta$ จึงทำให้ได้ $E\left[L_1^2(k)\right]$ ดังนี้

$$E\left[L_1^2(k)\right] = E(\hat{\beta}'Z'Z\hat{\beta}) - 2E(\beta'Z'Z\hat{\beta}) + \beta'Z'Z\beta + \beta'Z'Z\beta - 2\beta'Z\beta + \beta'\beta$$

$$= E\left[\hat{\beta}'Z'Z\hat{\beta} - \hat{\beta}'Z'Z\beta - \beta'Z'Z\hat{\beta} + \beta'Z'Z\beta\right] + (\beta'Z'Z\beta - \beta'Z'\beta$$

$$- \beta'Z\beta + \beta'\beta)$$

$$= E\left[(\hat{\beta} - \beta)'Z'Z(\hat{\beta} - \beta)\right] + (Z\beta - \beta)'(Z\beta - \beta) \quad .....(37)$$

แต่ $(\hat{\beta} - \beta) = \left[\beta + (X'X)^{-1}X'U\right] - \beta = (X'X)^{-1}X'U$ ดังนั้น

$$E\left[L_1^2(k)\right] = E\left[U'\{X(X'X)^{-1} Z'Z(X'X)^{-1}\} U\right] + (Z\beta - \beta)'(Z\beta - \beta)$$

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

$$= \sigma^2 \text{ trace } (X'X)^{-1} Z'Z + \beta'(Z - I)'(Z - I)\beta$$

$$= a + b \qquad \text{..............} \quad (38)$$

จาก (23) และ (29) $Z = \left[I + k(X'X)^{-1}\right]^{-1} = I - k(X'X + kI)^{-1}$ และ

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

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

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

$$= \sigma^2 \text{ trace } \left[\{I + k(X'X)^{-1}\}^{-1} (X'X)^{-1} \{I + k(X'X)^{-1}\}^{-1}\right]$$

$$= \sigma^2 \text{ trace } \left[(X'X + kI)^{-1} \{I + k(X'X)^{-1}\}^{-1}\right]$$

$$= \sigma^2 \text{ trace } \left[(X'X + kI)^{-1} \{I - k(X'X + kI)^{-1}\}\right]$$

$$= \sigma^2 \text{ trace } \left[(X'X + kI)^{-1} \{I - k(X'X + kI)^{-1}\}\right]$$

$$= \sigma^2 \text{ trace } \left[(X'X + kI)^{-1} - k(X'X + kI)^{-2}\right]$$

และเนื่องจากสามารถคำนวณหาเมตริกซ์ P ที่ $P'(X'X)P = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$

เมื่อ λ_i คือ Eigen value ที่ i ของเมตริกซ์ X ; $i = 1,2,\ldots, p$

ดังนั้น
$$a = \sigma^2 \text{ trace } \left[(\Lambda + kI)^{-1} - k(\Lambda + kI)^{-2}\right]$$

$$= \sigma^2 \left[\sum_{i=1}^{p} 1/(\lambda_i + k) - \sum_{i=1}^{p} k/(\lambda_i + k)^2\right]$$

$$= \sigma^2 \sum_{i=1}^{p} \lambda_1/(\lambda_i + k)^2$$

พิจารณา
$$b = \beta'(Z - I)'(Z - I)\beta \qquad \text{..........} (39)$$

และจาก (30) จะพบว่า $Z = I - k(X'X + kI)^{-1}$

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

$$= \beta'[k^2(X'X + kI)^{-2}]\beta$$

$$= k^2\beta'(X'X + kI)^{-2}\beta \qquad \ldots\ldots\ldots\ldots\ldots (40)$$

$$\therefore \text{MSE} = E[L_1^2(k)] = \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+k)^2 + [k^2\beta'(X'X + kI)^{-2}\beta] \ldots (41)$$

$$= \gamma_1(k) + \gamma_2(k)$$

หรือ MSE $= \gamma_1(k) + \gamma_2(k)$ โดยที่ $\gamma_1(k) = \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+k)^2$ = Total variance ขณะที่ $\gamma_2(k) = k^2\beta'(X'X + kI)^{-2}\beta = (Z\beta - \beta)'(Z\beta - \beta)$

