บทที่ ๒

ทฤษฎีเกี่ยวกับริดจ์ รีเกรสชั่น

ในกรณีที่ตัวแปรอิสระในตัวแบบมีสภาพไม่เหมาะสม (ill-condition) ซึ่งในที่นี้จะหมายถึงตัวแปรอิสระมีค่าสหสัมพันธ์สูง การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยที่สุดจะไม่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด ดังจะเห็นได้จากการพิสูจน์ทฤษฎี ๒.๔.๒.๓ ต่อไป สำหรับการประมาณค่าสัมประสิทธิ์การถดถอยที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดนั้น

Hoerl และ Kennard ได้เสนอวิธีการที่ใช้การบวกค่าคงที่ที่มากกว่าศูนย์กับสมาชิกทุกตัวบนเส้นทะแยงมุมของเมทริกซ์ $X'X$ วิธีการดังกล่าวนี้จะได้ค่าประมาณของพารามิเตอร์มีลักษณะเอนเอียง และมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยที่สุด ซึ่งต่อมาในปี พ.ศ. ๒๕๑๘ Richard F. Gunst และ Robert L. Mason ได้นำวิธีการของ Hoerl และ Kennard ไปทดสอบกับตัวเลขจริง ซึ่งให้ผลเป็นจริงเหมือนกับที่กล่าวข้างต้น[๑]

รายละเอียดต่าง ๆ ในบทนี้จะกล่าวถึงการประมาณค่าพารามิเตอร์ β โดยวิธีริดจ์ รีเกรสชั่นพร้อมทั้งการพิสูจน์ทฤษฎีเหล่านี้ด้วย

---

[๑]Richard F. Gunst and Robert L. Mason, "Biased Estimation in Regression : An Evaluation Using Mean Square Error," Journal of the American Statistical Association, (September 1977, vol.72, No. 359), p. 666.

๒.๑ ตัวแบบทั่วไป

จากตัวแบบทั่วไปของการวิเคราะห์การถดถอยพหุเชิงเส้นซึ่งเป็นดังนี้

$$Y = X\beta + \varepsilon \qquad \text{..........(2.1.1)}$$

เมื่อ Y เป็นเวคเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมทริกซ์ของตัวแปรอิสระขนาด $n \times p$ และมี **full rank p**

β เป็นเวคเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด $p \times 1$

ε เป็นเวคเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$

โดยที่ $E(\varepsilon) = 0$

และ $E(\varepsilon'\varepsilon) = \sigma^2 I_n$

ถ้าประมาณค่าพารามิเตอร์ β ด้วยวิธีกำลังสองน้อยที่สุดจะได้ตัวประมาณ

$$\hat{\beta} = [X'X]^{-1}X'Y \qquad \text{..........(2.1.2)}$$

ในกรณีที่ไม่เกิดสภาพไม่เหมาะสมของตัวแปรอิสระ คือตัวแปรอิสระมีเมทริกซ์สหสัมพันธ์อยู่ในรูปเมทริกซ์เอกลักษณ์การคูณ (Identity matrix) การประมาณค่า β โดยวิธีกำลังสองน้อยที่สุดจะให้ตัวประมาณ $\hat{\beta}$ (ตามสมการ 2.1.2) ซึ่งมีลักษณะไม่เอนเอียง กล่าวคือ

$$\begin{aligned} E(\hat{\beta}) &= E[(X'X)^{-1} X'Y] \\ &= E[(X'X)^{-1}(X'X)\beta + (X'X)^{-1} X'\varepsilon] \\ &= E[\beta + (X'X)^{-1} X'\varepsilon] \\ &= E(\beta) + (X'X)^{-1} X'E(\varepsilon) \\ &= \beta \end{aligned}$$ [๑]

---

[๑]Franklin A. Graybill, Introduction to Linear Statistical Models, vol. I, p. 111.

แต่ถ้าตัวแปรอิสระมีค่าสหสัมพันธ์สูงกล่าวคือสมาชิกของเมทริกซ์สหสัมพันธ์ของตัวแปรอิสระที่ไม่อยู่บนเส้นทะแยงมุมมีค่าเข้าใกล้ ๑ (ไม่อยู่ในรูปเมทริกซ์เอกลักษณ์การคูณ) การประมาณค่า β ด้วยวิธีกำลังสองน้อยที่สุดจะมีผลทำให้เกิดข้อผิดพลาดของตัวประมาณ $\hat{\beta}$ แต่ในทางปฏิบัติสหสัมพันธ์ของตัวแปรอิสระที่มีลักษณะดังกล่าวข้างต้นนี้อาจหลีกเลี่ยงไม่ได้ในบางกรณี ดังนั้นจึงควรพิจารณาวิธีการประมาณค่า β ที่เหมาะสมกว่านี้ ซึ่งในที่นี้จึงควรจะพิจารณาวิธีริดจ์ รีเกรสชั่น แทนวิธีกำลังสองน้อยที่สุด

การคำนวณโดยวิธี ริดจ์ **รีเกรสชั่น** จะให้ตัวประมาณ $\hat{\beta}$ ดังนี้ ตัวประมาณค่าพารามิเตอร์ β

$$\hat{\beta}^* = (X'X + KI)^{-1} X'Y \; ; \quad K > 0 \;\ldots\ldots\ldots(2.1.3)$$[๑]

จากสมการ (2.1.3) ใช้หลักการคำนวณเหมือนกับสมการ (2.1.2) โดยที่ริดจ์ รีเกรสชั่น มีส่วนที่แตกต่างจากวิธีกำลังสองน้อยที่สุด คือ จะมี K เข้ามาปรับค่า $(X'X)^{-1}$ โดยนำค่า K บวกกับสมาชิกทุกตัวบนเส้นทะแยงมุมของเมทริกซ์ $X'X$ เพื่อให้ตัวประมาณ $\hat{\beta}^*$ มีคุณสมบัติที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดต่ำลง จากสมการ (2.1.3) กำหนดให้ $K > 0$ เราไม่อาจหาค่า K ที่แน่นอนได้นอกจากจะต้องมีการทดลองให้ K มีค่าเพิ่มจากศูนย์ทีละน้อยและแทนค่า K ในสมการ (2.1.3) จนกว่าจะได้ค่า K ที่เหมาะสม จะเห็นได้ว่าค่าตัวประมาณ $\hat{\beta}^*$ **จะได้จากการคำนวณหลายรอบ** (Iterative) เพราะขึ้นกับค่า K ที่เปลี่ยนแปลงไปเรื่อย ๆ ทุกครั้งที่กำหนดค่า K เพื่อคำนวณค่าตัวประมาณ $\hat{\beta}^*$ จะนำค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดขึ้นจากวิธีนี้และวิธีกำลังสองน้อยที่สุดมาเปรียบเทียบกันไปเรื่อย ๆ จนกว่าจะพบว่า ค่า K ใดที่ทำให้ค่าเฉลี่ยความคลาด-

---

[๑]Arthur E. Hoerl and Robert W. Kennard, "Rige Regression : Application to Nonorthogonal Problem", <u>Technometrics</u>, (Vol. 12, No. 1, February 1970), p. 70

เคลื่อนกำลังสองของตัวประมาณค่า β ด้วยวิธี ริดจ์ รีเกรสชั่น ให้ค่าน้อยกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่า β ด้วยวิธีกำลังสองน้อยที่สุด เราจะถือว่าค่า K นั้น เริ่มจะเป็นค่า K ที่เหมาะสม แล้วจะค่อย ๆ เปลี่ยนค่า K ไปทีละน้อยอีกจนกว่าจะหาค่า K ที่เหมาะสมที่สุด ซึ่งจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดได้ ดังนั้นอาจกล่าวได้ว่าจุดอ่อนของริดจ์ รีเกรสชั่น คือไม่สามารถกำหนดค่า K ที่แน่นอนได้

๒.๒ คุณสมบัติของตัวประมาณค่า $\hat{\beta}$ ด้วยวิธีกำลังสองน้อยที่สุด

จากวิธีกำลังสองน้อยที่สุด ซึ่งให้ตัวประมาณค่า β เป็น

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

โดยมีคุณสมบัติว่าตัวประมาณ $\hat{\beta}$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดในกลุ่มของตัวประมาณค่า β ที่ไม่เอนเอียง จากตัวแบบสามารถเขียนค่าความผิดพลาดในรูป X,Y ดังนี้

$$\varepsilon = Y - X\beta \qquad \text{.........(2.2.1)}$$

ดังนั้นจะได้ผลบวกของความคลาดเคลื่อนกำลังสอง มีค่าเท่ากับ $\varepsilon'\varepsilon$ ถ้าให้ $\phi(\hat{\beta})$ เป็นผลบวกความคลาดเคลื่อนกำลังสองเมื่อใช้ตัวประมาณ $\hat{\beta}$ จะได้

$$\phi(\hat{\beta}) = (Y - X\hat{\beta})' (Y - X\hat{\beta}) \qquad \text{.........(2.2.2)}$$

เนื่องจากข้อมูลตัวอย่างที่นำมาศึกษาความถดถอยพหุเชิงเส้นส่วนมากมีสหสัมพันธ์ของตัวแปรอิสระที่ไม่อยู่ในรูปเมทริกซ์เอกลักษณ์การคูณ ถ้าประมาณ β ด้วย $\hat{\beta}$ จากวิธีกำลังสองน้อยที่สุดอาจมีผลทำให้เมทริกซ์ของค่าแปรปรวนและค่าแปรปรวนร่วมของตัวประมาณ $\hat{\beta}$ ไม่มีขนาดเล็กที่สุด ดังนั้นเราควรพิจารณาคุณสมบัติของตัวประมาณค่า β ๒ ประการ คือ เมทริกซ์ค่าแปรปรวนและค่าแปรปรวนร่วมของตัวประมาณ $\hat{\beta}$ **และเฉลี่ย**

กำลังสองจาก $\hat{\beta}$ ไปยัง β ซึ่งสามารถเขียนอยู่ในรูปฟังค์ชั่นของ $X'X$ และ σ^2 ได้ดังต่อไปนี้

กำหนดให้ความแปรปรวนของค่าตัวประมาณ $\hat{\beta}$ มีค่าเป็น VAR $(\hat{\beta})$ ดังนั้น

$$\text{VAR}\ (\hat{\beta}) = \sigma^2\ (X'X)^{-1} \qquad \dots\dots\dots(2.2.3)$$

และให้ระยะทางจาก $\hat{\beta}$ ถึง β มีค่าเป็น L_1

ดังนั้น $$L_1^2 = (\hat{\beta} - \beta)'\ (\hat{\beta} - \beta) \qquad \dots\dots\dots(2.2.4)$$

