บทที่ 2

สถิติที่ใช้ในการวิจัย

ในบทนี้จะกล่าวถึงวิธีการประมาณสัมประสิทธิ์การถดถอยพหุทั้ง 3 วิธี ซึ่งได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method (OLS)) วิธีริดจ์รีเกรสชัน (Ridge Regression Method (RR)) และวิธีที่ใช้หลักการของริดจ์และสไตน์ (Ridge and Stein Method (RS)) ในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงแบบเบ้จะทำการแปลงข้อมูลด้วยการแปลงที่อยู่ในรูปยกกำลัง (Power Transformation) ของ Box และ Cox เพื่อให้ข้อมูลมีการแจกแจงเข้าสู่ภาวะปกติเสียก่อน ซึ่งรายละเอียดต่าง ๆ มีดังนี้

การประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธี OLS

ตัวแบบเชิงเส้นของความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามอยู่ในรูปของ

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

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$

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

จากตัวแบบในสมการ (2.1) เราประมาณค่าเวกเตอร์พารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยวิธี OLS ซึ่งจะได้ค่าประมาณคือ $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$ โดย $\hat{\underset{\sim}{\beta}}$ เป็นตัวประมาณที่ไม่เอนเอียงที่มีค่าเฉลี่ยความคลาดเคลื่อนต่ำที่สุดในบรรดาตัวประมาณที่ไม่เอนเอียง การประมาณค่า $\hat{\underset{\sim}{\beta}}$ ด้วยวิธี OLS มีข้อสมมติที่จำเป็นข้อหนึ่งคือ ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันซึ่งในทางปฏิบัติเป็นไปได้น้อยมาก ในกรณีที่ตัวแปรอิสระมีความสัมพันธ์กัน คือ มีสภาพที่ไม่เหมาะสม (ill-condition)

การประมาณค่า $\underset{\sim}{\beta}$ ด้วยวิธี OLS อาจจะไม่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด ซึ่งการพิจารณาผลของตัวแปรอิสระที่มีพหุสัมพันธ์กันสามารถพิจารณาจากคุณสมบัติ 2 ประการ คือ เมตริกซ์ความแปรปรวนและค่าความแปรปรวนร่วมของตัวประมาณ $\underset{\sim}{\hat{\beta}}$ และ ค่าเฉลี่ยของความแตกต่างกำลังสองระหว่าง $\underset{\sim}{\hat{\beta}}$ และ $\underset{\sim}{\beta}$

กำหนดให้ความแปรปรวนของเวกเตอร์ตัวประมาณ $\underset{\sim}{\hat{\beta}}$ มีค่าเป็น $\mathrm{cov}(\underset{\sim}{\hat{\beta}})$ จะได้ว่า

(2.2) $$\mathrm{cov}(\underset{\sim}{\hat{\beta}}) = \sigma^2(X'X)^{-1}$$

และให้ L_1 คือความแตกต่างระหว่าง $\underset{\sim}{\hat{\beta}}$ กับ $\underset{\sim}{\beta}$ ดังนั้นค่าความแตกต่างกำลังสองระหว่าง $\underset{\sim}{\hat{\beta}}$ และ $\underset{\sim}{\beta}$ มีค่าเป็น

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

ค่าเฉลี่ยความแตกต่างกำลังสองระหว่าง $\underset{\sim}{\hat{\beta}}$ และ $\underset{\sim}{\beta}$ คือ $E(L_1^2)$ ซึ่งมีค่าดังนี้

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

โดยที่ $E(L_1^2)$ จะมีค่าสมมูล (equivalent) กับ $E[(\underset{\sim}{\hat{\beta}} - \underset{\sim}{\beta})'(\underset{\sim}{\hat{\beta}} - \underset{\sim}{\beta})]$ ซึ่งทำให้เราสามารถเขียนได้ว่า

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

$$= E[\underset{\sim}{\hat{\beta}}'\underset{\sim}{\hat{\beta}} - 2\underset{\sim}{\hat{\beta}}'\underset{\sim}{\beta} + \underset{\sim}{\beta}'\underset{\sim}{\beta}]$$

$$= E(\underset{\sim}{\hat{\beta}}'\underset{\sim}{\hat{\beta}}) - \underset{\sim}{\beta}'\underset{\sim}{\beta}$$

เมื่อแทนค่า $E(L_1^2)$ ในสมการ (2.3) จะสามารถหาค่า $E(\underset{\sim}{\hat{\beta}}'\underset{\sim}{\hat{\beta}})$ ได้คือ

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

ถ้า $\underset{\sim}{\varepsilon}$ มีการแจกแจงปกติจะได้ว่า

$$\text{(2.4)} \qquad \mathrm{Var}(L_1^2) = 2\sigma^4 \,\mathrm{trace}(X'X)^{-2}$$

เราเห็นได้ว่าจากสมการ (2.3) และ (2.4) จะอยู่ในรูปฟังก์ชันซึ่งเป็นผลบวกของสมาชิกในแนวทแยงมุมของเมตริกซ์ $X'X$ ($\mathrm{trace}(X'X)$) ดังนั้นเพื่อความสะดวกในการเปรียบเทียบการประมาณค่า $\underset{\sim}{\hat{\beta}}$ จึงควรแปลงให้อยู่ในรูปฟังก์ชันของค่าเจาะจง (eigenvalue) ของเมตริกซ์ $X'X$ โดยใช้คุณสมบัติที่สำคัญข้อหนึ่งของค่าเจาะจง กล่าวคือ ถ้า λ_i เป็นค่าเจาะจงของเมตริกซ์ $X'X$ เมื่อ $i = 1,2,3,\ldots, p$ แล้ว $\sum_{i=1}^{p}\lambda_i = \mathrm{trace}(X'X)$

สมมติให้ค่า λ_i เป็นค่าเจาะจงของเมตริกซ์ $X'X$ ซึ่ง λ_{max} คือค่าเจาะจงที่มีค่ามากที่สุดและ λ_{min} คือค่าเจาะจงที่มีค่าน้อยที่สุด ดังนั้น $(\lambda_{max} = \lambda_1) \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq (\lambda_p = \lambda_{min}) > 0$

จากสมการ (2.3) และ (2.4) เราสามารถเขียนในรูปฟังก์ชันของค่าเจาะจง ได้ดังนี้

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

$$\text{และ} \quad \mathrm{Var}(L_1^2) = 2\sigma^4 \sum_{i=1}^{p} \left(\frac{1}{\lambda_i}\right)^2$$

ดังนั้นในกรณีที่ตัวแปรอิสระมีสภาพที่ไม่เหมาะสม กล่าวคือ ค่าเจาะจงบางค่าของเมตริกซ์ $X'X$ จะมีค่าน้อยมากๆ ซึ่งมีผลทำให้ความแตกต่างระหว่าง $\underset{\sim}{\hat{\beta}}$ และ $\underset{\sim}{\beta}$ มีค่ามาก นั่นคือ การประมาณค่า $\underset{\sim}{\beta}$ ด้วยวิธี OLS จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่มีค่าสูงขึ้น ดังนั้นในกรณีที่เกิดปัญหาพหุสัมพันธ์ตัวประมาณ $\underset{\sim}{\hat{\beta}}$ จึงเป็นตัวประมาณที่ไม่เหมาะสม

การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธี RR

Hoerl and Kennard (1970:55-67) ได้เสนอวิธี RR เพื่อแก้ปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระซึ่งหลักการของตัวประมาณนี้คือ พยายามที่จะลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณ $\underset{\sim}{\hat{\beta}}$ ให้ต่ำลง เมื่อพิจารณาจากค่าเจาะจงของเมตริกซ์ $X'X$ พบว่าใน

กรณีที่เกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ ค่าเจาะจงบางค่าจะมีค่าน้อยมากๆ จึงทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณ $\hat{\underset{\sim}{\beta}}$ มีค่าสูงขึ้น วิธี RR จึงแก้ปัญหานี้โดยการบวกค่าคงที่ที่มากกว่าศูนย์กับสมาชิกทุกตัวบนเส้นทแยงมุมของเมตริกซ์ $X'X$ เพื่อจะทำให้ได้ค่าเจาะจงที่สูงขึ้น โดยที่เราสามารถเขียนสมการปกติของตัวประมาณด้วยวิธี RR ได้เป็น

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

ดังนั้นตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ จะอยู่ในรูปของ

$$\hat{\underset{\sim}{\beta}}_R = (X'X+kI)^{-1}\ X'\underset{\sim}{y} \qquad ; \quad k>0 \tag{2.5}$$

จากรูปแบบของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ ในสมการ (2.5) เราสามารถแปลงให้อยู่ในรูปฟังก์ชันของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ โดยกำหนดให้ $(X'X+kI)^{-1} = W$

$$\therefore \quad \hat{\underset{\sim}{\beta}}_R = W\,X'\underset{\sim}{y}$$

และจากสมการปกติ $X'X\underset{\sim}{\beta} = X'\underset{\sim}{y}$ เราจะสามารถเขียนตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ ได้ดังนี้

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

โดยที่ $Z=[\ I+k(X'X)^{-1}\ ]^{-1}$ ซึ่งค่า $Z, \hat{\underset{\sim}{\beta}}_R$ และ W จะนำไปพิจารณาเปรียบเทียบคุณสมบัติต่างๆ ของการประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ตลอดจนการพิสูจน์ทฤษฎีต่างๆ ที่สำคัญของการประมาณค่า $\hat{\underset{\sim}{\beta}}_R$ ต่อไป คุณสมบัติที่จำเป็นของ $Z, \hat{\underset{\sim}{\beta}}_R$ และ W มีดังนี้

ก) ให้ $\xi_i(W)$ และ $\xi_i(Z)$ เป็นค่าเจาะจงของ W และ Z ตามลำดับ ซึ่งคำนวณได้จากสมการเจาะจง (characteristic equation)

$$|\mathbf{W}-\xi_i \mathbf{I}| = 0$$

และ $$|\mathbf{Z}-\xi_i \mathbf{I}| = 0$$

จะได้ค่าเจาะจงของ $\mathbf{W}$ และ $\mathbf{Z}$ คือ

$$\xi_i(\mathbf{W}) = \frac{1}{\lambda_i + k}$$

และ $$\xi_i(\mathbf{Z}) = \frac{\lambda_i}{\lambda_i + k}$$

ข) ค่า $\mathbf{Z}$ อาจเขียนในรูปฟังก์ชันของ $\mathbf{W}$ ได้ดังนี้

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

ค) ค่าตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ จะมีค่าน้อยกว่า $\hat{\underset{\sim}{\beta}}$ เมื่อ $k > 0$ กล่าวคือ