= Square length ของระยะห่างระหว่าง $Z\beta$ กับ β

อนึ่งสำหรับ $\gamma_2(k)$ เราทราบว่า $Z = I - k(X'X + kI)^{-1}$ ซึ่งถ้าหาก $k = 0$ จะพบว่า $Z = I$ หรือ $Z\beta = \beta$ แต่ถ้า $k > 0$ จะพบว่า $Z \neq I$ หรือ $Z\beta \neq \beta$ แสดงว่า $\gamma_2(k)$ ก็คือ $(\text{bias})^2$ นั่นเอง

2.1.4 Mean, Variance และ SSE

(1) จาก $$\hat{\beta}^* = (X'X + kI_p)^{-1}X'Y$$

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

$$= [I + k(X'X)^{-1}]\hat{\beta}$$ เมื่อ $\hat{\beta}$ คือ OLS - Extimator

$$= Z\hat{\beta}$$

แต่ $E(\hat{\beta}) = \beta$

ดังนั้น $E(\hat{\beta}^*) \neq \beta \qquad \ldots\ldots\ldots (42)$

แสดงว่า $\hat{\beta}^* = Z\hat{\beta} = [I - k(X'X + kI)^{-1}]\hat{\beta}$ เป็น Biased Estimator

(2) Variance ของ $\hat{\beta}^*$

$$\text{จาก} \quad \hat{\beta}^* = (X'X + kI)^{-1} X'Y$$

$$= WX'Y$$

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

แสดงว่า $\hat{\beta}^*$ เป็น Linear function ของ Y กล่าวคือ $\hat{\beta}^* = \left[Z(X'X)^{-1} X'\right] Y$

$$\text{ดังนั้น} \quad Var(\hat{\beta}^*) = \left[Z(X'X)^{-1} X'\right] Var(Y) \left[Z(X'X)^{-1} X'\right]'$$

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

$$= \sigma^2 \left[Z(X'X)^{-1} Z'\right] \qquad \text{..........(43)}$$

$$\text{(3)} \quad SSE(\hat{\beta}^*) = (Y - X\hat{\beta}^*)'(Y - X\hat{\beta}^*)$$

$$= Y'Y - Y'X\hat{\beta}^* - \hat{\beta}^{*\prime} X'Y + \hat{\beta}^{*\prime} X'X\hat{\beta}^*$$

$$= Y'Y - \hat{\beta}^{*\prime} X'Y - \left[Y'X\hat{\beta}^* - \hat{\beta}^{*\prime} X'X\hat{\beta}^*\right]$$

$$= Y'Y - Y'X\hat{\beta}^* - \left[Y'X - \hat{\beta}^{*\prime} X'X\right] \hat{\beta}^*$$

$$= Y'Y - Y'X\hat{\beta}^* - \left[X'Y - X'X\hat{\beta}^*\right]' \hat{\beta}^*$$

แต่จาก (20) เราทราบว่าสมการปกติ คือ $(X'X + kI)\,\hat{\beta}^* = X'Y$ ทำให้ทราบได้ว่า

$X'Y - X'X\hat{\beta}^* = k\hat{\beta}^*$

$$SSE(\hat{\beta}^*) = Y'Y - Y'X\hat{\beta}^* - k\hat{\beta}^{*\prime}\hat{\beta}^*$$

$$= (SST - SSR^*) - k\left[\text{Square Length of } \hat{\beta}^*\right] \quad \text{....(46)}$$