ค่าเฉลี่ยของกำลังสองระยะทางจาก β ไปยัง $\hat{\beta}$ เป็นดังนี้

$$E[\ L_1^2\ ] = \sigma^2\ \text{Trace}^{๑}\ (X'X)^{-1} \qquad \dots\dots\dots(2.2.5)$$

$E[\ L_1^2\ ]$ จะมีค่าสมบูรณ์ (Equivalent) กับ $E[\ \hat{\beta}'\ \hat{\beta}\ ]$

$$E[\ \hat{\beta}'\ \hat{\beta}\ ] = \beta'\beta + \sigma^2\ \text{Trace}\ (X'X)^{-1}\ ^{๒} \qquad \dots\dots\dots(2.2.5\text{ก})$$

เมื่อ ε มีการกระจายแบบปกติจะได้

$$\text{VAR}\ (L_1^2) = 2\sigma^4\ \text{Trace}\ (X'X)^{-2}\ ^{๓} \qquad \dots\dots\dots(2.2.6)$$

จะเห็นได้ว่าจากสมการ (2.2.3), (2.2.5) และ (2.2.6) ค่าความแปรปรวนของตัวประมาณ $\hat{\beta}$ [$V(\hat{\beta})$] ค่าเฉลี่ยของกำลังสองระยะทางจาก β ไปยัง $\hat{\beta}$ [$E(L_1^2)$] และค่าความแปรปรวนของกำลังสองระยะทางจาก β ไปยัง $\hat{\beta}$ [$V(L_1^2)$] เหล่านี้

---

[๑] Trace = ผลบวกของสมาชิกทุกตัวบนเส้นทะแยงมุมของเมทริกซ์ $(X'X)^{-1}$

[๒] Arthur E. Hoerl and Robert W. Kennard, "Ridge Regression Biased Estimation for Nonothogonal Problems" Technometrics (Vol 12, No.1, February 1970), p. 56.

[๓] Ibid.

เป็นฟังค์ชั่นของ $\hat{\beta}$ เพื่อความสะดวกในการเปรียบเทียบการประมาณค่า β ระหว่างสองวิธีนี้ควรแปลงเมทริกซ์ $(X'X)^{-1}$ ให้อยู่ในรูปฟังค์ชั่นของค่า eigenvalue ของเมทริกซ์ $X'X$ โดยใช้คุณสมบัติที่สำคัญข้อหนึ่งของค่า eigenvalue ของ $X'X$ กล่าวคือถ้า λ_i เป็นค่า eigenvalue ของ $X'X$ เมื่อ $i = 1,2,\ldots, n$ n คือขนาดของเมทริกซ์ $X'X$ แล้ว $\sum_{i=1}^{n} \lambda_i = \text{Trace } (X'X)$

สมมติให้ค่า eigenvalue $X'X$ มีค่าเป็น

$$(\lambda_{max} = \lambda_1) > \lambda_2 > \lambda_3 \ldots\ldots(\lambda_p = \lambda_{min}) \geq 0 \qquad \ldots\ldots\ldots(2.2.7)$$

จากสมการ (2.2.5) ค่าเฉลี่ยของระยะทางจาก β ถึง $\hat{\beta}$ กำลังสอง อาจเขียนอยู่ในรูปฟังค์ชั่นของค่า eigenvalue ของ $X'X$ ดังนี้

$$E[\, L_1^2 \,] = \sigma^2 \sum_{i=1}^{p} (1/\lambda_i) \qquad \ldots\ldots\ldots(2.2.8)$$

และจากสมการ (2.2.6) ค่าความแปรปรวนของกำลังสองระยะทางจาก β ถึง $\hat{\beta}$ อาจเขียนในรูปฟังค์ชั่นของ eigenvalue ของ $X'X$ ดังนี้

$$VAR\ (L_1^2) = 2\sigma^4 \sum_{i=1}^{p} (1/\lambda_i)^2 \qquad \ldots\ldots\ldots(2.2.9)$$

จากสมการ (2.2.8) และสมการ (2.2.9) จะได้ขีดจำกัดล่าง (Lower Bound) ของค่าเฉลี่ยของกำลังสองของระยะทางจาก β ไปยัง $\hat{\beta}$ หรือ $E[\, L_1^2 \,]$ และค่าความแปรปรวนของกำลังสองของระยะทางจาก β ไปยัง $\hat{\beta}$ หรือ $VAR[\, L_1^2 \,]$ มีค่าเป็น σ^2/λ_{min} และ $2\sigma^4/\lambda_{min}^2$ ตามลำดับ

แต่ในกรณีที่ตัวแปรอิสระเกิดมีสภาพที่ไม่เหมาะสม ค่า eigenvalue ของเมทริกซ์ $X'X$ จะมีค่าต่ำมาก ๆ มีผลทำให้ระยะทางจาก β ไปยัง $\hat{\beta}$ มีค่ามาก[๙] พิจารณาค่า $E[L_1^2]$ และ

---

[๙]Arthur E. Hoerl and Robert W. Kennard, "Ridge Regression Biased Estimation for Nonorthogonal Problems" <u>Technometrics</u>, (Vol 12, No.1, February 1970), p. 56.

$VAR[L_1^2]$ จะเห็นว่าค่าทั้งสองสูงขึ้น สาเหตุที่เกิดเช่นนี้อาจมีผลมาจากการทำให้ค่าต่ำสุดของค่าผลบวกของความคลาดเคลื่อนกำลังสอง $\phi(\hat{\beta})$ ยังขาดวิธีการทางคณิตศาสตร์ที่จะลดค่า $E[L_1^2]$ ให้ต่ำลงได้นั่นเอง

๒.๓ การประมาณค่า β โดยวิธีริดจ์ รีเกรสชั่น (Ridge Regression)

A.E. Hoerl (2506) ได้ใช้วิธีการทางคณิตศาสตร์มาควบคุมคุณสมบัติของค่าความแปรปรวนของตัวประมาณค่า $\hat{\beta}$ ที่คำนวณโดยวิธีกำลังสองน้อยที่สุดให้มีค่าลดต่ำลงได้ดังนี้คือ กำหนดให้ค่าประมาณของ β ด้วยวิธี ริดจ์ รีเกรสชั่น และวิธีกำลังสองน้อยที่สุดมีค่าเท่ากับ $\hat{\beta}^*$ และ $\hat{\beta}$ ตามลำดับ จากสมการ (2.1.3) ค่า $\hat{\beta}^* = [X'X + KI]^{-1} X'Y$ โดยที่ $K > 0$ สามารถแปลง $\hat{\beta}^*$ ให้อยู่ในรูปฟังค์ชั่นใหม่ โดยกำหนดให้ $[X'X + KI]^{-1} = W$ ดังนี้

$$\hat{\beta}^* = W X'Y \qquad \text{..........(2.3.1)}$$

เนื่องจาก K ไม่ใช่ค่าคงที่จะต้องกำหนดค่า K หลายค่า ดังนั้นจะเห็นได้ว่า $\hat{\beta}^*$ มีค่าหลายชุดซึ่งอาจเขียนค่า $\hat{\beta}^*$ ให้อยู่ในรูป family ได้ดังนี้

$$\text{family } \hat{\beta}^* = \left\{ \begin{array}{l} \{\hat{\beta}_i^*\}/\{\hat{\beta}_i^*\} = W_i X'Y \ ; \ W = [X'X + KI]^{-1} \\ ; K > 0 \ , \ i = 1,2,\ldots, q \text{ เมื่อ } q \text{ หมายถึงจำนวนชุดของ } \hat{\beta} \end{array} \right\} \qquad \text{..........(2.3.2)}$$

จากตัวประมาณค่า $\hat{\beta}^*$ สามารถเขียนให้อยู่ในรูปความสัมพันธ์กับ $\hat{\beta}$ เพื่อสะดวกในการเปรียบเทียบวิธีการประมาณค่า β ระหว่างวิธีทั้งสองดังนี้

$$\hat{\beta}^* = [I + K(X'X)^{-1}]^{-1} \hat{\beta} \qquad \text{.......(2.3.3)}$$

$$= Z \hat{\beta} \qquad \text{.......(2.3.4)}$$

จากความสัมพันธ์ของสมการ (2.3.1) และ (2.3.4) จะได้

$Z = [I + K(X'X)^{-1}]^{-1}$ และ $W = [X'X + KI]^{-1}$ และค่า Z , $\hat{\beta}^*$

และ W เหล่านี้จะนำไปใช้พิจารณาเปรียบเทียบคุณสมบัติต่าง ๆ ของการประมาณค่าพารามิเตอร์ β ตลอดจนการพิสูจน์ทฤษฎีต่าง ๆ ที่สำคัญของการประมาณค่าสัมประสิทธิ์ความถดถอยของวิธีริดจ์ รีเกรสชั่น ต่อไป คุณสมบัติที่จำเป็นของ $\hat{\beta}^*$, Z, W มีดังนี้

๒.๓.๑ ให้ $E_i(W)$ และ $E_i(Z)$ เป็นค่า eigenvalue ของ W และ Z ตามลำดับซึ่งจะคำนวณจาก Characteristic Equation

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

และ

$$|Z - E_i I| = 0$$

$$E_i(W) = 1/(\lambda_i + K) \qquad \ldots\ldots\ldots(2.3.5)$$

$$E_i(Z) = \lambda_i/(\lambda_i + K) \qquad \ldots\ldots\ldots(2.3.6)$$

๒.๓.๒ ได้ว่า Z อาจแสดงในรูปฟังค์ชั่นของ KW ได้ดังนี้

$$Z = I - K(X'X + KI)^{-1}$$

$$= I - KW \qquad \ldots\ldots\ldots(2.3.7)$$ [๑]

๒.๓.๓ ค่าตัวประมาณ $\hat{\beta}^*$ จะมีค่าน้อยกว่า $\hat{\beta}$ เมื่อ $K > 0$

ตัวอย่างเช่น

$$(\hat{\beta}^*)'(\hat{\beta}^*) < \hat{\beta}'\hat{\beta} \qquad \ldots\ldots\ldots(2.3.8)$$

พิสูจน์จากนิยาม $\hat{\beta}^* = Z\hat{\beta}$ และ $X'X$, Z มีคุณสมบัติเป็น Symmetric positive definite ฉะนั้นจะได้

---

[๑]Arthur E. Hoerl and Robert W. Kennard, "Ridge Regression Biased Estimation for Nonorthogonal Problems" <u>Technometrics</u>, (Vol 12, No.1, February 1970), p. 57.