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

เราพิจารณาตัวประมาณ $\hat{\underset{\sim}{\beta}}_R = \mathbf{Z}\hat{\underset{\sim}{\beta}}$, เมตริกซ์ $\mathbf{X}'\mathbf{X}$ และ $\mathbf{Z}$ ที่สมมาตร ซึ่งมีคุณสมบัติเป็นบวกแน่นอน (Symmetric Positive Definite) จะได้ว่า

$$\begin{aligned} (\hat{\underset{\sim}{\beta}}_R)'(\hat{\underset{\sim}{\beta}}_R) &= (\mathbf{Z}\hat{\underset{\sim}{\beta}})'(\mathbf{Z}\hat{\underset{\sim}{\beta}}) \\ &= \sum_{i=1}^{p} \xi_i^2(\mathbf{Z})\, \hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}} \\ &\le \xi^2_{\max}(\mathbf{Z})\, \hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}} \end{aligned}$$

โดยที่ $\xi_{\max}(\mathbf{Z}) = \frac{\lambda_{\max}}{\lambda_{\max} + k}$ ซึ่ง $\lambda_{\max}$ เป็นค่าเจาะจงที่มีค่ามากที่สุดของเมตริกซ์ $\mathbf{X}'\mathbf{X}$ จากคุณสมบัติของ $\mathbf{Z}$ ในข้อ ข) พบว่า $\mathbf{Z}=\mathbf{I}$ ณ จุดที่ค่า k เท่ากับศูนย์และ $\mathbf{Z}$ มีค่าเข้าใกล้ศูนย์ เมื่อค่า $k \to \infty$ ดังนั้นคุณสมบัติในข้อ ค) จึงเป็นจริง

สำหรับผลบวกของความคลาดเคลื่อนกำลังสองของค่าตัวประมาณ $\underline{\hat{\beta}}_R$ มีค่าเป็น

$$\phi(\underline{\hat{\beta}}_R) = (\underline{y} - X\underline{\hat{\beta}}_R)'(\underline{y} - X\underline{\hat{\beta}}_R)$$

และจากการประมาณค่าพารามิเตอร์ $\underline{\beta}$ ด้วยวิธี OLS เราสามารถเขียน $\phi(\underline{\hat{\beta}})$ ซึ่งเป็นผลบวกของความคลาดเคลื่อนกำลังสองของตัวประมาณ $\underline{\hat{\beta}}$ ดังสมการข้างล่างนี้

$$\phi(\underline{\hat{\beta}}) = \underline{y}'\underline{y} - \underline{\hat{\beta}}'X'\underline{y} \tag{2.6}$$

ดังนั้นเพื่อให้เห็นความแตกต่างของ $\phi(\underline{\hat{\beta}}_R)$ และ $\phi(\underline{\hat{\beta}})$ เราสามารถเขียน $\phi(\underline{\hat{\beta}}_R)$ ในเทอมของ $\underline{y}'\underline{y}$ และ $X'\underline{y}$ ได้ดังนี้

$$\phi(\underline{\hat{\beta}}_R) = \underline{y}'\underline{y} - (\underline{\hat{\beta}})'X'\underline{y} - k\,\underline{\hat{\beta}}_R'\,\underline{\hat{\beta}}_R \tag{2.7}$$

จากสมการ (2.6) และ (2.7) เราสามารถสรุปได้ว่าผลบวกของความคลาดเคลื่อนกำลังสองของวิธี RR ให้ค่าน้อยกว่าผลบวกของความคลาดเคลื่อนกำลังสองที่คำนวณด้วยวิธี OLS

1. คุณสมบัติของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธี RR

การประมาณค่าสัมประสิทธิ์การถดถอยพหุ $\underline{\beta}$ ด้วยวิธีกำลังสองน้อยที่สุดจะได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองดังนี้

$$\begin{aligned} MSE(\underline{\hat{\beta}}) &= cov(\underline{\hat{\beta}}) \\ &\equiv \hat{\sigma}^2 (X'X)^{-1} \end{aligned}$$

เนื่องจาก $E(\underline{\hat{\beta}}) = \underline{\beta}$ ดังนั้นการประมาณค่าพารามิเตอร์ $\underline{\beta}$ ด้วยวิธี OLS จะได้ตัวประมาณค่า $\underline{\beta}$ ที่ไม่เอนเอียง แต่ในการประมาณค่าพารามิเตอร์ด้วยวิธี RR จะได้ตัวประ-

มาณที่เอนเอียง กล่าวคือ $E(\hat{\underset{\sim}{\beta}}_R) = Z\underset{\sim}{\beta}$ ซึ่งค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\beta}$ ด้วยวิธี RR จะอยู่ในรูปของ

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

$$= E[(\hat{\underset{\sim}{\beta}}_R - \underset{\sim}{\beta})' Z'Z (\hat{\underset{\sim}{\beta}}_R - \underset{\sim}{\beta})]$$

$$+ [(Z\underset{\sim}{\beta} - \underset{\sim}{\beta})'(Z\underset{\sim}{\beta} - \underset{\sim}{\beta})]$$

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

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

$$+ k^2 \underset{\sim}{\beta}'(X'X+kI)^{-2}\underset{\sim}{\beta}$$

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

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

ผู้อ่านจะเห็นได้ว่า $E[L_1^2(k)]$ เป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองซึ่งอยู่ในรูปผลบวกของฟังก์ชัน $\gamma_1(k)$ และ $\gamma_2(k)$ เมื่อ $\gamma_1(k)$ เป็นค่าความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ $(\text{cov}(\hat{\underset{\sim}{\beta}}_R(k)))$ และ $\gamma_2(k)$ เป็นระยะทางจาก $Z\underset{\sim}{\beta}$ ไป $\underset{\sim}{\beta}$

เมื่อ $k = 0$ ค่า $\gamma_2(k)$ จะมีค่าเป็นศูนย์เพราะได้ Z เป็นเมตริกซ์เอกลักษณ์การคูณ ซึ่งสอดคล้องกับค่าความเอนเอียงกำลังสองของ $\hat{\underset{\sim}{\beta}}$ ที่มีค่าศูนย์ สำหรับกรณีที่ $k > 0$ เราอาจพิจารณา $\gamma_2(k)$ ในเทอมของความเอนเอียงของ $\hat{\underset{\sim}{\beta}}$ กำลังสอง กล่าวคือ $\gamma_2(k)$ เท่ากับ $(\text{bias}(\hat{\underset{\sim}{\beta}}_R))^2$

เราสามารถเขียนค่า $\hat{\underset{\sim}{\beta}}_R$ ในเทอมของ $\underset{\sim}{y}$ ได้ดังนี้

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

ดังนั้นค่าความแปรปรวนร่วมของ $\hat{\underset{\sim}{\beta}}_R$ เขียนได้เป็น

$$\mathrm{cov}(\hat{\underset{\sim}{\beta}}_R) = Z(X'X)^{-1}X'\,\mathrm{cov}(\underset{\sim}{y})\,X(X'X)^{-1}Z'$$

(2.8)
$$= \sigma^2\ \mathrm{trace}\ Z(X'X)^{-1}Z'$$

จากความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ ในสมการ (2.8) เราจะเขียนให้อยู่ในรูปฟังก์ชันของค่าเจาะจงเพื่อให้สอดคล้องกับเทอม $\gamma_1(k)$ โดยกำหนดให้ λ_i เป็นค่าเจาะจงของเมตริกซ์ $X'X$ และ Z เมื่อ i = 1,2,...,p ซึ่งจากคุณสมบัติของค่าเจาะจงทำให้เราสามารถเขียนผลบวกของสมาชิกในแนวทแยงมุมของเมตริกซ์ $(X'X)^{-1}$ และ Z ได้ดังนี้

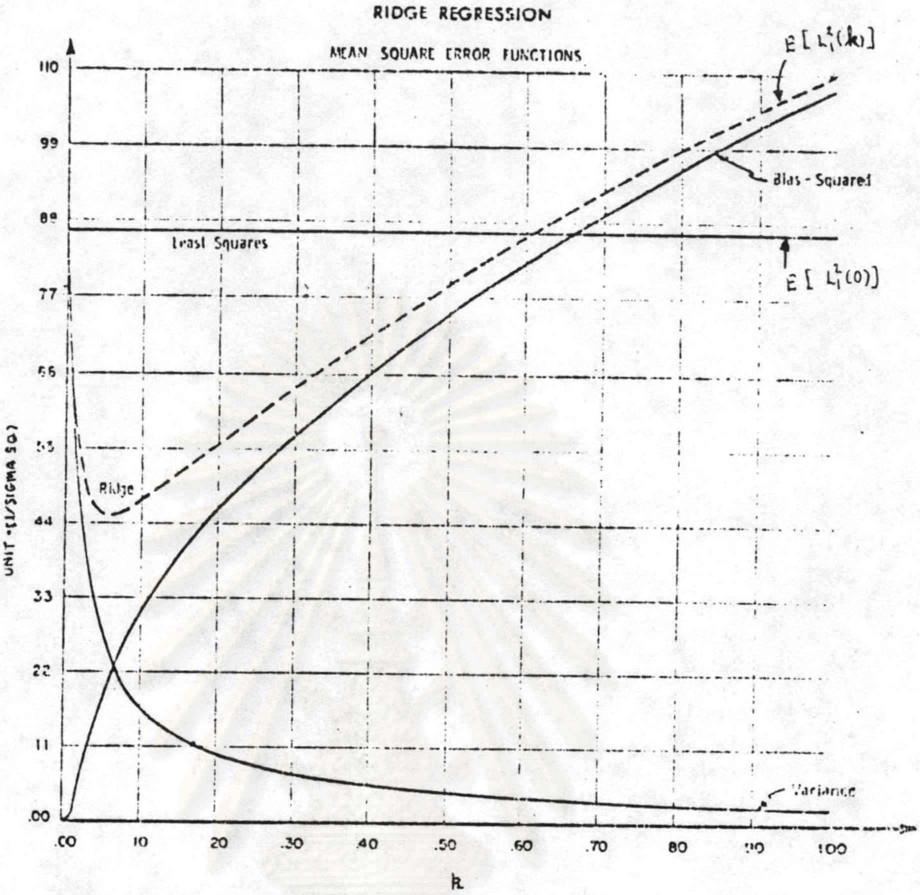
$$\mathrm{trace}(X'X)^{-1} = \sum_{i=1}^{p}\frac{1}{\lambda_i}$$
$$\mathrm{trace}(Z) = \sum_{i=1}^{p}\frac{\lambda_i}{\lambda_i + k}$$

และจากสมการ (2.8) จะได้ว่า

$$\sigma^2\ \mathrm{trace}\ Z(X'X)^{-1}Z' = \sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + k)^2}$$

ดังนั้น $E[L_1^2(k)] = \mathrm{cov}(\hat{\underset{\sim}{\beta}}_R(k)) + \left\{\mathrm{bias}(\hat{\underset{\sim}{\beta}}_R(k))\right\}^2$

เราจะพิจารณาคุณสมบัติของความสัมพันธ์ระหว่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ กับค่าพารามิเตอร์ k จากรูปที่ 2.1 เพื่อที่จะนำไปใช้ในการพิจารณาหาค่าพารามิเตอร์ k ในหัวข้อที่ 2 ต่อไป



รูปที่ 2.1

แสดงคุณสมบัติของรูปแบบความสัมพันธ์ระหว่างความแปรปรวนร่วม $\mathrm{cov}(\hat{\underset{\sim}{\beta}}_R(k))$