โดยที่ SSR = Sum of Square Regression

SSR^* = Sum of Square Regression ของ $\hat{\beta}^*$

2.1.5 คุณสมบัติของ MSE

จาก (41) $MSE(\hat{\beta}^{*}) = \gamma_1(k) + \gamma_2(k)$

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

จะมีคุณสมบัติที่สำคัญดังนี้

1) $\gamma_1(k)$ เป็น Continuous monotonic decressing function ของตัวคงที่ k

<u>พิสูจน์</u> $\gamma_1(k) = \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^2$

$$\gamma_1'(k) = \frac{\partial}{\partial k}\gamma_1(k) = -2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^3 \quad \ldots\ldots(47)$$

$$\lim_{k \to k'} \left[f(k) - f(k')\right] = \lim_{k \to k'} \left[\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^2 - \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k')^2\right]$$

$$= 0$$

ดังนั้น $\gamma_1(k)$ เป็น Continuous function ของ k และ $\gamma_1'(k) = -\,ve$ ซึ่งแสดงว่า $\gamma_1(k)$ เป็น Decressing function ของ k

เมื่อ $\gamma_1'(k) \to -\infty$ เมื่อ $k \to 0^+$ และ $\lambda_{(p)} \to 0$ เมื่อ $\lambda_{(p)} = \lambda_{min}$

<u>พิสูจน์</u>

$$\gamma_1'(k) = -2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k_p)^3$$

ดังนั้น $\lim_{k \to 0^+} \gamma_1'(k) = \lim_{k \to 0} + \left[-2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^3\right]$

$$= -2\sigma^2 \sum_{i=1}^{p} 1/\lambda_i^2$$

$$= -2\sigma^2\left[1/\lambda_1^2 + 1/\lambda_2^2 + \ldots + 1/\lambda_p^2\right] \ldots\ldots\ldots(48)$$

แต่จาก (48) จะพบว่า $\lim_{k\to 0} + \gamma_1'(k)$ มีค่าผันแปรไปตาม $\lambda_{(p)}$ กล่าวคือ ถ้า $\lambda_{(p)} \to 0$ จะมีผลให้ $1/\lambda_{(p)} \to \infty$

2) Square bias คือ $\gamma_2(k)$ เป็น Monotonic increasing function ของตัวคงที่ k

<u>พิสูจน์</u> จากสมการ $Y = X\beta + U$ เราสามารถคำนวณหา Orthogonal matrix P ในลักษณะที่ $P'(X'X)P = \Lambda = \text{diag.}(\lambda_1, \lambda_2, \ldots, {}_p)$ เมื่อ λ_i คือ Eigen value ค่าที่ i ของเมตริกซ์ $X'X$

ดังนั้นสมการถดถอย $Y = X\beta + U$ จึงเปลี่ยนรูปใหม่ได้ดังนี้

$$Y = XPP'\beta + U$$

หรือ $Y = X^*\alpha + U$ โดยที่ $X^* = XP$ และ $\alpha = P'\beta$

และพบว่า Normal equation สำหรับ Ridge Regression

$$(X^{*\prime}X^* + kI)\hat{\alpha} = X^{*\prime}Y \qquad \ldots\ldots\ldots\ldots(49)$$

จากสมการ (49) จะพบว่า

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

$$X^{*\prime}X^* = P'(X'X)P = \Lambda$$

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

$$= \gamma_1(k) + \gamma_2(k) \qquad \ldots\ldots\ldots\ldots(50)$$

จากสมการ (50) จะพบว่า

$$\gamma_2(k) = k^2\alpha'(\Lambda + kI)^{-2}\alpha$$

$$= k^2\alpha' D^{-1}\alpha$$

โดยที่ $D^{-2} = D^{-1}D^{-1}$ $\quad D^{-1} = \text{diag.}\left[1/(\lambda_1 + k), 1/(\lambda_2 + k), \ldots, 1/\lambda_p + k)\right]$

และจะเห็นได้ว่า $D^{-2} = \text{diag.}\left[1/(\lambda_1 + k)^2, 1/(\lambda_2 + k)^2, \ldots, 1/(\lambda_p + k)^2\right]$

$$\gamma_2(k) = k^2 \sum_{i=1}^{p} \alpha_i^2 \;/(\lambda_i + k)^2$$

ดังนั้น $$\gamma_2(k) = \sum_{i=1}^{p} \alpha_i^2 / \left[1 + (\lambda_i/k)\right]^2 \qquad \ldots\ldots\ldots\ldots(51)$$