$$(\hat{\beta}^*)'(\hat{\beta}^*) = (\mathcal{Z}\,\hat{\beta})'(\mathcal{Z}\,\hat{\beta})$$

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

$$< E^2_{i(max)}(\mathcal{Z})\ \hat{\beta}'\hat{\beta} \qquad ....(2.3.9)$$

เมื่อ $E_{(max)}(\mathcal{Z}) = \lambda_1/(\lambda_1 + K)$ เมื่อ λ_1 เป็นค่า eigenvalue ที่มีค่ามากที่สุดของ $X'X$

ดังนั้น $(\hat{\beta}^*)'(\hat{\beta}^*) < \hat{\beta}'\hat{\beta}$

เนื่องจากผลบวกของความคลาดเคลื่อนกำลังสองของค่าตัวประมาณ $\hat{\beta}^*$ เป็นดังนี้

$$\phi^*(K) = (Y - X\,\hat{\beta}^*)'\,(Y - X\,\hat{\beta}^*) \qquad ....(2.3.10)$$

จากการประมาณค่าพารามิเตอร์ β ด้วยวิธีกำลังสองน้อยที่สุดเราสามารถเขียน $\phi(\hat{\beta})$ ซึ่งเป็นผลบวกของความคลาดเคลื่อนกำลังสองของตัวประมาณ $\hat{\beta}$ เป็น $Y'Y-\hat{\beta}'X'Y$ ดังนั้น เพื่อให้เห็นความแตกต่างของ $\phi^*(K)$ และ $\phi(\hat{\beta})$ เราสามารถเขียน $\phi^*(K)$ ในเทอมของ $Y'Y$ และ $X'Y$ ได้ดังนี้

$$\phi^*(K) = Y'Y - (\hat{\beta}^*)'X'Y - K(\hat{\beta}^*)'(\hat{\beta}^*) \quad ......(2.3.11)$$

จาก $\phi(\hat{\beta})$ และ $\phi^*(K)$ สรุปได้ว่าผลบวกของความคลาดเคลื่อนกำลังสองของวิธีริดจ์ รีเกรสชั่นจะให้ค่าน้อยกว่าผลบวกของความคลาดเคลื่อนกำลังสองที่คำนวณโดยวิธีกำลังสองน้อยที่สุด

๒.๔ รีดจ์เทรซ (The Ridge Trace)

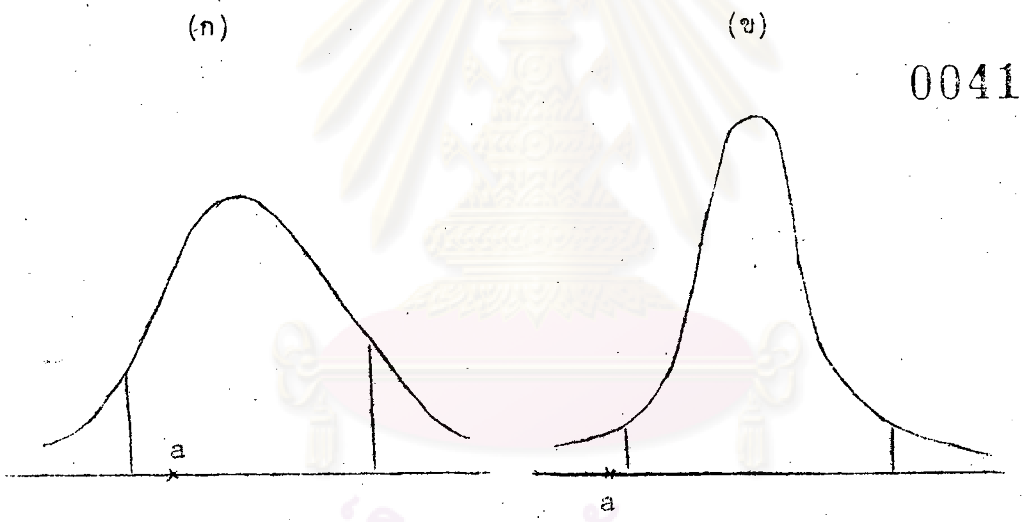
กรณีที่เมทริกซ์ของตัวแปรอิสระเกิดสภาพไม่เหมาะสมจะมีผลทำให้

๑. ตัวประมาณค่าพารามิเตอร์ β จะให้ค่าสัมบูรณ์มากเมื่อเปรียบเทียบกับค่า-สังเกตแล้วพบว่าอาจเป็นไปไม่ได้[๑]

๒. ตัวประมาณ $\hat{\beta}$ จะไม่สมเหตุสมผล (Unreasonable) เมื่อ $K = 0$ แต่ตัวประมาณค่า $\hat{\beta}$ อาจสมเหตุสมผล (reasonable) เมื่อคำนวณโดยกำหนดให้ $K > 0$ ดังตัวอย่างข้อมูลรูปข้างล่างนี้

รูปที่ ๑

แสดงการเปรียบเทียบการกระจายตัวประมาณค่าพารามิเตอร์ β เมื่อประมาณค่าโดยวิธีริดจ์ รีเกรสชั่น



จากรูป (ก) แสดงการกระจายของตัวประมาณค่า $\hat{\beta}$ พบว่าที่จุด a ซึ่งเป็นจุดที่ตกอยู่ในช่วงที่ยอมรับแต่ในทางปฏิบัติแล้วอาจไม่เป็นจริง ดังนั้น อาจถือว่าเป็นการเสี่ยงหรือมีข้อผิดพลาดสูงหากยอมรับค่าสังเกต แต่หากเราคำนวณตัวประมาณ $\hat{\beta}^*$ ด้วยวิธี

---

[๑]Arthur E. Hoerl and Robert W. Kennard, "Ridge Regression Biased Estimation for Nonorthogonal Problems", Technometric, (Vol 12, No.1, February 1970), p. 55-56.

ริดจ์ รีเกรสชั่น ดังรูป (ข) จะเห็นว่าการกระจายของตัวประมาณ $\hat{\beta}^*$ อยู่ในรูปปกติโด่ง ซึ่งมีความแปรปรวนต่ำ และจุด a จะอยู่ในช่วงที่เราไม่ยอมรับซึ่งหากเราไม่ยอมรับย่อมมีผลทำให้เกิดข้อผิดพลาดได้น้อย[๑]

เมื่อตัวแปรอิสระเกิดสภาพไม่เหมาะสมการประมาณค่าพารามิเตอร์ β ด้วยวิธีกำลังสองน้อยที่สุดจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูง[๒]

หลักการในการหาค่าประมาณของพารามิเตอร์ β คือจะเปลี่ยนค่า K ทีละน้อย โดยให้ $K > 0$ เสมอแล้วคำนวณตัวประมาณค่า $\hat{\beta}^*$ และค่าเฉลี่ยความคลาดเคลื่อนกำลัง-สองสำหรับค่า K ทุกค่า เพื่อให้ได้ค่า $\hat{\beta}^*$ ที่เหมาะสม โดยจะหยุดที่ค่า K ที่ทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าค่าที่ได้ด้วยวิธีกำลังสองน้อยที่สุด ทั้งนี้ Arthur และ Robert W. Kennard[๓] และ Donald W. Marquart และ Ronald S. Snee[๔] ได้เสนอแนะค่า ที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดว่าส่วนใหญ่จะอยู่ระหว่าง (0 , 1] ดังรูปที่ ๒

---

[๑]O.L. Obechain, "Classical F-test and Confidence Region for Ridge Regression" Technometrics, (Vol.10, No.4, November 1977), p. 429.

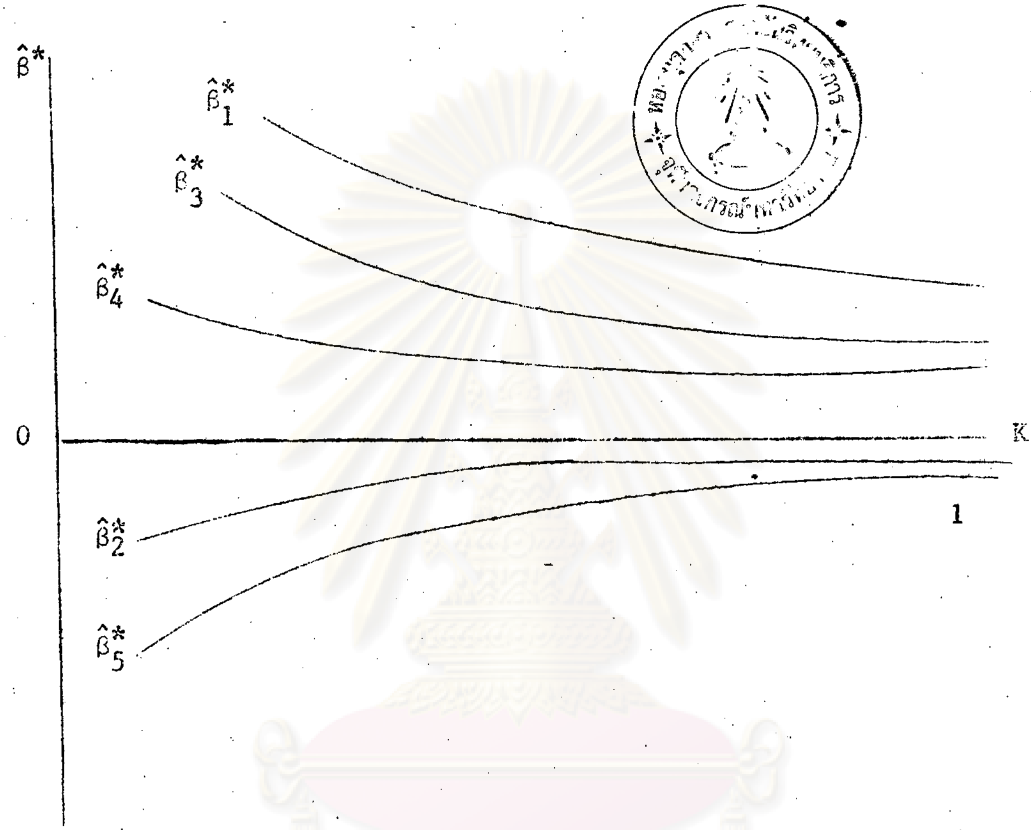
[๒]R.W. Farebrother, "The Minimium Mean Square Error Linear Estimator and Ridge Regression" Technometrics, (Vol.17, No.1, February 1975), p. 127.