ความเอนเอียงกำลังสอง $\left(\mathrm{bias}(\hat{\underset{\sim}{\beta}}_R(k))\right)^2$ และค่าพารามิเตอร์ k

รูปที่ 2.1 เป็นกราฟแสดงความสัมพันธ์ระหว่างความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ กับค่าพารามิเตอร์ k และความเอนเอียงกำลังสองกับค่าพารามิเตอร์ k ซึ่งความสัมพันธ์ระหว่างความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ กับค่าพารามิเตอร์ k คือ เมื่อ k มีค่าเพิ่มขึ้นค่าความแปรปรวนร่วมจะมีค่าลดลงเรื่อยๆ กล่าวคือ ค่าความแปรปรวนร่วมแปรผกผันกับค่าพารามิเตอร์ k แต่ความสัมพันธ์ระหว่างความเอนเอียงกำลังสองของตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ กับค่าพารามิเตอร์ k จะมีลักษณะตรงกันข้าม นั่นคือค่าความเอนเอียงแปรผันตามค่าพารามิเตอร์ k สำหรับรูปเส้นไข่ปลาเป็นกราฟแสดงค่า $E[L_1^2(k)]$ ซึ่งเป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดจากตัวประมาณ $\hat{\underset{\sim}{\beta}}_R$ เมื่อ $k>0$ และเมื่อ $k=0$ จะแสดงด้วยค่า $E[L_1^2(0)]$ ซึ่งคือตัวประมาณสัมประสิทธิ์การถดถอยพหุด้วยวิธี OLS

เนื่องจาก $\gamma_1(\mathbf{k})$ เป็นค่าความแปรปรวนร่วมของตัวประมาณ $\hat{\beta}_R$ ดังนั้นฟังก์ชัน $\gamma_1(\mathbf{k})$ จะมีลักษณะเป็นฟังก์ชันที่ลดลงทางเดียว (monotonically decreasing function) ในรูปฟังก์ชันของ k และ $\gamma_2(\mathbf{k})$ เป็นค่าความเอนเอียงกำลังสองของตัวประมาณ $\hat{\beta}_R$ จะได้ว่าฟังก์ชัน $\gamma_2(\mathbf{k})$ มีลักษณะเป็นฟังก์ชันที่เพิ่มขึ้นทางเดียว (monotonically increasing function)ในรูปฟังก์ชันของ k จากรูปเราจะพบว่า $E[L_1^2(\mathbf{k})]$ มีค่าน้อยกว่า $E[L_1^2(0)]$ ณ บางค่าของ k ดังนั้นถ้าเราสามารถเลือก k ซึ่งทำให้ $E[L_1^2(\mathbf{k})]$ น้อยกว่า $E[L_1^2(0)]$ จะทำให้ได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุจากวิธี RR มีค่าน้อยกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุจากวิธี OLS

เราสามารถหาค่าจำกัดของความแปรปรวนร่วมและความเอนเอียงกำลังสองของตัวประมาณ $\hat{\beta}_R$ โดยใช้ฟังก์ชันของ $\gamma_1(\mathbf{k})$ และ$\gamma_2(\mathbf{k})$ ซึ่งค่าจำกัดของความแปรปรวนร่วมและค่าความเอนเอียงกำลังสองหาได้จากอนุพันธ์ของ $\gamma_1(\mathbf{k})$ เทียบกับ k และ $\gamma_2(\mathbf{k})$ เทียบกับ k ตามลำดับ สำหรับค่าจำกัดของแต่ละฟังก์ชันมีดังนี้

(2.9) $$\lim_{k\to 0^+}\left(\frac{\partial \gamma_1}{\partial k}\right) = -2\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i^2}$$

(2.10) $$\lim_{k\to 0^+}\left(\frac{\partial \gamma_2}{\partial k}\right) = 0$$

ดังนั้นอนุพันธ์ของความแปรปรวนร่วมของตัวประมาณ $\hat{\beta}_R$ จะมีค่าเป็นลบและลู่เข้าสู่ $-2\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i^2}$ เมื่อ $k \to 0^+$ โดยที่ผลคูณของเมตริกซ์ของตัวแปรอิสระ $X'X$ มีลักษณะเชิงตั้งฉาก กล่าวคือ ตัวแปรอิสระ X จะต้องไม่มีความสัมพันธ์กัน และอนุพันธ์ของความแปรปรวนร่วมของตัวประมาณ $\hat{\beta}_R$ จะมีค่าเข้าสู่ $-\infty$ เมื่อเมตริกซ์ของตัวแปรอิสระ $X'X$ มีสภาพไม่เหมาะสม ส่วนอนุพันธ์ของความเอนเอียงกำลังสองมีค่าลดลงและจะเป็นศูนย์ ณ จุดกำเนิด เมื่อ $k \to 0^+$

คุณสมบัติของสมการ (2.9) และ (2.10) จะเป็นจริงเมื่อ $k > 0$ ซึ่งสรุปได้ว่าจะมีความเอนเอียงเล็กน้อย และค่าความแปรปรวนร่วมจะมีค่าลดลงมาก จึงส่งผลให้ได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่มีค่าลดลงและมีค่าน้อยกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากวิธี OLS

2. การประมาณค่าพารามิเตอร์ k

วิธีในการประมาณค่า k มีหลายวิธีการเช่น วิธีของ Hoerl Kennard and Baldwin (1975) วิธีของ TZE-SAN-LEE และวิธีของ McDonald Galarneau (1975) ซึ่งจากผลงานวิจัยของเจษฎาพร ยุทธนพิบูลย์ชัย(พ.ศ. 2533) ไม่สามารถสรุปได้แน่นอนว่าวิธีการประมาณค่า k วิธีการใดที่ให้ผลสรุปชัดเจน แต่จากผลงานวิจัยของจิรายุส พู่มนตรี(พ.ศ.2534) ได้มีการเสนอวิธีการประมาณค่า k โดยใช้วิธีการ Binary Search ซึ่งเป็นวิธีการหนึ่งในทางคอมพิวเตอร์ที่ใช้ในการค้นหาข้อมูลและผลสรุปที่ได้คือ วิธีการ Binary Search จะให้ผลในการประมาณค่า k ดีที่สุด แต่วิธีการนี้มีข้อจำกัดคือ ใช้กับข้อมูลที่มีลักษณะแบบไม่ต่อเนื่องและใช้ในการค้นหาข้อมูลที่มีลักษณะเรียงลำดับ ดังนั้นจึงต้องมีการเรียงข้อมูลจากน้อยไปหามากหรือจากมากไปหาน้อย จากข้อจำกัดดังกล่าวจึงทำให้วิธีการแบบ Binary Search ไม่เหมาะสมที่จะนำมาใช้ ผู้วิจัยจึงพยายามหาวิธีการค้นหาข้อมูลอื่นๆ ซึ่งพบว่าวิธีการค้นหาข้อมูลทางคอมพิวเตอร์ที่สามารถนำมาประยุกต์ใช้ได้คือ วิธีการค้นหาข้อมูลแบบลำดับ (Sequential Search) เพราะวิธีการนี้สามารถใช้ได้กับข้อมูลที่มีลักษณะทั้งแบบต่อเนื่องและแบบไม่ต่อเนื่องและข้อมูลที่จะค้นหาจะเรียงลำดับหรือไม่เรียงลำดับก็ได้ แต่ข้อเสียของวิธีการนี้คือ เราอาจใช้เวลาในการค้นหานานเมื่อข้อมูลที่ต้องการค้นหาอยู่ ณ ลำดับไกลๆ ซึ่งเมื่อพิจารณาจากคุณสมบัติของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุดังที่กล่าวมาในหัวข้อที่ 1 ค่า k ที่เหมาะสมจะมีค่าไม่ห่างจากศูนย์มากนัก และในทางปฏิบัติค่าของ k จะอยู่ในช่วง (0 , 1) จากคุณสมบัติของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุลักษณะกราฟที่ได้จะมีจุดต่ำสุดเพียงจุดเดียวและไม่มีจุดวกกลับ ดังนั้นการพิจารณาที่จะใช้วิธีการค้นหาแบบลำดับจึงมีความเหมาะสมมากกว่า ผู้วิจัยจึงได้ใช้วิธีการในการประมาณค่า k คือวิธีการค้นหาแบบลำดับ (Sequential Search) โดยมีหลักการดังนี้

กำหนดให้ k_{opt} คือ ค่า k ที่ทำให้ $MSE(\hat{\underset{\sim}{\beta}}_R(k))$ มีค่าต่ำสุด

k_{int} คือ ค่า k ณ จุดเริ่มต้นในการค้นหาข้อมูล

k_{01} คือ ค่า k ณ จุดที่มีการเว้นช่วงห่างของค่า k เท่ากับ 0.01

k_{BEF01} คือ ค่า k ณ จุดที่อยู่ก่อนค่า k ที่เหมาะสมที่มีการเว้นช่วงห่างของค่า k เท่ากับ 0.01

k_{AFT01} คือ ค่า k ณ จุดที่อยู่หลังค่า k ที่เหมาะสมที่มีการเว้นช่วงห่างของค่า k เท่ากับ 0.01

k_{001} คือ ค่า k ณ จุดที่มีการเว้นช่วงห่างของค่า k เท่ากับ 0.001

k_{STOP} คือ ค่า k ที่ใช้ตรวจสอบหยุดการค้นหาค่า k ที่เหมาะสม เมื่อทำการค้นหาข้อมูลทางด้านขวาของค่า k ที่เหมาะสม

d คือ ค่าคงที่ที่กำหนดขึ้นเพื่อเป็นช่วงห่างของค่า k

<u>ขั้นที่ 1</u> กำหนดให้ค่า $k_{int} = 0.0$ และ ค่า $d = 0.01$

<u>ขั้นที่ 2</u> คำนวณค่า $k_{01} = k_{int} + d$

<u>ขั้นที่ 3</u> คำนวณหา $MSE(\hat{\underset{\sim}{\beta}}_R(k_{int}))$ และหา $MSE(\hat{\underset{\sim}{\beta}}_R(k_{01}))$ แล้วทำการเปรียบเทียบ

ก) ถ้า $MSE(\hat{\underset{\sim}{\beta}}_R(k_{01})) > MSE(\hat{\underset{\sim}{\beta}}_R(k_{int}))$

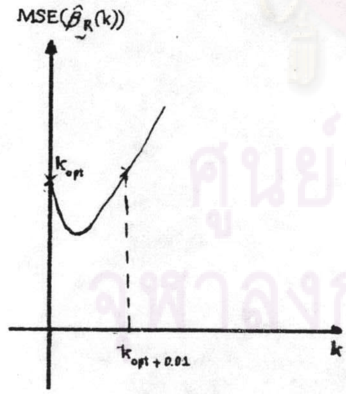
จะได้ว่า $k_{opt} = k_{int}$ ยุติการประมวลผล

ข) ถ้า $MSE(\hat{\underset{\sim}{\beta}}_R(k_{01})) \leq MSE(\hat{\underset{\sim}{\beta}}_R(k_{int}))$ เราจะกำหนดให้