จากสมการ (51) จะพบว่า $$\lim_{k\to k}, \left[k^2 \sum_{i=1}^{p} \alpha_i^2 /(\lambda_i+k)^2 - k^2 \sum_{i=1}^{p} \alpha_i^2/(\lambda_i+k')^2\right] = 0 \ldots(52)$$

$$\gamma_2'(k) = \frac{\partial}{\partial k}\,\gamma_2(k) = \frac{\partial}{\partial k}\left[\sum_{i=1}^{p} \alpha_i^2/\left[1+(\lambda_i/k)\right]^2\right.$$

$$= -2 \sum_{i=1}^{p} \alpha_i^2 /\left[1+(\lambda_i/k)\right]^3 (-\lambda_i/k^2)$$

$$= 2k \sum_{i=1}^{p} \lambda_i\, \alpha_i^2/(\lambda_i + k)^3$$

$$= +\,ve \qquad \ldots\ldots\ldots\ldots(53)$$

จากสมการ (52) และ (53) ทำให้สรุปได้ว่า $\gamma_2(k)$ เป็น Continuous monotonic increasing function ของตัวคงที่ k เมื่อ $k > 0$ และ $\lambda_i > 0$

ขีดจำกัดบนของ $\gamma_2(k)$ คือ $\beta'\beta$

<u>พิสูจน์</u> $$\gamma_2(k) = \sum_{i=1}^{p} \alpha_i^2/\left[1 + (\lambda_i/k)\right]^2$$

$$\lim_{k\to\infty} \gamma_2(k) = \lim_{k\to\infty} \sum_{i=1}^{p} \alpha_i^2/\left[1+(\lambda_i/k)\right]^2 = \sum_{i=1}^{p} \alpha_i^2$$

$$= \alpha'\alpha$$

$$= (p'\beta)'(P'\beta)$$

$$= \beta' pp'\beta$$

$$= \beta'\beta$$

แสดงว่าขีดจำกัดบนของ $\gamma_2(k)$ คือ $\beta'\beta$ หรือ $\sum_{i=1}^{p} \beta_i^2$ $\qquad \ldots\ldots\ldots(54)$

ถ้า $k \to 0^+$ จะมีผลให้ $\gamma_2'(k) \to 0$

<u>พิสูจน์</u> $$\gamma_2'(k) = 2k \sum_{i=1}^{p} \lambda_i \alpha_i^2/(\lambda_i+k)^3$$

$$\lim_{k\to 0^+} \gamma_2'(k) = \sum_{i=1}^{p} \lim_{k\to 0^+} \left[2k\lambda_i \alpha_i^2/(\lambda_i+k)^3\right] = 0$$

3) สามารถหาค่าคงที่ k ที่มีผลให้ $E\left[L_1^2(k)\right] < E\left[L_1^2(0)\right]$ ได้เสมอ

<u>พิสูจน์</u> จาก $$MSE = E\left[L_1^2(k)\right] = \gamma_1(k)+\gamma_2(k)$$

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

$$= \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+k)^2 + k^2 \sum_{i=1}^{p} \alpha_i^2/(\lambda_i+k)^2 \quad ....(56)$$

และ $$\frac{\partial}{\partial k}E\left[L_1^2(k)\right] = -2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+k)^3 + 2k \sum_{i=1}^{p} \lambda_i\alpha_i^2/(\lambda_i+k)^3 \quad ...(57)$$

จากสมการ (56) จะพบว่า $$E\left[L_i^2(0)\right] = \sigma^2 \sum_{i=1}^{p} 1/\lambda_i \quad ....(58)$$

จากสมการ (57) เราทราบจาก 1) ว่า $\gamma_1'(k)$ เป็น Monotonic decreasing function ของ k และจาก 2) เราทราบว่า $\gamma_2'(k)$ เป็น Monotonic increasing function ของ k (หรือ $\gamma_1'(k)$ เป็นลบเสมอ ขณะที่ $\gamma_2'(k)$ เป็นบวกเสมอ) ซึ่งจากสมการ (57) และ (58) ทำให้

$$\frac{\partial}{\partial k}E\left[L_1^2(k)\right] < \frac{\partial}{\partial k}E\left[L_1^2(0)\right] = 0$$