[๓]Arthur E. Hoerl and Robert W. Kennard, "Ridge Regression Biased Estimation for Nonorthogonal Problems" Technometrics, (Vol.12, No.1, February 1970), p. 65.

[๔]Donald W. Marquart and Ronald D. Snee, "Ridge Regression in Practice" The American Statistician, (Vol. 29, No. 1, February 1975), p. 5.

รูปที่ ๒

แสดงความสัมพันธ์ระหว่าง $\hat{\beta}^*$ และ K เมื่อ $K \in (0,1]$ ($\hat{\beta}^*$ หมายถึงสัมประสิทธิ์ของ X_1)



ริดจ์เทรซหมายถึงกราฟที่แสดงความสัมพันธ์ระหว่าง $\hat{\beta}^*$ และ K จากรูปที่ ๒ ริดจ์เทรซเป็นกราฟเส้นโค้ง ซึ่งในช่วงแรกกราฟของค่า $\hat{\beta}^*$ จะเป็นเส้นโค้งที่มีค่า Slope สูงเมื่อ $K \longrightarrow 0$ และค่า Slopeจะค่อย ๆ ลดลงจนถึงจุดคงที่ และเมื่อ $K \longrightarrow 1$ จะสังเกตได้ว่าเส้นโค้งจะไม่มีการเปลี่ยนแปลง (ค่า Slope มีค่าเป็นศูนย์) จะสังเกตได้ว่า ณ. K เริ่มจะคงที่ มีผลทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ประมาณค่าโดยริดจ์ รีเกรสชั่น ให้ค่าน้อยกว่าวิธีกำลังสองน้อยที่สุด และเมื่อ K มีค่าเข้าใกล้ 1 หรือเท่ากับ 1 จะมีผลทำให้ค่าตัวประมาณ $\hat{\beta}^*$ เปลี่ยนแปลงน้อยมาก และเมื่อ K มีค่ามากกว่า 1 ค่าตัวประมาณ $\hat{\beta}^*$ จะให้ค่าไม่แตกต่างจากค่า $\hat{\beta}^*$ เมื่อ K = 1 หรือเข้าใกล้ 1

ฉนั้นในทางปฏิบัติ เมื่อนำวิธีริดจ์รีเกรสชั่นมาวิเคราะห์กับข้อมูลทั่ว ๆ ไป จึงคาดได้ว่าค่า K ที่เหมาะสมส่วนใหญ่ควรจะอยู่ในช่วง (0 , 1]

ดังนั้นจะเห็นว่าเส้นโค้งริดจ์เทรซ จึงเป็นตัวแทนที่สำคัญวิธีหนึ่งของการประมาณสัมประสิทธิ์ความถดถอยของวิธีริดจ์ รีเกรสชั่น ในบางครั้ง เพราะสามารถช่วยในการตัด-สินใจว่าจะใช้ค่า K ใด $(K \varepsilon (0,1])$ ที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยลงเมื่อเปรียบเทียบกับวิธีกำลังสองน้อยที่สุด โดยพิจารณาการเปลี่ยนแปลงของตัวประมาณ $\hat{\beta}^*$ เป็นเกณฑ์ด้วย

สมมติให้ B เป็นค่าใด ๆ ที่ได้จากการประมาณค่าพารามิเตอร์ β การพิจารณาผลบวกความคลาดเคลื่อนกำลังสองของตัวประมาณ B หรือ $\phi(B)$ เมื่อเปรียบเทียบกับค่า $\phi(\hat{\beta})$ ซึ่งมีค่าเท่ากับ $(Y - X\hat{\beta})'(Y - X\hat{\beta})$ สามารถเขียนได้ดังนี้

$$\phi(B) = (Y - XB)'(Y - XB) \qquad \ldots\ldots\ldots\ldots(2.4.1)$$

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

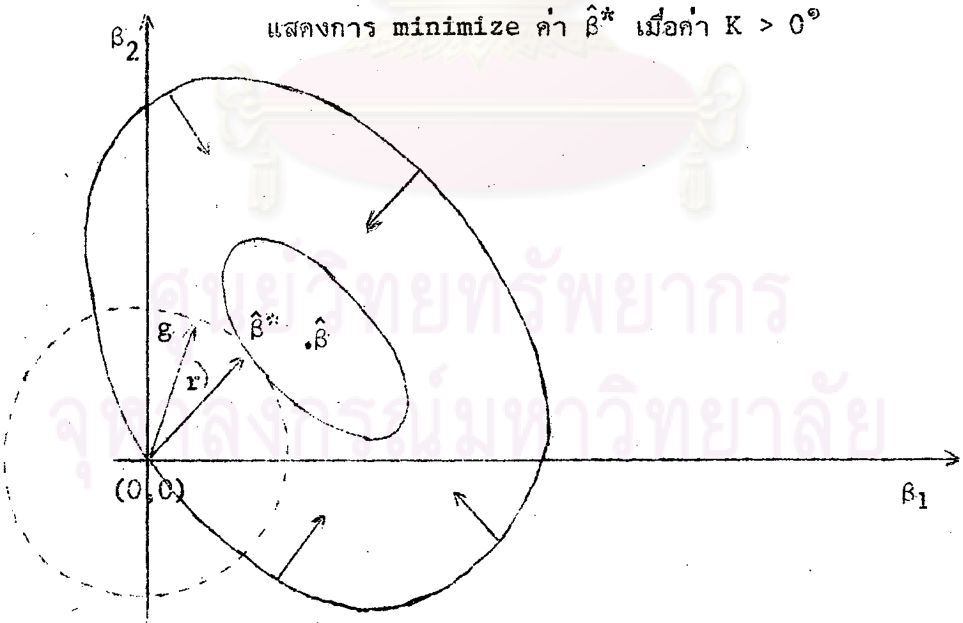
$$= \phi_{min} + \phi_B \qquad \ldots\ldots\ldots\ldots(2.4.2)$$

จากสมการ (2.4.2) $\phi(B)$ เขียนอยู่ในเทอมของผลบวกของ ϕ_{min} (มีค่าเท่ากับ $\phi(\hat{\beta})$) กับฟังค์ชั่นที่อยู่ในรูปกำลังสองของ $(B-\hat{\beta})$ โดยที่ ϕ_{min} เป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ $\hat{\beta}$ ที่คำนวณโดยวิธีกำลังสองน้อยที่สุด และถ้ากำหนดให้ B_o เป็นตัวประมาณพารามิเตอร์ β ที่สามารถคำนวณค่าได้ และเป็นตัวประมาณที่เหมาะสมที่สุด เมื่อพิจารณาผลบวกกำลังสองของความคลาดเคลื่อนของ B_o จะเห็นว่ามีค่าเท่ากับ ϕ_{min} กับ ϕ_{B_o} โดยที่ $\phi_{B_o} > 0$ ดังนั้นจะสังเกตได้ว่า ϕ_{min} เป็นค่าคงที่ที่เกิดขึ้นเสมอ ไม่ว่าจะเป็นการประมาณค่าพารามิเตอร์ β ด้วยวิธีใด ๆ ส่วน ϕ_{B_o} เป็นฟังค์ชั่นกำลังสองของ $(B_o - \beta)$ มีค่าเพิ่มขึ้น จากสมการ (2.2.5 ก) ในกรณีตัวแปรอิสระเกิดสภาพไม่เหมาะสม Hoerl และ Kennard ได้ชี้จุดบกพร่องของวิธีกำลังสองน้อยที่สุดโดยให้ข้อสังเกตว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\hat{\beta}$ จะมีค่าเท่ากับ

$\beta'\beta + \sigma^2$ Trace $(X'X)^{-1}$ โดยที่มีค่า $\beta'\beta$ เพิ่มขึ้นจากเดิม (σ^2 Trace $(X'X)^{-1}$) ค่า $\beta'\beta + \sigma^2$ Trace $(X'X)^{-1}$ หรือ $\beta'\beta + \sigma^2 \sum_{i=1}^{p} 1/\lambda_i$ เมื่อนำมาเปรียบเทียบกับวิธีริดจ์ รีเกรสชั่น จะสังเกตได้ว่าค่า eigenvalue ของ $X'X$ จะมีค่าต่ำมาก ๆ เมื่อตัวแปรอิสระเกิดสภาพไม่เหมาะสม ถ้าให้ λ_{min} เป็นค่า eigenvalue ของ $X'X$ ที่มีค่าต่ำสุด อาจกล่าวได้ว่าค่า $\beta'\beta + \sigma^2 \sum_{i=1}^{p} 1/\lambda_i > \beta'\beta + \sigma^2\ 1/\lambda_{min}$ เสมอ เมื่อ $\beta'\beta + \sigma^2/\lambda_{min}$ เป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่คำนวณโดยวิธีริดจ์ รีเกรสชั่น และผลที่ตามมาอีกคือ เวคเตอร์ $\hat{\beta}$ จะเพิ่มขึ้น วิธีที่จะลด $\beta'\beta$ ให้ลดลงนี้คือพยายามลดเวคเตอร์ $\hat{\beta}$ ให้เล็กลง โดยไม่มีผลทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ($\phi(B)$) เพิ่มขึ้น วิธีการที่ทำให้ $\hat{\beta}$ มีค่าเล็กลงจึงควรมุ่งไปพิจารณาที่ริดจ์เทรซ ซึ่งอาจหมายถึงระยะทางจากจุดกำเนิดใด ๆ ของเวคเตอร์ $\hat{\beta}^*$ ไปยังพื้นที่ผิวของรูปวงรี $\phi(B)$ ให้มีค่าสั้นที่สุดดังรูปที่ ๓

<u>รูปที่ ๓</u>

แสดงการ minimize ค่า $\hat{\beta}^*$ เมื่อค่า $K > 0$[๑]



---

[๑]Donald, W. Marquart and Ronald, D. Snee, "Ridge Regression in Practice" <u>The American Statistician</u>, (Vol.29, No.1, February 1975), p. 6.