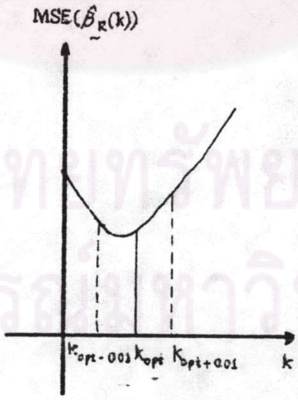
$$k_{int} = k_{01}$$

และทำการวิเคราะห์ขั้นที่ 2 ใหม่จนกว่าจะเข้ากรณี ก)

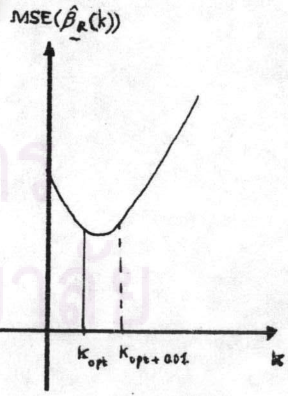
เนื่องจากลักษณะของ $MSE(\hat{\underset{\sim}{\beta}}_R(k))$ เป็นรูปโค้งหงาย การหาจุดที่ต่ำสุดจึงขึ้นอยู่กับการกำหนดช่วงห่างของค่า k ซึ่งจากการวิจัยพบว่าลักษณะความเป็นไปได้ของการจะพบจุดที่ต่ำสุดเมื่อช่วงห่างไม่ละเอียดพออาจแสดงได้ดังกรณีที่ 1-3



<u>กรณีที่ 1</u>



<u>กรณีที่ 2</u>



<u>กรณีที่ 3</u>

<u>กรณีที่ 1</u> ถ้าเรากำหนดช่วง d ห่างเกินไป จุดที่จะได้จุดต่ำสุดอาจจะเป็นจุดเริ่มแรกที่กำหนดคือ $k_{opt} = 0.0$ ซึ่งกรณีนี้ผู้วิจัยจะทำการแก้ปัญหาโดยให้มีการกำหนดช่วงห่าง

ของการค้นหาให้ละเอียดขึ้นแล้วทำการค้นหาข้อมูลแบบลำดับจากจุดเริ่มต้น $k_{opt} = 0.0$ จนถึงจุดที่ $k_{opt} = 0.0 + 0.01$ ใหม่อีกครั้งโดยใช้ช่วงห่างของการค้นหาเป็น $d = 0.001$

<u>กรณีที่ 2</u> ถ้าจุดของค่า k ที่ค้นหาได้ตกอยู่หลังค่า k ที่เหมาะสมจริงๆ จึงจะเกิดกรณีนี้ขึ้น ผู้วิจัยได้แก้ปัญหาโดยการเพิ่มช่วงห่างของการค้นหาให้ละเอียดขึ้นและจะค้นหาข้อมูลทางด้านซ้ายของค่า k_{opt} อีกครั้งโดยจะค้นหาข้อมูลแบบลำดับจากจุด $k_{opt} - 0.01$ ไปจนถึงจุด k_{opt} โดยใช้ช่วงห่างของการค้นหาเป็น $d = 0.001$

<u>กรณีที่ 3</u> ถ้าจุดของค่า k ที่ค้นหาได้ตกอยู่ก่อนหน้าค่า k ที่เหมาะสมจริงๆ จึงจะเกิดกรณีนี้ ผู้วิจัยได้แก้ปัญหาโดยการเพิ่มช่วงห่างของการค้นหาให้ละเอียดขึ้นและจะค้นหาข้อมูลทางด้านขวาของค่า k_{opt} อีกครั้งโดยจะค้นหาข้อมูลแบบลำดับจากจุด k_{opt} ไปจนถึงจุด $k_{opt} + 0.01$ โดยใช้ช่วงห่างของการค้นหาเป็น $d = 0.001$

ดังนั้นผู้วิจัยจึงจะทำการค้นหาค่า k ที่เหมาะสมอีกครั้งโดยกำหนดช่วงความห่างของการค้นหาใหม่ คือ ให้ $d = 0.001$ โดยจะทำการตรวจสอบก่อนว่าค่า k_{opt} ที่ได้มีค่าเท่ากับศูนย์หรือไม่ ถ้ามีค่าเท่ากับศูนย์ก็คือกรณีที่ 1 ผู้วิจัยจะทำการค้นหาข้อมูลแบบลำดับเฉพาะทางด้านขวาของค่า k_{opt} แต่ถ้าค่า k_{opt} มีค่าไม่เท่ากับศูนย์ ผู้วิจัยจะทำการค้นหาข้อมูลแบบลำดับทางด้านซ้ายของค่า k_{opt} ก่อนถ้าไม่พบก็จะทำการค้นหาทางด้านขวาของค่า k_{opt} อีกครั้ง ซึ่งขั้นตอนต่างๆ ในการค้นหามีดังนี้

ให้ $d = 0.001$

<u>ขั้นที่ 1</u> ตรวจสอบว่าค่า k_{opt} เท่ากับศูนย์หรือไม่

ก) ถ้า $k_{opt} = 0$ ให้ทำการวิเคราะห์ขั้นที่ 3 (ค้นหาทางด้านขวาของค่า k_{opt})

ข) ถ้า $k_{opt} \neq 0$ (มากกว่าศูนย์) ให้ทำการวิเคราะห์ขั้นที่ 2 (ค้นหาทางด้านซ้ายของค่า k_{opt})

<u>ขั้นที่ 2</u> ให้ $k_{BEF01} = k_{opt} - 0.01$

ก) ให้ $k_{001} = k_{BEF01} + d$

เปรียบเทียบค่า k_{001} กับ k_{opt}

ถ้า $k_{001} \geq k_{opt}$ ให้ทำการวิเคราะห์ขั้นที่ 3

ถ้า $k_{001} < k_{opt}$

คำนวณหา $MSE(\hat{\underset{\sim}{\beta}}_R(k_{BEF01}))$ และหา $MSE(\hat{\underset{\sim}{\beta}}_R(k_{001}))$ แล้วทำการเปรียบเทียบ

-ถ้า $\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{001})) > \mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{BEF01}))$

จะได้ว่า $k_{opt} = k_{BEF01}$

ยุติการประมวลผล

-ถ้า $\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{001})) \le \mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{BEF01}))$

เราจะกำหนด $k_{BEF01} = k_{001}$

แล้วทำการวิเคราะห์ข้อ ก) ในขั้นที่ 2 จนกว่าจะเข้ากรณีที่

$k_{001} \ge k_{opt}$ หรือ $\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{001})) > \mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{BEF01}))$

<u>ขั้นที่ 3</u> ให้ $k_{STOP} = k_{opt} + 0.01$

ก) ให้ $k_{001} = k_{opt}$

ข) ให้ $k_{AFT01} = k_{001} + d$

เปรียบเทียบค่า k_{AFT01} กับ k_{STOP}

ถ้า $k_{AFT01} \ge k_{STOP}$ ให้ยุติการประมวลผล

ถ้า $k_{AFT01} < k_{STOP}$

เราจะคำนวณหา $\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{AFT01}))$ และคำนวณหา

$\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{001}))$ แล้วทำการเปรียบเทียบ

-ถ้า $\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{AFT01})) > \mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{001}))$

จะได้ว่า $k_{opt} = k_{001}$ และยุติการประมวลผล

-ถ้า $\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{AFT01})) \le \mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{001}))$

เราจะกำหนด $k_{001} = k_{AFT01}$

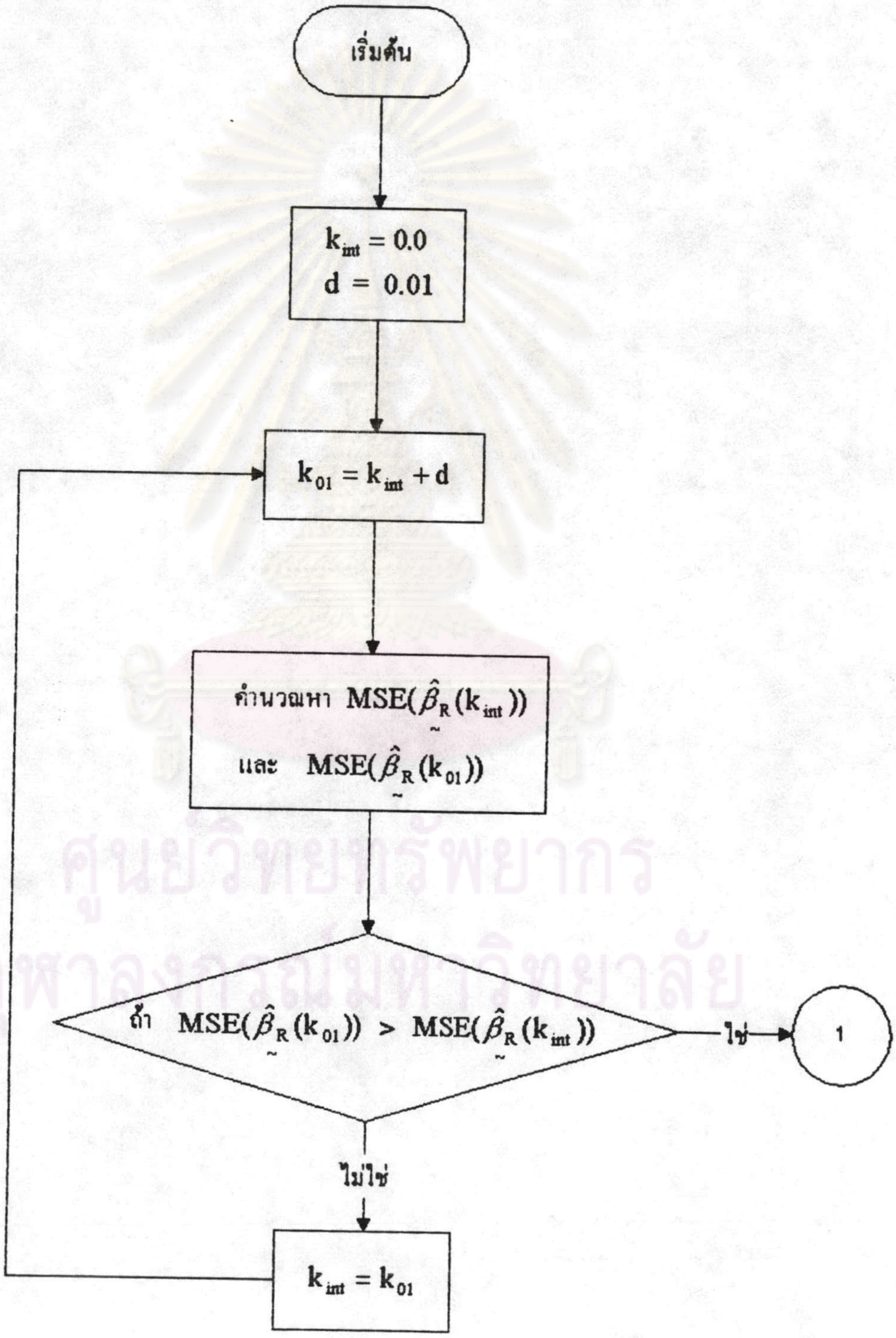
และทำการวิเคราะห์ข้อ ข) ในขั้นที่ 3 จนกว่าจะเข้ากรณีที่