หรือ $$-2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+k)^3 + 2k \sum_{i=1}^{p} \lambda_i\alpha_i^2/(\lambda_i+k)^3 < 0 \quad ....(59)$$

หรือเรามีความจำเป็นต้องพิสูจน์ให้เห็นว่าเราสามารถหาค่าคงที่ $k \; ; \; k < 0$ ที่มีผลให้ $E\left[L_1^2(k)\right]$ เป็น Monotonic decreasing function

จากสมการ (59) จะพบว่า

$$2k \sum_{i=1}^{p} \alpha_i^2 \lambda_i/(\lambda_i + k)^3 < 2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^3 \qquad ....(60)$$

พิจารณา Order value ของ α_i^2 ; i = 1,2,..., P คือ

$$\alpha_{max} = \alpha_{(1)} > \alpha_{(2)} > \alpha_{(3)} > ...>\alpha_{(p)} = \alpha_{min}$$

จะพบว่า เมื่อแทน $\alpha_{(i)}$ ด้วย α_{min} และถ้า $\alpha_{min} \rightarrow 0$ จะทำให้ k ไม่เป็น Finite value วิธีที่จะให้ k เป็น Finite value (หรือ Bounded value) คือใช้ α_{max} เป็นตัวแทนของ $\alpha_{(i)}$; i = 1,2,..., P

ดังนั้นจากสมการ (60) เมื่อแทนที่ $\alpha_{(i)}$ ด้วย α_{max} จะทำให้ได้ค่าสูงสุดของ k ที่มีผลให้ $E\left[L_1^2(k)\right] < E\left[L_1^2(0)\right]$

$$k < \sigma^2/\alpha_{max}^2 \qquad ....(61)$$

นั่นคือ ในกรณีของ Ridge Regression นั้น สามารถคำนวณหาค่าคงที่ k ที่มีผลให้ Ridge estimator มี MSE ต่ำกว่า OLS-estimator ได้เสมอ และค่าคงที่ k ดังกล่าวจะเป็นค่าบวกที่มีค่าไม่เกินไปกว่า σ^2/α_{max}^2

2.1.6 <u>Generalized Ridge Regression</u>

พิจารณาสมการถดถอย $Y = X\beta + U$ จะพบว่าเราสามารถคำนวณหา เมตริกซ์ P ที่มีผลให้ $P'(X'X)P = \Lambda = \text{diag}(\lambda_i, \lambda_2, ..., \lambda_p)$ ได้เสมอ

ดังนั้นสมการถดถอย $Y = X\beta + U$ สามารถแปลงรูปไปสู่รูปใหม่ได้ดังนี้คือ

$$Y = XPP'\beta + U$$

$$= X^*\alpha + U \text{ เมื่อ } X^* = XP \text{ และ } \alpha = P'\beta \qquad ....(62)$$

จาก (62) จะเห็นได้ว่า

$$X^{*'}X^* = P'X'XP = \Lambda \text{ , } \alpha'\alpha = \beta'PP'\beta = \beta'\beta$$

โดยอาศัยหลักเกณฑ์ทำนองเดียวกันกับตอน 2.2.2 จะพบว่า

$$\hat{\alpha} = [X^{*\prime} X^{*} + K]^{-1} X^{*\prime} Y$$

โดยที่ $K = (\delta_{ij} k_{ij})_{pxp}$ เมื่อ δ_{ij} = kronecker delta กล่าวคือ

$\delta_{ij} = 0 ; i \neq j$, $\delta_{ij} = 1 ; i = j$ หรือนัยหนึ่ง $K = \text{diag.}(k_1 , k_2, \ldots, k_p)$