จากรูปที่ ๓ แสดงให้เห็นว่าค่า $\phi(B)$ เป็นรูปวงรีที่มีจุดศูนย์กลางที่ $\hat{\beta}$ จะถูกทำให้มีค่าต่ำสุด เมื่อกำหนดให้ $\phi(B)$ เป็นรูปวงรีใหญ่จะมีค่าลดลงเรื่อย ๆ จนกระทั่งเป็นรูปวงรีขนาดเล็ก และเมื่อลากเส้นตรงจากจุดกำเนิดไปยังเส้นรอบวงรีขนาดเล็กนี้พบว่าเส้นตรงนี้มีค่าเท่ากับเวคเตอร์ $\hat{\beta}^*$ เป็นเส้นตรงที่สั้นที่สุดและอยู่ระหว่าง $\hat{\beta}$ กับ g เสมอ เมื่อ g เป็นเวคเตอร์ที่ลากจากจุดกำเนิดไปตั้งฉากกับรูปวงรี $\phi(B)$ และ g ทำมุมกับเวคเตอร์ $\hat{\beta}^*$ เท่ากับ γ ทุกครั้งที่ K มีค่าเพิ่มขึ้นค่าของมุม γ จะมีค่าเล็กลง ๆ ดังนั้นในการคำนวณค่า B จึงพิจารณาจากการหาค่าต่ำสุดของระยะทางของเวคเตอร์ B ไปยัง $\phi(B)$ นั่นเอง การหาค่าต่ำสุดของ $B'B$ เป็นดังนี้

กำหนดให้ $(B - \hat{\beta})' (X'X) (B - \hat{\beta}) = \phi_B$ ....(2.4.3)

การหาค่าต่ำสุดของ $B'B$ จะใช้ทฤษฎีและหลักของ Lagrangian โดยกำหนดให้ F ซึ่งเป็นฟังค์ชั่นที่จะหาค่าต่ำสุดดังนี้

$$F = B'B + 1/K \left[(B-\hat{\beta})'(X'X)(B-\hat{\beta})-\phi_B\right] \quad \text{...........(2.4.4)}$$

เมื่อหาอนุพันธ์ F เทียบกับ B แล้วกำหนดให้มีค่าเท่ากับศูนย์จะได้

$$\frac{\partial F}{\partial B} = 2B + 1/K \left[2(X'X)B-2(X'X)\hat{\beta}\right] = 0 \quad \text{.......(2.4.5)}$$

ฉนั้น $B = \hat{\beta}^* = [X'X + KI]^{-1} X'Y$ ....(2.4.6)

โดยที่ K เป็นค่าที่ทำให้ ϕ_B ต่ำสุดและมีผลทำให้ $B'B$ ต่ำสุดและผลบวกของความคลาดเคลื่อนกำลังสองของตัวประมาณค่า $\hat{\beta}^*$ สอดคล้องกับสมการ (๒.๔.๖) จากสมการ (๒.๔.๒) อาจเขียนอยู่ในเทอมของ $\hat{\beta}^*$ ได้ดังนี้

$$\text{จาก } \phi^*(K) = (Y - X\hat{\beta}^*)' (Y - X\hat{\beta}^*)$$

$$= \phi_{min} + (\hat{\beta}^* - \hat{\beta})'(X'X)(\hat{\beta}^*-\hat{\beta}) \quad \text{.......(2.4.7)}$$

จาก $(\hat{\beta}^*-\hat{\beta})'(X'X)(\hat{\beta}^*-\hat{\beta}) = \hat{\beta}^{*\prime}(I-Z^{-1})'(X'X)(I-Z^{-1})\hat{\beta}^*$ ......(2.4.8)

$$= \hat{\beta}^{*\prime} \left[I-I-K(X^{\prime}X)^{-1} \right](X^{\prime}X)\left[I-I-K(X^{\prime}X)^{-1} \right]\hat{\beta}^{*} \quad ...(2.4.9)$$

$$= K^{2}\, \hat{\beta}^{*\prime}(X^{\prime}X)^{-1}\, \hat{\beta}^{*} \quad ....(2.4.10)$$

ดังนั้นสมการ (2.4.7) สามารถเขียนใหม่ได้ดังนี้

$$\phi^{*}(K) = \phi_{min} + K^{2}\, \hat{\beta}^{*\prime}(X^{\prime}X)^{-1}\, \hat{\beta}^{*} \quad ....(2.4.11)$$

โดยที่ $\phi^{*}(K)$ เป็นผลบวกของความคลาดเคลื่อนกำลังสองของตัวประมาณ $\hat{\beta}^{*}$

จากสมมติฐานเบื้องต้นของการประมาณค่าพารามิเตอร์ β คือ $\varepsilon \sim N(0,\sigma^{2} I_{n})$ ฟังค์ชั่นภาวะน่าจะเป็น (Likelihood function) ของความคลาดเคลื่อนคือ

$$(2\pi\sigma^{2})^{-n/2}\exp\{(-1/2\sigma^{2})(Y-X\beta)^{\prime}(Y-X\beta)\} \quad ....(2.4.12)$$

$$(2\pi\sigma^{2})^{-n/2} \exp \{(-1/2\sigma^{2})\left[(Y-X\hat{\beta})^{\prime}(Y-X\hat{\beta})+(\beta-\hat{\beta})^{\prime}(X^{\prime}X)(\beta-\hat{\beta})\right]\}.(2.4.13)$$

เมื่อ $\hat{\beta}^{*}$ เป็นตัวประมาณค่าใด ๆ ของพารามิเตอร์ ที่ประมาณค่าโดยวิธีริดจ์ รีเกรสชั่น ภาวะน่าจะเป็นคือ

$$(2\pi\sigma^{2})^{-n/2} \exp \{(-1/2\sigma^{2})\left[(Y-X\hat{\beta})^{\prime}(Y-X\hat{\beta})+K^{2}\hat{\beta}^{*\prime}(X^{\prime}X)^{-1}\hat{\beta}^{*}\right]\} \quad ..(2.4.14)$$

จากสมการ (2.4.13) อาจกล่าวได้ว่าฟังค์ชั่นภาวะน่าจะเป็นมีค่าเพิ่มขึ้นในเทอมของ $(\beta-\hat{\beta})^{\prime}(X^{\prime}X)(\beta-\hat{\beta})$ แต่เมื่อพิจารณาสมการ (2.4.14) จะสังเกตได้ว่า $K^{2}(\hat{\beta}^{*})^{\prime}(X^{\prime}X)^{-1}\hat{\beta}^{*}$ มีค่าน้อยกว่าเทอม $(\beta-\hat{\beta})^{\prime}(X^{\prime}X)(\beta-\hat{\beta})$ ดังนั้นจึงสรุปได้ว่าฟังค์ชั่นภาวะน่าจะเป็นเมื่อประมาณค่าพารามิเตอร์ β ด้วยวิธีริดจ์ รีเกรสชั่น จะมีค่าลดลง (ซึ่งสังเกตได้จากรูปที่ ๓ เมื่อขนาดของรูปวงรีมีขนาดลดลงนั่นเอง) และฟังค์ชั่นภาวะน่าจะเป็น (การกระจายของค่าความคลาดเคลื่อนของตัวประมาณ $\hat{\beta}^{*}$) จะมีค่าเฉลี่ย (Mean) เท่ากันและค่าความแปรปรวนเท่ากัน โดยที่ค่าความคลาดเคลื่อน (ε)มีการกระจายแบบปกติคือ $\varepsilon \sim N(0, \sigma^{2} I_{n})$

๒.๕ คุณสมบัติของผลบวกของความคลาดเคลื่อนกำลังสองของริดจ์ รีเกรสชั่น

๒.๕.๑ จากการประมาณค่าสัมประสิทธิ์การถดถอย β ด้วยวิธีกำลังสองน้อยที่สุดจะได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองดังนี้

$$MSE = VAR(\hat{\beta}) + (BIAS)^2 \qquad \ldots\ldots\ldots(2.5.1)$$

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

เนื่องจาก $E(\hat{\beta}) = \beta$ ดังนั้นการประมาณค่าพารามิเตอร์ β ด้วยวิธีกำลังสองน้อยที่สุดจะได้ตัวประมาณค่า $\hat{\beta}$ ที่ไม่เอนเอียง สำหรับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณสัมประสิทธิ์ความถดถอยด้วยวิธี ริดจ์ รีเกรสชั่น ซึ่งมีค่าเป็น $E\left[L_1^2(K)\right]$ อาจพิจารณาได้ดังนี้

ในการประมาณค่าพารามิเตอร์ β ด้วยวิธีริดจ์ รีเกรสชั่น จะได้ $E(\hat{\beta}^*) \neq \beta$[๑] แสดงว่าค่าตัวประมาณ $\hat{\beta}^*$ มีลักษณะเอนเอียง

$$E(\beta^*) = E(Z\hat{\beta}) \qquad \ldots\ldots\ldots(2.5.2)$$

$$= ZE(\hat{\beta}) \qquad \ldots\ldots\ldots(2.5.3)$$

$$= Z\beta \qquad \ldots\ldots\ldots(2.5.4)$$

จาก

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

$$= E\left[(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta)\right] + (Z\beta-\beta)'(Z\beta-\beta) \qquad \ldots\ldots\ldots(2.5.5)$$

$$= \sigma^2 \text{ Trace } (X'X)Z'Z + \beta'(Z-I)'(Z-I)\beta \qquad \ldots\ldots\ldots(2.5.6)$$

---

[๑]Donald W. Marquart and Ronald D. Snee, "Ridge Regression in Practice" <u>The American Statisticien</u>, (Vol.29, No.1, February 1975) p. 5.

$$= \sigma^2 \left[\text{Trace } (X'X + KI)^{-1} - K \text{ Trace } (X'X + KI)^{-2} \right]$$

$$+ K^2 \beta'(X'X + KI)^{-2} \beta \qquad \ldots\ldots.(2.5.7)$$

$$= \sigma^2 \sum_{i=1}^{P} \lambda_i/(\lambda_i+K)^{-2} + K^2 \beta'(X'X+KI)^{-2}\beta \quad \ldots..(2.5.8)$$

$$= \gamma_1(K) + \gamma_2(K) \qquad \ldots\ldots.(2.5.9)$$

จะเห็นว่า $E[L_1^2(K)]$ เป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองซึ่งอยู่ในรูปผลบวกของฟังค์ชั่น $\gamma_1(K)$ และ $\gamma_2(K)$

โดยที่ $\gamma_1(K)$ เป็นค่าความแปรปรวนของตัวประมาณ $\hat{\beta}^*$ และ $\gamma_2(K)$ เป็นค่าความเอนเอียงของตัวประมาณค่า $\hat{\beta}^*$ กำลังสอง เมื่อพิจารณาเทอม $\gamma_2(K)$ ซึ่งเป็นระยะทางจาก $Z\beta$ ไปยัง β จะเห็นว่ามีค่าเป็นศูนย์เมื่อ $K = 0$ โดยที่ Z เป็นเมทริกซ์เอกลักษณ์การคูณ ซึ่งสอดคล้องกับค่าความเอนเอียงกำลังสองของ $\hat{\beta}$ ที่มีค่าเป็นศูนย์ สำหรับกรณีที่ $K > 0$ อาจพิจารณา $\gamma_2(K)$ ในเทอมของความเอนเอียงของ $\hat{\beta}$ กำลัง-สอง กล่าวคือ $\gamma_2(K) = (\text{BIAS})^2$