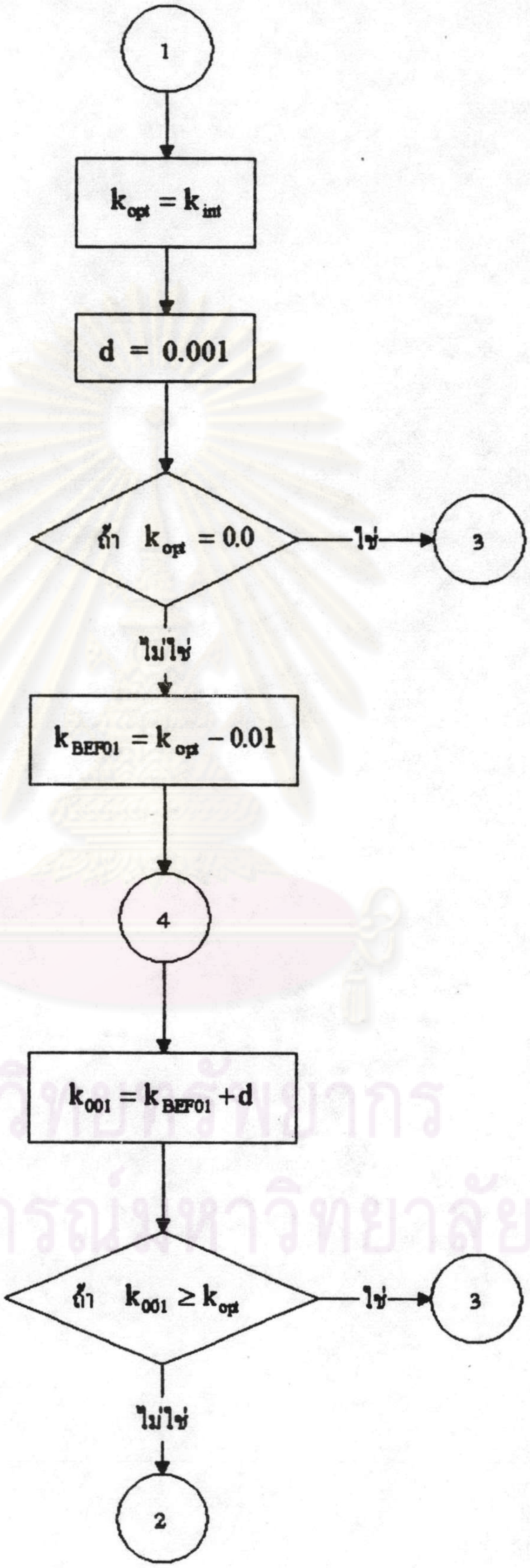
$k_{AFT01} \ge k_{STOP}$ หรือ

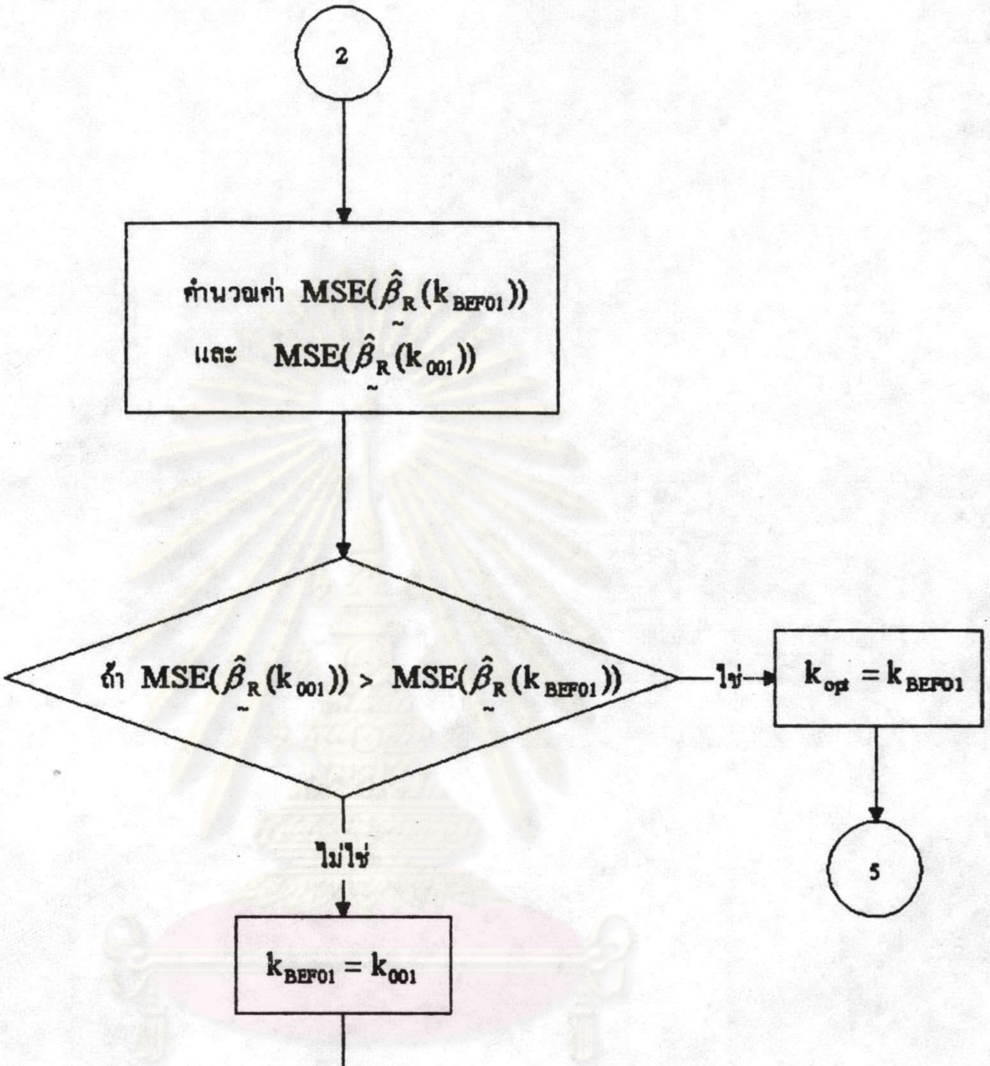
$\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{AFT01})) > \mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k_{001}))$

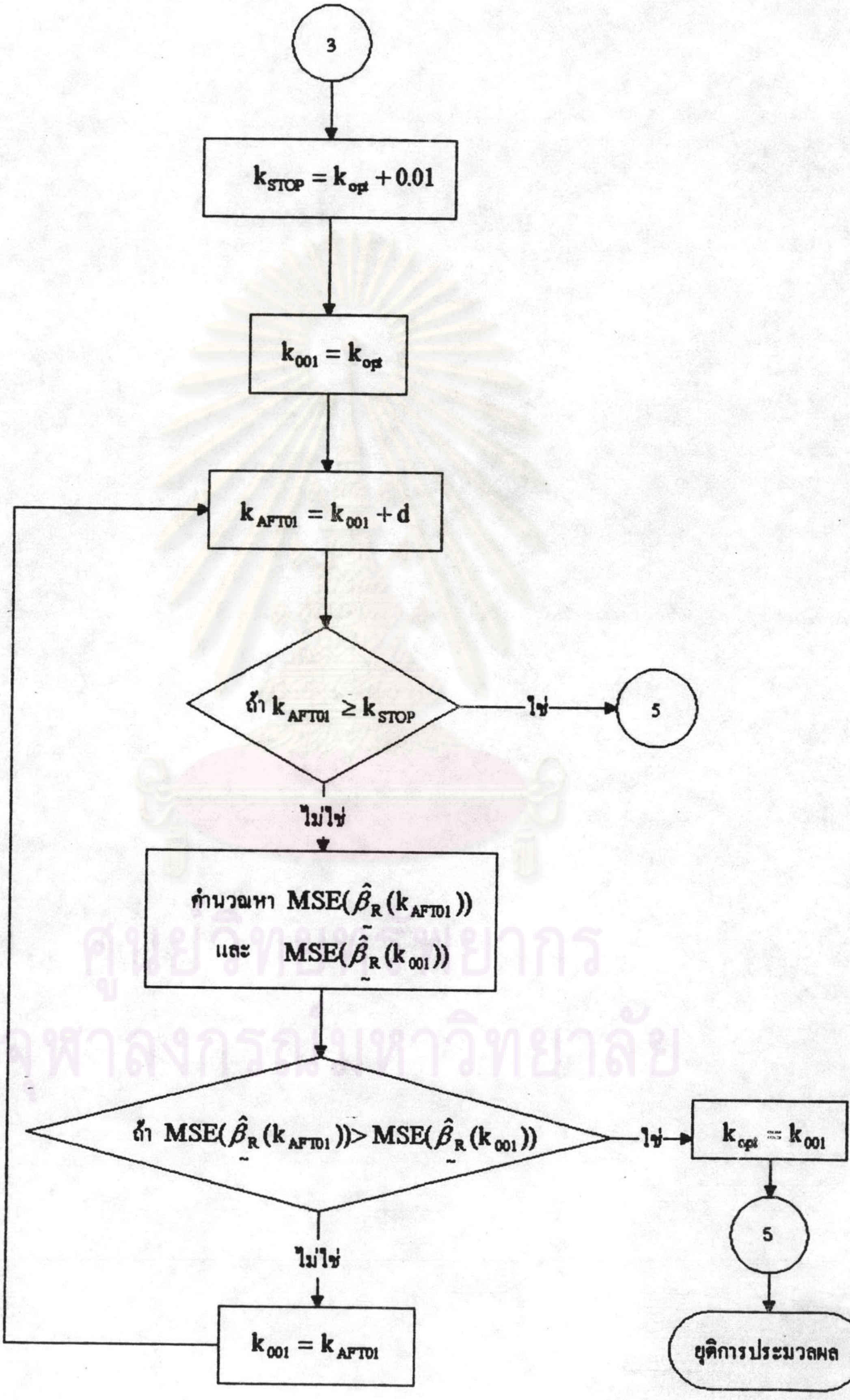
เนื่องจาก ณ จุดที่ค่า $d = 0.001$ และ $d = 0.0001$ จะให้ค่า $\mathrm{MSE}(\underset{\sim}{\hat{\beta}}_R(k))$ ที่มีค่าใกล้เคียงกัน ดังนั้นผู้วิจัยจึงไม่ทำการขยายช่วงค่า d จาก $d = 0.001$ เป็น $d = 0.0001$ สำหรับขั้นตอนต่างๆ ที่อธิบายมาแล้วสามารถเขียนเป็นผังงานเพื่อให้เห็นลักษณะการทำงานที่ชัดเจนได้ดังผังงานที่ 1