และพบว่าค่าที่พอเหมาะ (optimal) ของ k_i คือ

$$k_i = \sigma^2/\alpha_i^2 ; i = 1,2,\ldots, P$$

และ $$\hat{k}_i = \hat{\sigma}^2/\hat{\alpha}_i^2 ; i = 1,2,\ldots, P \qquad \ldots\ldots\ldots(63)$$

จากการศึกษาพบว่า ค่าประมาณที่ดีที่สุดของ $\hat{\beta}^*$ คือใช้ $k_i = k ; i = 1,2,\ldots, P$ หรือ $K = kI_p$

2.1.7 การประมาณค่าคงที่

จากการศึกษาของวิชเชอร์นและเชอร์ชีล (Wichern, D.W. and Churchill, G.A., 1978), กิบบอน (Gibbons, D.G., 1981) แมคโดนาล์ดและกาลานู (Mcdonald, G.C. and Galarneau, D.I., 1975) พบว่า ในบรรดาวิธีประมาณค่า k กว่า 10 วิธีนั้น วิธีดีที่สุดและเป็นที่นิยมใช้กันทั่วไปมีอยู่ 2 วิธีคือ

1. วิธี Hoerl, Kennard and Baldwin
2. วิธี Lawless and Wang

ซึ่งการศึกษาครั้งนี้จะเปรียบเทียบโดยใช้อัลกอริทึมของการประมาณค่า k สำหรับ Ridge Estimator ตามวิธีทั้ง 2 ประการนี้เท่านั้น วิธีประมาณค่า k สำหรับแต่ละวิธีดังกล่าวมีหลักการโดยสรุปดังนี้

1. วิธี Hoerl, Kennard and Baldwin (HKB)

วิธีนี้เสนอโดยโฮเอร์ล เคนนาร์ด และบอล์ดวิน (Hoerl, A.E., Kennard, R.W. and Baldwin, K.F., 1975) ซึ่งแนะนำให้เลือกค่า k ดังนี้

$$k = PS^2/\hat{\beta}'\hat{\beta}$$

โดยที่ $S^2 = \hat{\sigma}^2 = \frac{1}{n-P} (Y - X\hat{\beta})' (Y - X\hat{\beta})$ P = จำนวนตัวแปรอิสระ

$$\hat{\beta} = (X'X)^{-1} X'Y \quad \text{.............(64)}$$

สำหรับวิธี HKB นี้เหมาะสำหรับการวิเคราะห์ความถดถอยเมื่อ $P \geqslant 3$ เพราะถ้า $P \nless 2$ แล้ว HKB จะไม่ดีไปกว่า OLS

สังเกตว่า $k = PS^2/\hat{\beta}'\hat{\beta}$ คือ Harmonic mean ของ k_i ; $i = 1,2,\ldots, P$ ในตอน 2.1.6 นั่นเอง

2. วิธี Lawless and Wang (LW)

พิจารณาสมการถดถอย $Y = X\beta + U$

ให้ P คือ Orthogonal modal matrix ที่ $P'P = PP' = I$

$P'X'XP = \Lambda = \text{diag.}\ (\lambda_1, \lambda_2, \ldots, \lambda_p)$

ดังนั้น เมื่อจัดสมการ $Y = X\beta + U$ ให้อยู่ในรูปของ Canonical form

$$Y = XPP'\beta + U$$

$$= X^*\alpha + U$$

เมื่อ $X^* = XP$ และ $\alpha = P'\beta$

ดังนั้น OLS-estimator ของ α คือ

$$\hat{\alpha} = (X^{*\prime}X^*)^{-1} X^{*\prime}Y$$

$$= (P'X'XP)^{-1} P'X'Y$$

$$= \Lambda^{-1} X^{*\prime}Y$$

ลอว์เลสและแวง (Lawless and Wang, 1976) ใช้วิธีการของเบย์แล้วใช้สัดส่วนของความแปรปรวน คือ $PS^2/\hat{\alpha}'\Lambda\hat{\alpha}$ เป็นตัวประมาณค่าของ k