ส่วนเทอม $\gamma_1(K)$ นั้นเป็นผลบวกของความแปรปรวนของตัวประมาณค่า $\hat{\beta}^*$ ซึ่งเขียนในเทอมของ Y ได้ดังนี้

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

$$= Z(X'X)^{-1} X'Y \qquad \ldots\ldots..(2.5.10)$$

$$V(\hat{\beta}^*) = Z(X'X)^{-1} X'\text{VAR}[Y] X(X'X)^{-1} Z'$$

$$= \sigma^2 \text{ Trace } Z(X'X)^{-1} Z' \qquad \ldots\ldots..(2.5.11)$$

เนื่องจากความแปรปรวนของตัวประมาณ $\hat{\beta}^*$ ทั้งหมด $V(\hat{\beta}^*)$ คำนวณได้จากผลบวกของสมาชิกทุกตัวบนเส้นทะแยงมุมของเมทริกซ์ $\sigma^2 Z(X'X)^{-1}Z$ เพื่อให้สอดคล้องกับ

$\gamma_1(K)$ ของสมการ (2.5.9) ซึ่งเป็นฟังค์ชั่นที่อยู่ในเทอมของ eigenvalue ของ $X'X$ และ K ดังนั้นเมทริกซ์ $X'X$ และ Z สามารถปรับให้เป็นเมทริกซ์อยู่ในรูปของ λ_i ซึ่งเป็นค่า eigenvalue ของ $X'X$ เมื่อ $i = 1, 2, \ldots, p$ และ p หมายถึงจำนวนตัวแปรอิสระ X

$$\text{Trace } (X'X)^{-1} = \sum_{i=1}^{p} (1/\lambda_i) \qquad \ldots\ldots(2.5.12)$$

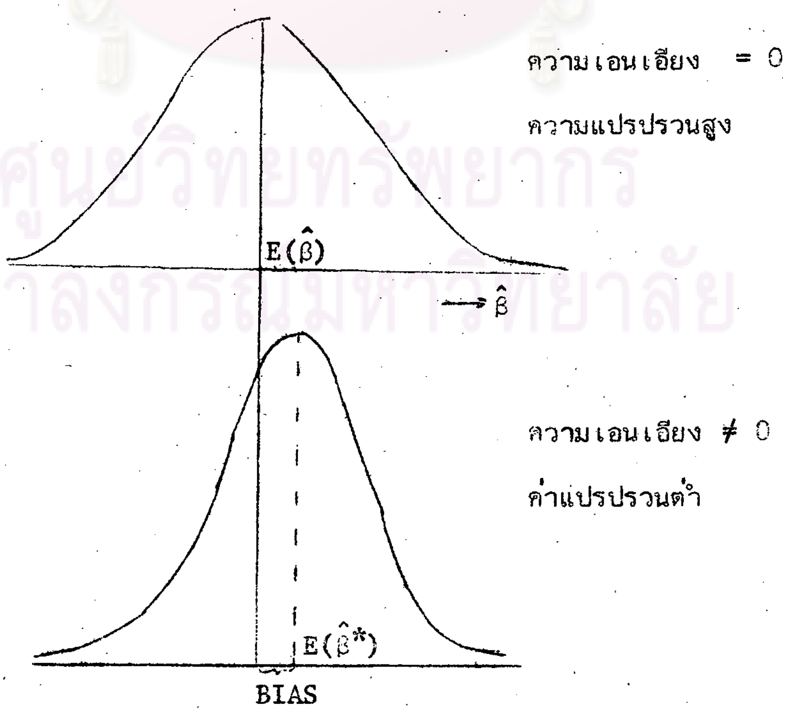
และ
$$\text{Trace } (Z) = \sum_{i=1}^{p} (\lambda_i/\lambda_i+K) \qquad \ldots\ldots(2.5.13)$$

จากสมการ (2.5.11) จะได้ $\sigma^2 \text{ Trace } Z(X'X)^{-1} Z = \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+K)^2$

ดังนั้นกล่าวได้ว่า $E[\, L_1^2(K) \,] = VAR(\hat{\beta}^*) + (BIAS)^2$

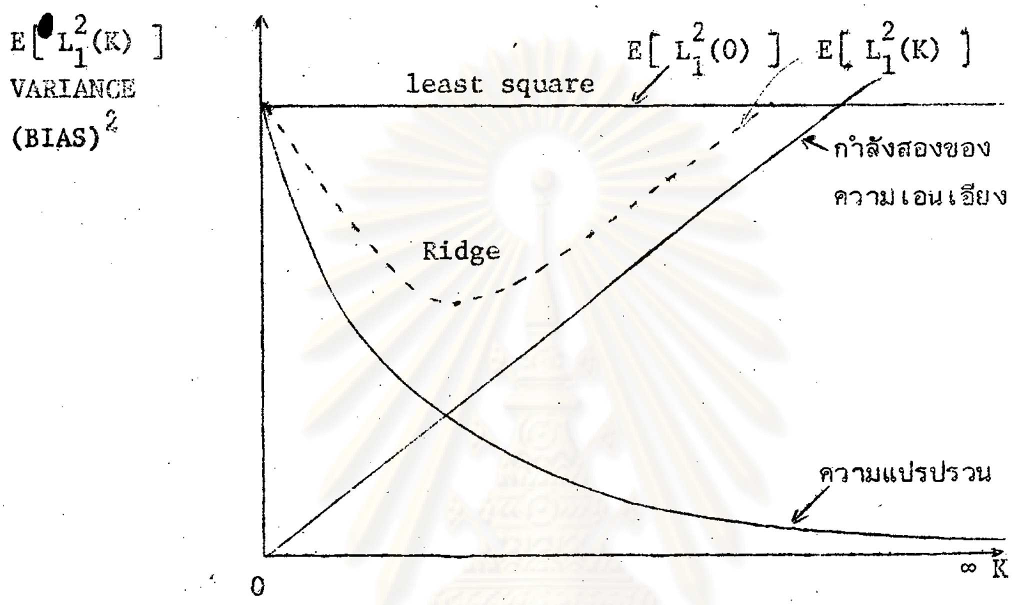
<u>รูปที่ ๔</u>

แสดงการกระจายของตัวประมาณพารามิเตอร์ β ทั้ง ๒ วิธี



รูปที่ ๕

แสดงกราฟของ $E[\,L_1^2(K)\,]$



จากรูปที่ ๕ เป็นกราฟแสดงความสัมพันธ์ระหว่างความแปรปรวนของตัวประมาณ $\hat{\beta}^*$ ความเอนเอียงกำลังสองกับค่า K เมื่อกำหนดให้ $K\varepsilon(0\ ,\ 1]$ สรุปได้ดังนี้ คือความแปรปรวนทั้งหมดจะมีค่าลดลงเรื่อย ๆ ขณะที่ K มีค่าเพิ่มขึ้น และความเอนเอียงกำลังสองจะมีค่าแปรผันตาม K จากรูปเส้นไข่ปลาเป็นกราฟที่แสดงผลบวกของความแปรปรวนทั้งหมดของ $\hat{\beta}^*$ และความเอนเอียงกำลังสองของ $\hat{\beta}^*$ จะมีค่าเท่ากับ $E[\,L_1^2(K)\,]$ ซึ่งเป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดจากตัวประมาณค่า $\hat{\beta}^*$ เมื่อ $K > 0$

กราฟของฟังก์ชั่น $\gamma_1(K)$ มีลักษณะเป็น monotonically decreasing ในรูปฟังก์ชั่นของ K และกราฟของฟังก์ชั่น $\gamma_2(K)$ มีลักษณะเป็น monotonically increasing ในรูปฟังก์ชั่นของ K ดังนั้นจากรูปที่ ๕ จะเห็นว่า $E[\,L_1^2(K)\,]$ มีค่าน้อยกว่า $E[\,L_1^2(0)\,]$ โดยที่ $E[\,L_1^2(0)\,]$ เป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดจากการประมาณค่าพารามิเตอร์ β โดยวิธีกำลังสองน้อยที่สุด

จากรูปที่ ๕ แสดงให้เห็นการเปลี่ยนแปลงของค่าต่าง ๆ ที่กล่าวมาแล้วข้างต้น ซึ่งสามารถคำนวณหาขีดจำกัด (Limit) ของความแปรปรวนของตัวประมาณ $\hat{\beta}^*$ และความเอนเอียงกำลังสองของตัวประมาณ $\hat{\beta}^*$ ได้จากอนุพันธ์ของแต่ละฟังก์ชั่นดังนี้

$$\lim_{k \longrightarrow o^+} (\partial\gamma_1/\partial K) = -2\sigma^2 \sum_{i=1}^{p} (1/\lambda_i^2) \qquad ....(2.5.14)$$

$$\lim_{k \longrightarrow o^+} (\partial\gamma_2/\partial K) = 0 \qquad ....(2.5.15)$$

ดังนั้นอนุพันธ์ของความเอนเอียงกำลังสองมีค่าเป็นลบ และเข้าใกล้ $-2P\sigma^2$ ขณะที่ $k \longrightarrow o^+$ เมื่อผลคูณของเมทริกซ์ของตัวแปรอิสระ ($X'X$) เป็นออร์ธอโทนอล และ $\gamma_1(K)$ มีค่าเข้าใกล้ค่าอนันต์ลบ ($-\infty$) กรณีที่เมทริกซ์ของตัวแปรอิสระเกิดมีสภาพไม่เหมาะสม ค่า eigenvalue $\lambda_p \longrightarrow 0$

ส่วน $\lim_{k \longrightarrow o^+} (\partial\gamma_2/\partial K)$ มีค่าค่อย ๆ ลดลงและเป็นศูนย์ที่จุดกำเนิดเมื่อ $k \longrightarrow o^+$ คุณสมบัติของสมการ (2.5.14) และ (2.5.15) จะเป็นจริงเมื่อ $K > 0$ ซึ่งสรุปได้ว่าค่าความเอนเอียงเพิ่มขึ้นเรื่อย ๆ ขณะที่ความแปรปรวนของตัวประมาณ $\hat{\beta}^*$ ลดลง ซึ่งจะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลงด้วย

๒.๔.๒ ทฤษฎีต่าง ๆ ที่สำคัญของฟังก์ชั่นค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