ผังงานที่ 1 แสดงการค้นหาค่า k โดยใช้วิธีการค้นหาแบบลำดับ (Sequential Search)




1
$k_{opt} = k_{int}$
$d = 0.001$
ถ้า $k_{opt} = 0.0$
ใช่
3
ไม่ใช่
$k_{BEF01} = k_{opt} - 0.01$
4
$k_{001} = k_{BEF01} + d$
ถ้า $k_{001} \ge k_{opt}$
ใช่
3
ไม่ใช่
2


2
คำนวณค่า $MSE(\hat{\underset{\sim}{\beta}}_R(k_{BEF01}))$
และ $MSE(\hat{\underset{\sim}{\beta}}_R(k_{001}))$
ถ้า $MSE(\hat{\underset{\sim}{\beta}}_R(k_{001})) > MSE(\hat{\underset{\sim}{\beta}}_R(k_{BEF01}))$
ใช่
$k_{opt} = k_{BEF01}$
5
ไม่ใช่
$k_{BEF01} = k_{001}$
4


3
$k_{STOP} = k_{opt} + 0.01$
$k_{001} = k_{opt}$
$k_{AFT01} = k_{001} + d$
ถ้า $k_{AFT01} \geq k_{STOP}$
ใช่
5
ไม่ใช่
คำนวณหา $MSE(\hat{\beta}_R(k_{AFT01}))$
และ $MSE(\hat{\beta}_R(k_{001}))$
ถ้า $MSE(\hat{\beta}_R(k_{AFT01})) > MSE(\hat{\beta}_R(k_{001}))$
ใช่
$k_{opt} = k_{001}$
5
ไม่ใช่
$k_{001} = k_{AFT01}$
ยุติการประมวลผล

การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธี RS

Liu Keijian (1993:343-402) ได้เสนอวิธี RS ในการแก้ปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระโดยใช้หลักการของตัวประมาณริดจ์ กล่าวคือการบวกค่าคงที่ค่าหนึ่งให้กับสมาชิกในแนวทแยงมุมของเมตริกซ์ $X'X$ เนื่องจากตัวประมาณริดจ์มีปัญหาในการประมาณค่าพารามิเตอร์ k ดังนั้นตัวประมาณที่ใช้หลักการของริดจ์และสไตน์จึงแก้ปัญหาดังกล่าวโดยการบวกค่าคงที่ k ที่มีค่าเท่ากับหนึ่งให้กับสมาชิกทุกตัวในแนวทแยงมุมของเมตริกซ์ $X'X$ เพราะ-ว่าตัวประมาณสไตน์เป็นตัวประมาณที่หาค่าพารามิเตอร์ง่ายเราจึงนำตัวประมาณดังกล่าวมาบวกกับเทอม $X'\underset{\sim}{y}$ เพราะฉะนั้นตัวประมาณอยู่ในรูปของ $\underset{\sim}{\hat{\beta}}_S = c\underset{\sim}{\hat{\beta}}$ เมื่อ $0 < c < 1$

การประมาณสัมประสิทธิ์การถดถอยพหุโดยวิธี RS จะได้ตัวประมาณที่อยู่ในรูปของ

(2.11) $$\underset{\sim}{\hat{\beta}}_C = (X'X+I)^{-1}(X'\underset{\sim}{y}+c\underset{\sim}{\hat{\beta}}) \quad , \quad 0 < c < 1$$

เราสามารถเขียน $\underset{\sim}{\hat{\beta}}_C$ ให้อยู่ในรูปฟังก์ชันของ $\underset{\sim}{\hat{\beta}}$ ซึ่งเราจะแทนค่า $X'\underset{\sim}{y}$ ที่ได้จากสมการปกติ $X'X\underset{\sim}{\hat{\beta}} = X'\underset{\sim}{y}$ ในสมการ (2.11) จะได้ว่าตัวประมาณดังกล่าวอยู่ในรูปของ

$$\begin{aligned}\underset{\sim}{\hat{\beta}}_C &= (X'X+I)^{-1}(X'X\underset{\sim}{\hat{\beta}}+c\underset{\sim}{\hat{\beta}}) \quad , \quad 0 < c < 1 \\ &= (X'X+I)^{-1}(X'X+cI)\underset{\sim}{\hat{\beta}} \\ &= C\underset{\sim}{\beta}\end{aligned}$$

เมื่อ $C = (X'X+I)^{-1}(X'X+cI)$ ถ้า $c = 1$ จะได้ $\underset{\sim}{\hat{\beta}}_C = \underset{\sim}{\hat{\beta}}$ และ $\|\underset{\sim}{\hat{\beta}}_C\| < \|\underset{\sim}{\hat{\beta}}\|$, $0 < c < 1$

เมื่อพิจารณาค่าคาดหวังของตัวประมาณที่ใช้หลักการของริดจ์และสไตน์จะได้ว่า

$$\begin{aligned}E(\underset{\sim}{\hat{\beta}}_C) &= E(C\underset{\sim}{\hat{\beta}}) \\ &= C\,E(\underset{\sim}{\hat{\beta}})\end{aligned}$$

$$= C\underset{\sim}{\beta}$$
$$= (X'X+I)^{-1}(X'X+cI)\underset{\sim}{\beta}$$

ความเอนเอียง(bias) ของ $\hat{\underset{\sim}{\beta}}_C = [E(\hat{\underset{\sim}{\beta}}_C) - \underset{\sim}{\beta}]$
$$= [C\underset{\sim}{\beta} - \underset{\sim}{\beta}]$$

และความเอนเอียงจะมีค่าเข้าใกล้ศูนย์เมื่อค่าของเมตริกซ์ C เข้าใกล้เมตริกซ์เอกลักษณ์การคูณ (Identity Matrix) ดังนั้นตัวประมาณที่ใช้หลักการของริดจ์และสไตน์จะเป็นตัวประมาณเอนเอียงที่หดตัว (bias shrinkage estimator)

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\hat{\underset{\sim}{\beta}}_C$ จะอยู่ในรูปของ

$$\mathrm{MSE}(\hat{\underset{\sim}{\beta}}_C) = E(\hat{\underset{\sim}{\beta}}_C - \underset{\sim}{\beta})^2$$

$$= \mathrm{trace}\{\mathrm{cov}(\hat{\underset{\sim}{\beta}}_C)\} + [E(\hat{\underset{\sim}{\beta}}_C) - \underset{\sim}{\beta}]^2$$

$$= \sigma^2 \mathrm{trace}(X'X+I)^{-1}(X'X+cI)(X'X)^{-1}(X'X+cI)(X'X+I)^{-1}$$

$$+ (c-1)^2 \underset{\sim}{\beta}'(X'X+I)^{-2}\underset{\sim}{\beta}$$

$$= \sigma^2 \sum_{i=1}^{p} \frac{(\lambda_i + c)^2}{\lambda_i(\lambda_i+1)^2} + (c-1)^2 \underset{\sim}{\beta}'(X'X+I)^{-2}\underset{\sim}{\beta}$$

$$= \sigma^2 \sum_{i=1}^{p} \frac{(\lambda_i + c)^2}{\lambda_i(\lambda_i+1)^2} + (c-1)^2 \sum_{i=1}^{p} \frac{\hat{\underset{\sim}{\beta}}^2}{(\lambda_i+1)^2}$$

จากรูปแบบของตัวประมาณที่ใช้หลักการของริดจ์และสไตน์พบว่าจะต้องทำการประมาณค่าพารามิเตอร์ c ซึ่งรายละเอียดดังกล่าวอยู่ในหัวข้อที่ 1

1. การประมาณค่าพารามิเตอร์ c

ตัวแบบในสมการ (2.1) สามารถเขียนให้อยู่ในรูปแบบใหม่โดยใช้เมตริกซ์เชิงตั้ง-ฉาก (Orthogonal Matrix) ได้ดังนี้

$$\underset{\sim}{y} = Z\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon} \tag{2.12}$$

เมื่อ $Z = XQ$, $\underset{\sim}{\alpha} = Q'\underset{\sim}{\beta}$ และ $Q'X'XQ = Z'Z = \Lambda$

และ X คือเมตริกซ์ของตัวแปรอิสระขนาด n x p

$\underset{\sim}{\beta}$ คือเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุขนาด p x 1

Q คือเมตริกซ์เชิงตั้งฉากขนาด p x p โดยที่แต่ละแนวตั้งของเมตริกซ์คือ เวกเตอร์เจาะจง (eigenvector) ที่สอดคล้องกับค่าเจาะจง (eigenvalue) $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p$ ตามลำดับ

Z คือเมตริกซ์ซึ่งแต่ละแนวตั้งของ Z คือ ตัวแปรอิสระชุดใหม่ซึ่งเป็นผลบวกเชิงเส้น (linear combination) ของตัวแปรอิสระ

Λ คือเมตริกซ์ทแยงมุมขนาด p x p โดยที่สมาชิกในแนวทแยงมุมเป็นค่าเจาะจงของเมตริกซ์ $X'X$

จากสมการ (2.12) เราสามารถใช้วิธี OLS เพื่อประมาณค่า $\underset{\sim}{\alpha}$ จะได้ว่า

$$\hat{\underset{\sim}{\alpha}} = (Z'Z)^{-1}Z'\underset{\sim}{y} = \Lambda^{-1}Z'\underset{\sim}{y}$$

เมตริกซ์ความแปรปรวนของ $\underset{\sim}{\alpha}$ มีค่าเป็น

$$\mathrm{Var}(\hat{\underset{\sim}{\alpha}}) = \sigma^2(Z'Z)^{-1} = \sigma^2\Lambda^{-1}$$

ในทำนองเดียวกันจะได้ว่าตัวประมาณริดจ์ คือ

$$\hat{\underset{\sim}{\alpha}}_R = (Z'Z+kI)^{-1}Z'\underset{\sim}{y}$$

และตัวประมาณที่ใช้หลักการของริดจ์และสไตน์ คือ

$$\underset{\sim}{\hat{\alpha}_C} = (Z'Z+I)^{-1}(Z'\underset{\sim}{y}+c\underset{\sim}{\hat{\alpha}})$$
$$= (\Lambda+I)^{-1}(Z'\underset{\sim}{y}+c\underset{\sim}{\hat{\alpha}})$$

พิจารณาคุณสมบัติ $\underset{\sim}{\alpha} = Q'\underset{\sim}{\beta}$ เราสามารถเขียนสมการได้เป็น $\underset{\sim}{\beta} = Q\underset{\sim}{\alpha}$ เนื่องจาก Q เป็นเมตริกซ์เชิงตั้งฉาก ดังนั้น $\underset{\sim}{\hat{\beta}}$ จะอยู่ในรูปของสมการข้างล่างนี้

(2.13) $$\underset{\sim}{\hat{\beta}} = Q\underset{\sim}{\hat{\alpha}}\ ,\ \underset{\sim}{\hat{\beta}_R} = Q\underset{\sim}{\hat{\alpha}_R} \text{ และ } \underset{\sim}{\hat{\beta}_C} = Q\underset{\sim}{\hat{\alpha}_C}$$