ผลจากการศึกษาเปรียบเทียบพบว่า ค่าประมาณของ β ตามวิธีของ LW คือ

$$k = PS^2/\hat{\alpha}'\Lambda\hat{\alpha}$$

โดยที่ P = จำนวนพารามิเตอร์ และ $S^2 = \frac{1}{n-p}\left[Y'Y - \hat{\beta}'X'Y\right]$ เป็นวิธีประมาณที่ดีที่สุดวิธีหนึ่งในจำนวนวิธีประมาณค่าหลายวิธี

2.2 การประมาณค่าสังเกตที่สูญหาย

2.2.1 เหตุผลและความสำคัญ

ในการวิเคราะห์ความถดถอยเชิงเส้น

$$Y = X\beta + U$$

ปัญหาที่พบอยู่เสมอก็คือ ปัญหาอันเนื่องมาจากค่าสังเกตสูญหายไปในบางช่วงของอนุกรมเวลา ทางออกสำหรับปัญหานี้ในทางปฏิบัติก็คือ ผู้วิเคราะห์ยังคงประมาณค่าสมการถดถอยโดยอาศัย Ordinary Least Square อยู่เช่นเดิมโดยตัดค่าตัวอย่างที่ไม่สมบูรณ์ทิ้งไป การกระทำดังกล่าวมีผลเสียที่เด่นชัด 2 ประการ ประการแรกค่าสังเกตขาดความต่อเนื่อง ประการที่สองค่าMean Square Error (MSE) มีค่าสูงกว่าความเป็นจริง เนื่องจากการสูญเสีย degree of freedom ด้วยเหตุผลดังกล่าวจึงได้มีการสร้างข้อมูลทดแทนขึ้นมาสำหรับข้อมูลที่สูญหายไป

2.2.2 วิธีการประมาณค่าสังเกตที่สูญหาย

จากวิธีการประมาณค่าสังเกตที่สูญหายซึ่งมีอยู่หลายวิธีนั้น วิธีที่ได้รับความนิยมนำไปใช้อยู่เสมอ มีอยู่ 2 วิธีคือ

1. วิธีแทนด้วยค่าเฉลี่ย (Mean Subatitution)
2. วิธีสมการถดถอย (Regression)

ซึ่งวิธีทั้ง 2 มีหลักการดังนี้

1. วิธีแทนด้วยค่าเฉลี่ย

เป็นวิธีที่ง่ายที่สุดโดยใช้ค่าเฉลี่ยของตัวแปรแต่ละตัวจากข้อมูลเท่าที่มีอยู่แล้ว แทนที่ข้อมูลที่สูญหาย กล่าวคือ

แทนที่ค่าของตัวแปร X_i ที่สูญหายไป m_i ค่าด้วย

$$X_{i.}^{(n_i)} = \frac{1}{n_i} \sum_{\ell}^{n_i} X_{i\ell}$$

โดยที่ n_i = จำนวนค่าสังเกตของตัวแปร X_i ที่ปรากฏ , $i = 1,2,\ldots, P$

m_i = จำนวนค่าสังเกตของตัวแปร X_i ที่สูญหาย คือ $m_i = n - n_i$

n = ขนาดตัวอย่างทั้งสิ้นของค่าสังเกต

$X_{i\ell}$ = คือ ค่าสังเกตที่ปรากฏของตัวแปรที่ i และตัวอย่างที่ ℓ

2. วิธีสมการถดถอย

เป็นวิธีที่อาศัยความสัมพันธ์ของสมการถดถอย มาใช้ในการประมาณค่าสังเกตที่สูญหายโดยมีแนวคิดพื้นฐาน คือ วิเคราะห์สมการถดถอยเชิงเส้น

$$X_i = f(X_j) \; ; \; i \neq j = 1,2,\ldots, P$$

แล้วใช้ค่าประมาณ $\hat{X}_{i\ell}$ จาก Regression Function $\hat{X}_{i\ell} = f(X_{j\ell})$; $i \neq j = 1,2,\ldots,P$ ประมาณค่า $X_{i\ell}$ ที่สูญหาย โดยใช้ข้อมูลที่สมบูรณ์เท่านั้นมาสร้างสมการถดถอย