ทฤษฎี ๒.๔.๒.๑ ความแปรปรวนทั้งหมด $\gamma_1(K)$ เป็นฟังก์ชั่นต่อเนื่องและมีลักษณะเป็น monotonically decreasing ฟังก์ชั่นในรูปของ K

พิสูจน์ จาก (2.5.8) $\gamma_1(K) = \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + K)^2$ ....(2.5.16)

หาอนุพันธ์ของ $\gamma_1(K)$ เทียบกับ K

$$\frac{\partial\gamma_1(K)}{\partial K} = \sigma^2 \sum_{i=1}^{p} \frac{\partial\lambda_i (\lambda_i + K)^{-2}}{\partial K} \qquad ......(2.5.17)$$

$$= \sigma^2 \sum_{i=1}^{p} \lambda_i(-2)(\lambda_i + K)^{-3}$$

$$\gamma_1'(K) = -2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + K)^3 \qquad \ldots\ldots(2.5.18)$$

จาก $K > 0$, $\gamma_1(K)$ จะเป็นฟังค์ชั่นต่อเนื่องของ K และพบว่า $\gamma_1'(K)$ เป็นฟังค์ชั่นที่มีค่าเป็นลบ ดังนั้น $\lim_{K \to \infty} \gamma_1'(K)$ จะมีลักษณะเป็น monotonically decreasing ฟังค์ชั่น K

ดังนั้นอนุพันธ์อันดับหนึ่งของ $\gamma_1(K)$ เมื่อเทียบกับ K มีค่าเป็น $\gamma_1'(K)$ จะเข้าใกล้ $-\infty$ เมื่อ $k \to o^+$ และ $\lambda_p \to o$

ดังนี้คือ $\lim_{k \to o^+} \gamma_1'(K) = -\infty$

ทฤษฎี ๒.๕.๒.๒ ความเอนเอียงกำลังสองของ $\gamma_2(K)$ เป็นฟังค์ชั่นต่อเนื่องและ monotonically increasing ฟังค์ชั่นของ K

พิสูจน์

จาก (2.5.8) $\gamma_2(K) = K^2\beta'(X'X + KI_p)^{-2}\beta$

ถ้า λ เป็นเมทริกซ์ของ eigenvalue $X'X$ และ p เป็น orthogonal เมทริกซ์ที่มีคุณสมบัติ $X'X = P\lambda P$ ดังนั้น $\gamma_2(K)$ อาจเขียนใหม่ได้ดังนี้

$$\gamma_2(K) = K^2 \sum_{i=1}^{p} \alpha_i^2 / (\lambda_i + K)^2 \qquad \ldots\ldots(2.5.19)$$

เมื่อ $\alpha = P\beta$ ......(2.5.20)

เนื่องจาก $\lambda_i > 0$ ทุกค่าของ i และ $K > 0$ ดังนั้น แต่ละเทอมของ $(\lambda_i + K)$ จะเป็นบวก จาก (2.5.19) จะได้ $\gamma_2(0) = 0$ เมื่อ $K = 0$ แล้วจะได้ $\gamma_2(K)$ เป็นฟังค์ชั่นต่อเนื่องเมื่อ $K > 0$ อาจเขียน $\gamma_2(K)$ ใหม่ดังนี้

$$\gamma_2(K) = \sum_{i=1}^{p} \alpha_i^2/[1 + (\lambda_i/K)]^2 \qquad \ldots\ldots(2.5.21)$$

ทุกค่าของ i, $\lambda_i > 0$ ดังนั้น λ_i/K จะเป็น monotonically decreasing เมื่อ K มีค่าเพิ่มขึ้น และแต่ละเทอมของ $\gamma_2(K)$ เป็น monotonically increasing ดังนั้น $\gamma_2(K)$ เป็น monotonically increasing ฟังค์ชั่นของ K

บทแทรก ๒.๕.๒.๒ (๑) ความเอนเอียงกำลังสอง $\gamma_2(K)$ จะมีค่าเข้าใกล้ $\beta'\beta$ ซึ่งเป็นขีดจำกัดบน

พิสูจน์

จาก (2.5.21) $\lim\limits_{k\to\infty} \gamma_2(K) = \sum\limits_{i=1}^{p} \alpha_i^2$

$$= \alpha'\alpha$$

$$= \beta'P'P\beta$$

$$= \beta'\beta$$

บทแทรก ๒.๕.๒.๒ (๒) อนุพันธ์ $\gamma_2'(K)$ จะมีค่าเข้าใกล้ 0 ขณะที่ $k \to o^+$

พิสูจน์

จาก (2.4.19) $\gamma_2(K) = K^2 \sum\limits_{i=1}^{p} \alpha_i^2/(\lambda_i + K)^2$

$$\frac{\partial\gamma_2(K)}{\partial K} = 2K \sum_{i=1}^{p} \lambda_i\ \alpha_i^2/(\lambda_i + K)^3 \quad .....(2.5.22)$$

ผลบวกของแต่ละเทอม $2K\alpha_i^2\ \lambda_i/(\lambda_i + K)^3$; $i = 1,2,...,p$ เป็นฟังค์ชั่นต่อเนื่อง และขีดจำกัด (Limit) ของแต่ละเทอมจะเข้าใกล้ 0 ขณะที่ $k \to o^+$ ดังนั้น $\gamma_2'(K)$ จะเข้าใกล้ 0 ขณะที่ $k \to o^+$

ทฤษฎี ๒.๕.๒.๓ สำหรับทุกค่าของ $K > 0$

$E[\ L_1^2(K)\ ] < E[\ L_1^2(0)\ ]$ เมื่อ $E[\ L_1^2(0)] \ = \ \sigma^2 \sum\limits_{i=1}^{p} (1/\lambda_i)$

พิสูจน์

จาก (2.5.8), (2.5.19) และ (2.5.22

คำนวณค่าอนุพันธ์ $E[\ L_1^2(K)\ ]$ เทียบกับ K ได้ดังนี้

$$\frac{\partial E[\ L_1^2(K)\ ]}{\partial K} = \frac{\partial \gamma_1(K)}{\partial K} + \frac{\partial \gamma_2(K)}{\partial K}$$

$$= -2\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+K)^3 + 2K \sum_{i=1}^{p} \lambda_i\alpha_i^2/(\lambda_i+K)^3$$

.......(2.5.23)

$$\frac{\partial E[\ L_1^2(0)\ ]}{\partial K} = \frac{\partial \gamma_1(K)}{\partial K} + \frac{\partial \gamma_2(K)}{\partial K}$$

$$= \frac{\partial \left[\sigma^2 \sum_{i=1}^{p} (1/\lambda_i)\right]}{\partial K} + 0$$

$$= 0$$

.......(2.5.24)

เนื่องจาก K = 0 จะได้ $\gamma_1(0) = \sigma^2 \sum_{i=1}^{p} (1/\lambda_i)$ และ $\gamma_2(0) = 0$

การที่จะพิสูจน์ว่า $E[\ L_1^2(K)\ ] < E[\ L_1^2(0)\ ]$ ต้องแสดงให้

เห็นจริงว่า $\frac{\partial E[\ L_1^2(K)\ ]}{\partial K} < \frac{\partial E[\ L_1^2(0)\ ]}{\partial K}$; $K > 0$

จาก (2.5.24) $\frac{\partial E[\ L_1^2(0)\ ]}{\partial K} = 0$ ดังนั้น $\frac{\partial E[\ L_1^2(K)\ ]}{\partial K} < 0$

จึงจะสรุปได้ว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่คำนวณโดยวิธีริดจ์รีเกรสชั่นจะให้ค่าน้อยกว่า คำนวณโดยวิธีกำลังสองน้อยที่สุด

เนื่องจาก $K < \sigma^2/\hat{\alpha}^2$ max จะเป็น K ที่เหมาะสม (optimium)

ดังนั้นแทนค่า K ในสมการ (2.5.23) จะได้ $\frac{\partial E[\ L_1^2(K)\ ]}{\partial K} < 0$

นั่นคือ $E[\,L_1^2(K)\,] < E[\,L_1^2(0)\,]$

๒.๖ Generalized Ridge Regression

หลักการของ Generalized Ridge Regression จะแปลงผลคูณเมทริกซ์ของตัวแปรอิสระให้อยู่ในรูป diagonal เมทริกซ์ โดยสมมติให้ P เป็น orthogonal เมทริกซ์ซึ่งมีความสัมพันธ์กับผลคูณเมทริกซ์ของตัวแปรอิสระดังนี้คือ $X'X = P'\lambda P$ เมื่อ λ เป็นเมทริกซ์ของค่า eigenvalue ที่เป็น diagonal เมทริกซ์ ดังนั้น จากสมการถดถอยพหุเชิงเส้นทั่วไปเราสามารถแปลงให้อยู่ในรูป canonical form ได้ดังนี้

$$Y = X\beta + \varepsilon \qquad \ldots\ldots(2.6.1)$$

กำหนดให้

$$X^* = XP \qquad \ldots\ldots(2.6.2)$$

$$\alpha = P'\beta \qquad \ldots\ldots(2.6.3)$$

นำ (2.6.2) และ (2.6.3)แทนค่าใน (2.6.1) จะได้ Y ในรูปแบบ canonical form ดังนี้

$$Y = X^*\alpha + \varepsilon \qquad \ldots\ldots(2.6.4)$$

เมื่อ α เป็นเวคเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด $P \times 1$

X^* เป็นเมทริกซ์ของตัวแปรอิสระที่อยู่ในรูป Cononical ที่มีคุณสมบัติ $(X^*)'(X^*) = \lambda$

จาก (2.6.4) ประมาณค่าพารามิเตอร์ α โดยวิธีริดจ์ รีเกรสชั่น จะได้ดังนี้

$$\hat{\alpha}(K) = [\,X^{*\prime}X^* + KI_p\,]^{-1}(X^*)'Y \qquad \ldots\ldots(2.6.5)$$

จาก (2.6.5) เมื่อ $K = 0$, $\hat{\alpha}(K)$ คือตัวประมาณค่าพารามิเตอร์ β ที่ประมาณค่าด้วยวิธีกำลังสองน้อยที่สุด โดยที่ $\hat{\alpha}_i(0) = \hat{\alpha}_i$ ซึ่งพิจารณาได้จาก

$$\hat{\alpha}_i(K) = \frac{\lambda_i}{\lambda_i+K} \cdot \hat{\alpha}_i \quad i = 1,2,\ldots,p \quad \ldots.(2.6.6)$$

$$= \hat{\alpha}_i \quad \text{เมื่อ} \quad K = 0 \quad \ldots\ldots\ldots(2.6.7)$$