จากสมการ (2.13) แสดงว่าทุกๆ ค่าของ $\underset{\sim}{\hat{\alpha}}$ เราสามารถหาค่า $\underset{\sim}{\hat{\beta}}$ ได้ นั่นคือคุณสมบัติของ $\underset{\sim}{\hat{\beta}}$ จะสอดคล้องกับ $\underset{\sim}{\hat{\alpha}}$ ดังนั้นเมื่อเราศึกษาคุณสมบัติของ $\underset{\sim}{\hat{\alpha}}$ ก็จะสามารถทราบถึงคุณสมบัติของ $\underset{\sim}{\hat{\beta}}$ ได้เช่นเดียวกัน ในกรณีนี้จึงศึกษาคุณสมบัติของ $\underset{\sim}{\hat{\beta}_C}$ โดยอาศัยคุณสมบัติของ $\underset{\sim}{\hat{\alpha}_C}$

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\hat{\alpha}_C}$ คือ

$$\text{MSE}(\underset{\sim}{\hat{\alpha}_C}) = \text{trace}\left\{\text{cov}(\underset{\sim}{\hat{\alpha}_C})\right\} + (\text{bias})^2$$

ซึ่ง $\text{cov}(\underset{\sim}{\hat{\alpha}_C}) = \sigma^2(\Lambda+I)^{-1}(\Lambda+cI)\Lambda^{-1}(\Lambda+cI)(\Lambda+I)^{-1}$

และความเอนเอียงของ $\underset{\sim}{\hat{\alpha}_C} = E(\underset{\sim}{\hat{\alpha}_C}) - \underset{\sim}{\alpha}$

จะได้ว่า

$$\text{MSE}(\underset{\sim}{\hat{\alpha}_C}) = \sigma^2\sum_{i=1}^{p}\frac{(\lambda_i+c)^2}{\lambda_i(\lambda_i+1)^2} + (c-1)^2\sum_{i=1}^{p}\frac{\underset{\sim}{\alpha_i}^2}{(\underset{\sim}{\alpha_i}+1)^2}$$

เราสามารถหาค่าพารามิเตอร์ c โดยการหาอนุพันธ์อันดับที่ 1 ของ $\text{MSE}(\underset{\sim}{\hat{\alpha}_C})$ เทียบกับค่า c ดังนั้นค่า c มีค่าเท่ากับ

$$(2.14) \qquad c = \sum_{i=1}^{p} \frac{\underset{\sim}{\alpha_i}^2 - \sigma^2}{(\lambda_i + 1)^2} \Bigg/ \sum_{i=1}^{p} \frac{\sigma^2 + \lambda_i \ \underset{\sim}{\alpha_i}^2}{\lambda_i(\lambda_i + 1)^2}$$

เมื่อแทนค่าตัวประมาณที่ไม่เอนเอียงของ $\underset{\sim}{\alpha}^2$ และ σ^2 ซึ่งมีค่าเท่ากับ $\underset{\sim}{\hat{\alpha}}^2 - \left(\frac{\hat{\sigma}^2}{\lambda_i}\right)$ และ $\hat{\sigma}^2$ ตามลำดับ ลงในสมการ (2.14) จะได้ว่าค่า c มีค่าเท่ากับ

$$(2.15) \qquad \hat{c}_{mm} = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^{p} \frac{1}{\lambda_i(\lambda_i + 1)}}{\sum_{i=1}^{p} \frac{\underset{\sim}{\hat{\alpha}_i}^2}{(\lambda_i + 1)^2}} \right]$$

จากสมการ (2.15) เราสามารถเขียนเป็นกรณีทั่วๆ ไปได้ คือ

$$(2.16) \qquad \hat{c}_{mmh} = 1 - h\hat{\sigma}^2 \left[\frac{\sum_{i=1}^{p} \frac{1}{\lambda_i(\lambda_i + 1)}}{\sum_{i=1}^{p} \frac{\underset{\sim}{\hat{\alpha}_i}^2}{(\lambda_i + 1)^2}} \right] \quad ; \quad h > 0$$

ค่าพารามิเตอร์ c สามารถใช้หลักการของการหาค่าพารามิเตอร์ k ได้ ซึ่ง Liu Keijian (1993:343-402) ได้ทำการศึกษาโดยใช้วิธีการทำซ้ำกับหลักวิธีการของ Hoerl and Kennard , GCV(Generalized Cross Validation) และวิธีที่ใช้เกณฑ์ $C_L(C_L - \text{Criterion})$ ซึ่งผลสรุปที่ได้ คือ วิธีการทำซ้ำของ Hoerl and Kennard จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยที่มีค่าต่ำสุด ดังนั้นในการวิจัยครั้งนี้จะใช้ค่า c ที่ได้จากการทำซ้ำของ Hoerl and Kennard ซึ่งมีขั้นตอนการทำซ้ำคือ

<u>ขั้นที่ 1</u> $\quad \underset{\sim}{\hat{\alpha}}$

$$c_0 = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^{p} \frac{1}{\lambda_i(\lambda_i + 1)}}{\sum_{i=1}^{p} \frac{\underset{\sim}{\hat{\alpha}_i}^2}{(\lambda_i + 1)^2}} \right]$$

<u>ขั้นที่ 2</u> $\underset{\sim}{\hat{\alpha}}_C(c_0)$ $c_1 = 1-\hat{\sigma}^2\left[\dfrac{\sum_{i=1}^{p}\dfrac{1}{\lambda_i(\lambda_i+1)}}{\sum_{i=1}^{p}\dfrac{\underset{\sim}{\hat{\alpha}}_{Ci}^{2}(c_0)}{(\lambda_i+1)^2}}\right]$

<u>ขั้นที่ 3</u> $\underset{\sim}{\hat{\alpha}}_C(c_1)$ $c_2 = 1-\hat{\sigma}^2\left[\dfrac{\sum_{i=1}^{p}\dfrac{1}{\lambda_i(\lambda_i+1)}}{\sum_{i=1}^{p}\dfrac{\underset{\sim}{\hat{\alpha}}_{Ci}^{2}(c_1)}{(\lambda_i+1)^2}}\right]$

...

ผู้วิจัยจะยุติการทำซ้ำถ้า $\dfrac{(c_j - c_{j+1})}{c_j} \leq 20\,T^{-13}$ ซึ่ง $T = \dfrac{\text{trace}(X'X)^{-1}}{p}$

2. คุณสมบัติทางสถิติของตัวประมาณที่ใช้หลักการของริดจ์และสไตน์

การศึกษาคุณสมบัติของตัวประมาณที่ใช้หลักการของริดจ์และสไตน์มีทฤษฎีต่างๆ ที่เกี่ยวข้องดังนี้

<u>ทฤษฎีที่ 2.1</u> ถ้า $0 < c < 1$ จะได้ว่า $MSE(\underset{\sim}{\hat{\beta}}_C) < MSE(\underset{\sim}{\hat{\beta}})$

<u>พิสูจน์</u> จากคุณสมบัติในสมการ (2.13) ถ้าเราสามารถพิสูจน์ได้ว่า $MSE(\underset{\sim}{\hat{\alpha}}_C) < MSE(\underset{\sim}{\hat{\alpha}})$ ก็จะสามารถสรุปได้ว่า $MSE(\underset{\sim}{\hat{\beta}}_C) < MSE(\underset{\sim}{\hat{\beta}})$

จาก $\text{cov}(\underset{\sim}{\hat{\alpha}}_C) = \sigma^2(\Lambda+I)^{-1}(\Lambda+cI)\Lambda^{-1}(\Lambda+cI)(\Lambda+I)^{-1}$

และ $E(\underset{\sim}{\hat{\alpha}}_C) = (\Lambda+I)^{-1}(\Lambda+cI)\underset{\sim}{\alpha}$

$$MSE(\underset{\sim}{\hat{\alpha}}_C) = \text{trace}\{\text{cov}(\underset{\sim}{\hat{\alpha}}_C)\} + \left\|E(\underset{\sim}{\hat{\alpha}}_C) - \underset{\sim}{\hat{\alpha}}\right\|^2$$

$$= \sigma^2 \sum_{i=1}^{p} \frac{(\lambda_i + c)^2}{\lambda_i(\lambda_i+1)^2} + (c-1)^2 \sum_{i=1}^{p} \frac{\underset{\sim}{\alpha_i}^2}{(\underset{\sim}{\alpha_i}+1)^2}$$

$$= g(c)$$

อนุพันธ์อันดับที่ 1 ของ g(c) คือ

$$g'(c) = 2\sigma^2 \sum_{i=1}^{p} \frac{(\lambda_i + c)}{\lambda_i(\lambda_i+1)^2} + 2(c-1)^2 \sum_{i=1}^{p} \frac{\underset{\sim}{\alpha_i}^2}{(\underset{\sim}{\alpha_i}+1)^2} \tag{2.17}$$

เมื่อ $c = 1$ ตัวประมาณที่ใช้หลักการของริดจ์และสไตน์ก็คือ $\hat{\underset{\sim}{\alpha}}$ ซึ่งจะได้ว่า

$$g'(1) = 2\sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i(\lambda_i+1)} > 0 \tag{2.18}$$

จากสมการ (2.17) และ (2.18) จะเห็นได้ว่าเมื่อ $0 < c < 1$ อัตราการเพิ่มขึ้นของ $MSE(\hat{\underset{\sim}{\alpha}}_C)$ เมื่อเทียบกับ c จะมีค่าน้อยกว่าอัตราการเพิ่มขึ้นของ $MSE(\hat{\underset{\sim}{\alpha}})$ เมื่อเทียบกับ c แสดงว่าถ้า $0 < c < 1$ จะทำให้ $g(c) < g(1)$ นั่นก็คือ ทำให้ $MSE(\hat{\underset{\sim}{\alpha}}_C) < MSE(\hat{\underset{\sim}{\alpha}})$ ทำให้สรุปได้ว่า ถ้า $0 < c < 1$ จะได้ว่า $MSE(\hat{\underset{\sim}{\beta}}_C) < MSE(\hat{\underset{\sim}{\beta}})$

พิจารณาสมการ (2.16) เราพบว่าการประมาณค่า c จะต้องมีการประมาณค่า h ขึ้นมาก่อน ซึ่งค่า h จะต้องอยู่ในขอบเขตที่กำหนดจึงจะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากวิธี RS มีค่าน้อยกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากวิธี OLS ซึ่งมีทฤษฎีที่เกี่ยวข้องดังนี้

<u>ทฤษฎีที่ 2.2</u> สมมติว่า $\underset{\sim}{\varepsilon} \sim N(0, \sigma^2 I)$ ค่า h ที่เหมาะสมคือ

$$0 < h < \frac{2(n-p)}{n-p+2} \left[1 - \frac{2}{\sum_{i=1}^{p} \frac{\lambda_p(\lambda_p+1)}{\lambda_i(\lambda_i+1)}} \right]$$ สำหรับทุกค่าของ $\underset{\sim}{\beta_i}$ และ σ^2 จะทำให้ได้ว่า

$$\text{MSE}(\hat{\beta}_C(\hat{c}_{mmh})) < \text{MSE}(\hat{\beta})$$

<u>พิสูจน์</u> เราเพียงพิสูจน์ว่า $\text{MSE}(\hat{\alpha}_C(\hat{c}_{mmh})) < \text{MSE}(\hat{\alpha})$