เมื่อ $\lambda_1, \lambda_2, \ldots, \lambda_p$ เป็นค่า eigenvalue ของ $X'X$

ดังนั้นจะได้ค่าสัมประสิทธิ์ความถดถอยของ β คือ $\hat{\beta}^*$ เขียนอยู่ในรูป Generalized Ridge Regression คือ

$$\hat{\beta}^* = P.\hat{\alpha}(K) \quad \ldots\ldots\ldots(2.6.8)$$[๑]

จะเห็นว่า $\hat{\beta}^*$ เป็นฟังค์ชั่นของ P และ $\hat{\alpha}(K)$ การคำนวณตัวประมาณ $\hat{\beta}^*$ นั้น จะต้องกำหนดค่า K ที่เหมาะสม เพื่อให้การประมาณค่า β มีค่าเฉลี่ยความคลาดเคลื่อนกำลัง-สองต่ำสุด

สำหรับการกำหนดค่า K นี้ มีหลายวิธีดังต่อไปนี้

<u>วิธีที่ ๑</u> (Hoerl and Kennard [5])

$$K = \sigma^2/\max(\hat{\alpha}_i^2(0)) \quad \ldots\ldots\ldots(2.6.9)$$

ค่า K สามารถคำนวณค่าได้โดยที่ $\sigma^2, \hat{\alpha}_i(0)$ เป็นค่าที่ทราบ

<u>วิธีที่ ๒</u> (Lawless and Wang [8])

$$K = P\sigma^2 / \sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2(0) \quad \ldots\ldots\ldots(2.6.10)$$

ค่า K สามารถคำนวณค่าได้โดยที่ $\sigma^2, \lambda_i, \hat{\alpha}_i(0)$, P เป็นค่าที่ทราบ

---

[๑] W.J. Hermerle and T.F. Brantle, "Explicit and Constrain Generized Ridge Estimation; <u>Technometrics</u>, (Vol. 20 No. 2 May 1978), p. 110.

<u>วิธีที่ ๓</u> (Hoerl, Kennard and Badwin [7])

$$K = P\sigma^2 / \sum_{i=1}^{P} \hat{\alpha}_i^2(0) \qquad \text{.......(2.6.11)}$$

ค่า K สามารถคำนวณค่าได้โดยที่ P, σ^2, $\hat{\alpha}_i(0)$ เป็นค่าที่ทราบ

<u>วิธีที่ ๔</u> (McDonald and Galarneau [13])

$$Q = \hat{\underset{\sim}{\alpha}}'(0)\ \hat{\underset{\sim}{\alpha}}(0) - \hat{\sigma}^2 \left(\sum_{i=1}^{P} 1/\lambda_i \right) \text{ และเลือก K ...(2.6.12)}$$

ที่ให้ค่า $\hat{\underset{\sim}{\alpha}}'(K)\ \hat{\underset{\sim}{\alpha}}'(K) = Q$ โดยที่ $\hat{\underset{\sim}{\alpha}}$, $\hat{\underset{\sim}{\alpha}}(K)$, $\hat{\sigma}^2$, λ_i เป็นค่าที่ทราบ

สำหรับวิธีต่าง ๆ ดังกล่าวข้างต้นนี้ DEAN W. WICHERN AND GILBERT A. CHURCHILL ได้นำข้อมูลชุดเดียวกันซึ่งมีพหุสัมพันธ์ในตัวแปรอิสระทดลองใช้กับวิธีทั้งสี่ เพื่อเลือกหา K ที่เหมาะสม พบว่าวิธีที่ ๑ จะให้ค่า K ที่เหมาะสมที่มีผลทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่คำนวณโดยวิธีริดจ์ รีเกรสชั่น น้อยกว่าวิธีกำลังสองน้อยที่สุด[๑] ซึ่งค่า K ที่กำหนดในวิธีที่ ๑ คือ

$K = \sigma^2/\alpha^2(\max\ (0))$ สามารถพิสูจน์ได้ดังนี้

$$\hat{\alpha}^* = [\,D + K\,]^{-1}\ (X^*)'Y \qquad \text{.......(2.6.13)}$$

เมื่อ $\hat{\alpha}^*$ เป็นตัวประมาณค่าพารามิเตอร์ α ขนาด $P \times 1$

D เป็นไดเอโกนอลเมทริกซ์ที่สมาชิกทุกตัวมีค่าเป็น λ_i โดยที่ λ_i เป็นค่า eigenvalue ของ $X'X$, $i = 1,2,\ldots,P$ และเมทริกซ์ D มีขนาด $n \times n$

K เป็นค่าคงที่ที่มากกว่าศูนย์

---

[๑] Dean W. Wichern and Gilbert A. Churchill, "Comparison of Ridge Estimators"; <u>Technometrics</u>, (Vol.20, No.3, August 1978), p. 303.

W.J. Hermerle และ T.F. Brantle ให้ข้อสังเกตไว้ คือ การเลือก K ที่เหมาะสมนั้นได้จากการหาค่าต่ำสุดของค่าเฉลี่ยของระยะทางจาก $\hat{\alpha}^*$ ไปยัง α กำลังสองดังนี้คือ

จากสมการ (2.6.13) ค่าเฉลี่ยของระยะทางของตัวประมาณ ค่า $\hat{\alpha}^*$ ห่างจากค่าพารามิเตอร์ α กำลังสอง คือ

$$E\left[(\hat{\alpha}^* - \hat{\alpha})'(\hat{\alpha}^* - \hat{\alpha}) \right] = Q \qquad \ldots\ldots.(2.6.14)$$

[๑]

และจากสมการ (2.6.14) William J. Hermerle ได้มีการพิสูจน์การหาค่า K ที่เหมาะสมดังนี้คือเทอม $E\left[(\hat{\alpha}^* - \alpha)'(\hat{\alpha}^* - \hat{\alpha}) \right]$ สามารถกระจายให้อยู่ในรูปฟังค์ชั่นของ λ_i, α_i, K และ σ^2 ได้

$$Q = \sum_{i=1}^{p} (\sigma^2\lambda_i + \alpha_i^2 K_i^2)/(\lambda_i + K_i)^2 \qquad \ldots\ldots.(2.6.15)$$

เพื่อคำนวณค่า K ที่เหมาะสมจะหาค่าต่ำสุดของสมการ (2.6.15) ได้ โดยหาอนุพันธ์ Q เทียบกับ K และกำหนดให้มีค่าเท่ากับศูนย์ จะได้

$$\frac{\partial Q}{\partial K_{(i)}} = \frac{2\lambda_i(\lambda_i + K)(K\alpha_i^2 - \sigma^2)}{(\lambda_i + K_i)^4} = 0 \qquad \ldots\ldots.(2.6.16)$$

---

[๑] W.J. Hermerle and T.F. Brantle "Explicit and Censtrain Generized Ridge Estimation", Technometrics, (Vol. 20, No.2, May 1978), p. 109.

$$K_i = \sigma^2/\hat{\alpha}_i^2 \quad ; \quad i = 1, 2, \ldots, P \qquad \ldots\ldots\ldots(2.6.17)$$ [๑]

เนื่องจาก $\hat{\alpha}_i$ เป็นตัวประมาณค่าพารามิเตอร์ β ที่ประมาณค่าโดยวิธีกำลังสองน้อยที่สุด ซึ่งมีค่า $\hat{\alpha}_i(0)$ มีค่าเท่ากับ $\hat{\beta}_i$ ดังนั้นจะสรุปได้ว่าค่า K เป็นฟังค์ชั่นที่ขึ้นกับค่า σ^2, $\hat{\beta}$, $(K = \sigma^2/\hat{\beta}'\hat{\beta})$ ดังสมการ (2.6.17) เนื่องจากค่า $\hat{\alpha}_i$ และ K_i มีหลายค่าเพื่อสะดวกในการกำหนดค่า K เราอาจกำหนดค่า K เพียงค่าเดียวที่สามารถปรับเมทริกซ์ $X'X$ ที่มีผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองให้มีค่าต่ำสุดจากสมการ (2.6.17) กำหนดค่า K ใหม่ดังนี้

$$K = \sigma^2/\max \hat{\alpha}_i^2(0)$$

แต่ส่วนมากจะสังเกตได้ว่าค่า K optimum เมื่อ $K = 2\sigma^2/\hat{\beta}'\hat{\beta}$ [๒]

---

[๑] William J. Hemmerl, "An Explicit for Generalized Ridge Regression", Technometrics, (Vol. 17, No.13, August 1975), p. 309 - 310.

[๒] R.W. Farebrother, "Partitioned Ridge Regression". Technometrics, (Vol. 20, No. 2, May 1978), p. 121.

ในการวิจัยนี้ใช้สูตรในการคำนวณเพื่อเปรียบเทียบการประมาณค่าพารามิเตอร์ β ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น[๙] ดังนี้

ตารางที่ ๑

แสดงการเปรียบเทียบการประมาณค่า β ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีริดจ์ รีเกรสชั่น

	วิธีกำลังสองน้อยที่สุด	วิธีริดจ์ รีเกรสชั่น
๑) ค่าประมาณค่าสัมประสิทธิ์ความถดถอย β	$\hat{\beta}=(X'X)^{-1}X'Y$	$\hat{\beta}^{*}=(X'X+KI_p)^{-1}X'Y \;;\; K>0$
๒) ค่าประมาณความเอนเอียงของตัวประมาณ β	ไม่เอนเอียงไปจากประชากร โดยที่ $E(\hat{\beta}) = \beta$	เอนเอียงไปจากประชากร $E(\hat{\beta}^{*})= Z\beta$ โดยที่ $Z = (X'X + KI)^{-1}$
๓) ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง	$MSE(\hat{\beta})=\sigma^2 Trace(X'X)^{-1}$	$MSE(\hat{\beta}^{*})=Trace\left[VAR(\hat{\beta}^{*})\right]+\beta'(Z-I)(Z-I)\beta$
๔) ความแปรปรวนของสัมประสิทธิ์ความถดถอย β	$VAR(\hat{\beta})=\sigma^2_{y.x}(X'X)^{-1}$	$VAR(\hat{\beta}^{*})=\sigma^2_{y.x}\left[X'X+KI\right]^{-1}(X'X)\left[X'X+KI\right]$
๕) ความเอนเอียงกำลังสอง	$(BIAS)^2 = 0$	$(BIAS)^2= \beta'(Z-I)'(Z-I)\beta$

[๙]Donald W. Marquart and Ronald D. Snee, "Ridge Regression in Practice", *The American Statistician*, (Vol. 29, No. 1, February 1975) p. 6.