ให้ $\hat{\alpha}_C(\hat{c}_{mmh}) = \tilde{\alpha} = (\Lambda+1)^{-1}(\Lambda+\hat{c}_{mmh}I)\hat{\alpha}$

ดังนั้น $\tilde{\alpha}_i = \dfrac{\lambda_i+\hat{c}_{mmh}}{\lambda_i+1}\hat{\alpha}_i \quad , \quad i = 1,2,\ldots,p$

และ
$$E(\tilde{\alpha}_i-\alpha_i)^2 = E(\hat{\alpha}_i-\alpha_i+\frac{\hat{c}_{mmh}-1}{\lambda_i+1}\hat{\alpha}_i)^2$$

$$= E(\tilde{\alpha}_i-\alpha_i)^2+\frac{2}{\lambda_i+1}\; E(\hat{c}_{mmh}-1)\hat{\alpha}_i(\hat{\alpha}_i-\alpha_i)$$

$$+\frac{1}{(\lambda_i+1)^2}E(\hat{\alpha}_{mmh}-1)\hat{\alpha}^2$$

จะได้ว่า $E(\hat{c}_{mmh}-1)\hat{\alpha}_i(\hat{\alpha}_i-\alpha_i)$

$$= -h\,E\left[\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Big/\sum_{i=1}^{p}\frac{\hat{\alpha}_i^2}{(\lambda_i+1)^2}\right]\hat{\alpha}_i\;(\hat{\alpha}_i-\alpha)$$

เนื่องจาก $\hat{\alpha}_i \sim N(\alpha_i,\frac{\sigma^2}{\lambda_i})$ จะได้ว่า

$$E\left[\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Big/\sum_{i=1}^{p}\frac{\hat{\alpha}_i^2}{(\lambda_i+1)^2}\right]\hat{\alpha}_i\;(\hat{\alpha}_i-\alpha)$$

$$= \frac{\sigma^2}{\lambda_i}\,E\left[\;\left[\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Big/\sum_{i=1}^{p}\frac{\hat{\alpha}_i^2}{(\lambda_i+1)^2}\right]-\right.$$

$$\left[\frac{2\underset{\sim}{\hat{\alpha}}_i^{\,2}}{(\lambda_i+1)^2}\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Bigg/\left[\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{\,2}}{(\lambda_i+1)^2}\right]^2\right]\Bigg]$$

และ $E(\underset{\sim}{\tilde{\alpha}}_i-\underset{\sim}{\alpha}_i)^2$

$$= E(\underset{\sim}{\hat{\alpha}}_i-\underset{\sim}{\alpha})^2 - \frac{2h\sigma^2}{\lambda_i(\lambda_i+1)}E\Bigg[\left[\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Bigg/\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{2}}{(\lambda_i+1)^2}\right]$$

$$-\left[\frac{2\underset{\sim}{\hat{\alpha}}_i^{\,2}}{(\lambda_i+1)^2}\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Bigg/\left[\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{\,2}}{(\lambda_i+1)^2}\right]^2\right]\Bigg]$$

$$+ \frac{h^2}{(\lambda_i+1)^2}E\left[\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Bigg/\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{\,2}}{(\lambda_i+1)^2}\right]^2\underset{\sim}{\hat{\alpha}}_i^{\,2}$$

ดังนั้น $MSE(\underset{\sim}{\tilde{\alpha}})-MSE(\underset{\sim}{\hat{\alpha}})$

$$= E\Bigg[\left[\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Bigg/\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{2}}{(\lambda_i+1)^2}\right]\Bigg[\sum_{i=1}^{p}\frac{-2\sigma^2h}{\lambda_i(\lambda_i+1)} +$$

$$\left[4\sigma^2h\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{\,2}}{\lambda_i(\lambda_i+1)}\Bigg/\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{\,2}}{(\lambda_i+1)^2}\right]+\sum_{i=1}^{p}\frac{h^2\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Bigg]\Bigg]$$

$$= E\Bigg[\left[\sum_{i=1}^{p}\frac{h\hat{\sigma}^2}{\lambda_i(\lambda_i+1)}\Bigg/\sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}}_i^{\,2}}{(\lambda_i+1)^2}\right]\Bigg[\sum_{i=1}^{p}\frac{-2\sigma^2}{\lambda_i(\lambda_i+1)} +$$

$$\frac{4\sigma^2}{\lambda_p(\lambda_p+1)} + h\sum_{i=1}^{p}\frac{\hat{\sigma}^2}{\lambda_i(\lambda_i+1)} \Bigg]\Bigg]$$

เนื่องจาก $\underset{\sim}{\hat{\alpha}}$ และ $\hat{\sigma}^2$ เป็นอิสระซึ่งกันและกัน และค่าคาดหวังของ $\hat{\sigma}^2$, $\hat{\sigma}^4$ มีค่าเท่ากับ $E(\hat{\sigma}^2) = \sigma^2$, $E(\hat{\sigma}^4) = \frac{n-p+2}{n-p}\sigma^4$ ตามลำดับ

จะได้ว่า $MSE(\underset{\sim}{\tilde{\alpha}}) - MSE(\underset{\sim}{\hat{\alpha}})$

$$= E\left[\sum_{i=1}^{p}\frac{h}{\lambda_i(\lambda_i+1)} \Bigg/ \sum_{i=1}^{p}\frac{\underset{\sim}{\hat{\alpha}_i}^{2}}{(\lambda_i+1)^2}\right]\Bigg[\sum_{i=1}^{p}\frac{-2\sigma^4}{\lambda_i(\lambda_i+1)} +$$

$$\frac{4\sigma^4}{\lambda_p(\lambda_p+1)} + h\sum_{i=1}^{p}\frac{n-p+2}{n-p}\,\frac{\hat{\sigma}^4}{\lambda_i(\lambda_i+1)}\Bigg]$$

ดังนั้นเมื่อ $0 < h < \frac{2(n-p)}{n-p+2}\left[1 - \frac{2}{\sum_{i=1}^{p}\frac{\lambda_p(\lambda_p+1)}{\lambda_i(\lambda_i+1)}}\right]$

จะได้ว่า $MSE(\underset{\sim}{\tilde{\alpha}}) < MSE(\underset{\sim}{\hat{\alpha}})$

นั่นก็คือ $MSE(\underset{\sim}{\hat{\beta}}_C(\hat{c}_{mmh})) < MSE(\underset{\sim}{\hat{\beta}})$

<u>บทแทรก 2.2.1</u> เมื่อ $\frac{2(n-p)}{n-p+2}\left[1 - \frac{2}{\sum_{i=1}^{p}\frac{\lambda_p(\lambda_p+1)}{\lambda_i(\lambda_i+1)}}\right] > 1$ สำหรับทุกค่า $\underset{\sim}{\beta}$ และ σ^2

จะได้ว่า $MSE(\underset{\sim}{\hat{\beta}}_C(\hat{c}_{mmh})) < MSE(\underset{\sim}{\hat{\beta}})$

การแปลงข้อมูลด้วยการแปลงที่อยู่ในรูปยกกำลัง(Power Transformation)

การแปลงข้อมูลด้วยการแปลงที่อยู่ในรูปยกกำลัง (Power Transformation) ได้รวมการแปลงซึ่งใช้ลอกการิทึมในวงศ์ (family) ของการแปลง ซึ่งการแปลงข้อมูลอาจจะทำให้ข้อสมมติของความเป็นปกติใช้ได้สำหรับกรณีข้อมูลที่ศึกษาการแจกแจงของค่าคลาดเคลื่อนเป็นแบบเบ้

Box และ Cox(ค.ศ 1962 : 211-243) ได้พิจารณาการแปลงในตัวแปรตามและสามารถทำในตัวแปรอิสระด้วย การแปลงของเขาอยู่ในรูปของ

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

ซึ่งเป็นข้อดีของการแปลง เมื่อเปรียบเทียบการแปลงอย่างง่าย y_i^{λ} ซึ่งจะต่อเนื่องใน $\lambda = 0$
ดังนั้น $L_{t_{\lambda \to 0}} \dfrac{y_i^{\lambda}-1}{\lambda} = \log_e y_i$

Box และ Cox พิจารณาตัวแบบของการถดถอย

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

เมื่อ $\underset{\sim}{\varepsilon_i} \sim I_n(0,\sigma^2)$ ดังนั้น ฟังก์ชันความหนาแน่นร่วมของค่าสังเกต $y_1, y_2, y_3, \ldots, y_n$ เป็น

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{p} (y_i^{(\lambda)} - \underset{\sim}{x}'\underset{\sim}{\beta})^2\right] * J$$

โดยที่ J เป็นจาโคเบียนของการแปลงจากตัวแปร y_i เป็น $y_i^{(\lambda)}$ ดังนั้น

$$J = \prod_{i=1}^{n}\left|\frac{dy_i^{(\lambda)}}{dy_i}\right| = \prod_{i=1}^{n} y_i^{\lambda-1}$$

จะได้ว่า ฟังก์ชัน log-likelihood อยู่ในรูปของ

$$\text{Log}_e L \; = \; \frac{n}{2}\log_e 2\pi \; - \; \frac{n}{2}\log_e \sigma^2 \; - \; \frac{1}{2\sigma^2}\sum_{i=1}^{p}\left[y_i^{(\lambda)} - \underset{\sim}{x}_i{}' \underset{\sim}{\beta} \right]^2 + (\lambda - 1)\sum_{i=1}^{n}\log_e y_i$$

ถ้าเราหารแต่ละ y_i ด้วย geometric mean ของ y แล้ว เราจะได้ $\sum_{i=1}^{p} \log_e y_i = 0$ ซึ่งทำให้เทอมสุดท้ายหายไป และค่าที่มากที่สุดของ L เท่ากับ $\frac{n}{2}\log_e \hat{\sigma}^2{}_{\varepsilon/\lambda}$ เมื่อ $\hat{\sigma}^2{}_{\varepsilon/\lambda}$ เป็นผลรวมกำลังสองของค่าคลาดเคลื่อนจากการถดถอยของ $y^{(\lambda)}$ บน x_i ดังนั้นขั้นตอนต่างๆ มีดังนี้

1. หารค่าแต่ละ y ด้วยค่ากลางเรขาคณิต y

2. สำหรับค่าของ λ ทำการถดถอย $y^{(\lambda)}$ บน x_i และคำนวณผลรวมกำลังสองของค่าความคลาดเคลื่อน $\hat{\sigma}^2{}_{\varepsilon/\lambda}$

3. เลือกค่า λ สำหรับ $\hat{\sigma}^2{}_{\varepsilon/\lambda}$ ที่น้อยที่สุด จะได้ว่า $\hat{\lambda}$ เป็น maximum likelihood ของ λ