บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณของตัวแบบการถดถอยเชิงเส้นด้วยวิธีกำลังสองน้อยสุด วิธีริดจ์รีเกรสชันที่ใช้ข้อสนเทศโดยหลักเกณฑ์และวิธีลิว ดีเจียนทั่วไป ซึ่งรายละเอียดต่าง ๆ มีดังนี้

2.1 ตัวแบบทั่วไป

ตัวแบบทั่วไปของการวิเคราะห์การถดถอยพหุคูณเชิงเส้นสามารถเขียนได้ดังนี้

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

เมื่อ $\underset{\sim}{y}$ คือ เวกเตอร์ของตัวแปรตามขนาด $n\times1$

X คือ เมทริกซ์ของตัวแปรอิสระขนาด $n\times(p+1)$ และ $X'X$ มีค่าลำดับชั้นเท่ากับ $p+1$

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยขนาด $(p+1)\times1$

$\underset{\sim}{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนขนาด $n\times1$ โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และ $Cov(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

n เป็นขนาดตัวอย่างของตัวแปรแต่ละตัว

และ p เป็นจำนวนตัวแปรอิสระ

จากตัวแบบทั่วไปจะได้ผลรวมความคลาดเคลื่อนกำลังสองอยู่ในรูปแบบดังนี้

$$S(\underset{\sim}{\beta}) = \underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} = (\underset{\sim}{y} - X\underset{\sim}{\beta})'(\underset{\sim}{y} - X\underset{\sim}{\beta})$$

$$= \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\beta}'X'\underset{\sim}{y} - \underset{\sim}{y}'X\underset{\sim}{\beta} + \underset{\sim}{\beta}'X'X\underset{\sim}{\beta}$$

(2.1.2) $$= \underset{\sim}{y}'\underset{\sim}{y} - 2\underset{\sim}{\beta}'X'\underset{\sim}{y} + \underset{\sim}{\beta}'X'X\underset{\sim}{\beta}$$

วิธีกำลังสองน้อยสุดมีหลักการที่ทำให้ผลรวมความคลาดเคลื่อนกำลังสองมีค่าต่ำที่สุด ดังนั้นเราจะได้ค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณจากการหาอนุพันธ์อันดับที่ 1 ของสมการที่ (2.1.2) เทียบกับ $\underset{\sim}{\beta}$ แล้วให้เท่ากับศูนย์ ซึ่งผลดังกล่าวอยู่ในรูปแบบดังนี้

$$\frac{\partial}{\partial\underset{\sim}{\beta}}S(\underset{\sim}{\beta}) = -2X'\underset{\sim}{y} + 2X'X\hat{\underset{\sim}{\beta}} = \underset{\sim}{0}$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

กล่าวคือ $X'X\hat{\underset{\sim}{\beta}} = X'\underset{\sim}{y}$

ซึ่งเราเรียกสมการดังกล่าวว่า สมการปกติ (normal equation) และจะได้ตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณในรูปของ

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

ตัวประมาณในสมการที่ (2.1.3) มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงของ $\underset{\sim}{\beta}$ กล่าวคือ

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

ตัวประมาณ $\hat{\underset{\sim}{\beta}}$ จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดในบรรดาตัวประมาณไม่เอนเอียงเชิงเส้น แต่ในการประมาณค่า $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุดมีข้อสมมุติฐานที่สำคัญข้อหนึ่งคือตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์กับตัวแปรอิสระตัวอื่นซึ่งในทางปฏิบัติจะเกิดขึ้นได้น้อยมาก เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันสูงจะทำให้เมทริกซ์ $X'X$ เกิดเงื่อนไขที่ไม่ดี (ill-condition) อาจมีผลทำให้การประมาณ $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุดไม่ได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด เราจึงควรตรวจสอบประสิทธิภาพของตัวประมาณที่ได้โดยเราพิจารณาประสิทธิภาพของตัวประมาณค่า $\underset{\sim}{\beta}$ สองส่วน กล่าวคือเมทริกซ์ความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ และค่าเฉลี่ยของกำลังสองระยะทางจาก $\hat{\underset{\sim}{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ ซึ่งเราสามารถเขียนเมทริกซ์ความแปรปรวนร่วมของ $\hat{\underset{\sim}{\beta}}$ ในรูปฟังก์ชันของ $X'X$ และ σ^2 ดังต่อไปนี้

(2.1.4) $$Cov(\hat{\underset{\sim}{\beta}}) = \sigma^2(X'X)^{-1}$$

ให้ L_1 คือระยะทางจาก $\hat{\underset{\sim}{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ ดังนั้น

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

และเราจะได้ค่าเฉลี่ยของกำลังสองระยะทางจาก $\hat{\underset{\sim}{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ ในรูปของ

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

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

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

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

เมื่อ $\underset{\sim}{\varepsilon}$ มีการแจกแจงแบบปกติจะได้ว่า

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

จากสมการที่ (2.1.4) , (2.1.6) และ (2.1.8) เราจะเห็นได้ว่า $\operatorname{Cov}(\hat{\underset{\sim}{\beta}})$, $E[L_1^2]$ และ $\operatorname{Var}[L_1^2]$ ต่างก็เป็นฟังก์ชันของเมทริกซ์ $X'X$ ดังนั้นเพื่อความสะดวกในการทำความเข้าใจเราจึงแปลงเมทริกซ์ $X'X$ ให้อยู่ในรูปของค่าเฉพาะ (eigenvalue) ของเมทริกซ์ $X'X$ โดยใช้ทฤษฎีที่สำคัญข้อหนึ่งคือ ถ้า λ_i เป็นค่าเฉพาะของเมทริกซ์ $X'X$ แล้ว $\sum_{i=1}^{p}\lambda_i = \operatorname{trace}(X'X)$; $i=1,2,\ldots\ldots\ldots,p$ เมื่อ p เป็นจำนวนตัวแปรอิสระ

กำหนดให้ค่าเฉพาะของเมทริกซ์ $X'X$ มีค่าเป็น

$$(\lambda_{max} = \lambda_1) \ge \lambda_2 \ge \ldots\ldots\ldots\ge (\lambda_p = \lambda_{min}) \quad ; \quad \lambda_1, \lambda_2, \ldots\ldots\ldots, \lambda_p > 0$$

จากสมการที่ (2.1.6) เราสามารถเขียนค่าเฉลี่ยของกำลังสองระยะทางจาก $\hat{\underset{\sim}{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ ในรูปฟังก์ชันของค่าเฉพาะได้ดังนี้

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

และจากสมการที่ (2.1.8) เราสามารถเขียนค่าความแปรปรวนของกำลังสองระยะทางจาก $\hat{\underset{\sim}{\beta}}$ ไปยัง $\underset{\sim}{\beta}$ อยู่ในรูปฟังก์ชันของค่าเฉพาะได้ดังนี้

(2.1.10) $$\operatorname{Var}(L_1^2) = 2\sigma^4 \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)^2$$

ในกรณีที่ตัวแปรอิสระมีสภาพไม่เหมาะสม กล่าวคือเกิดความสัมพันธ์ระหว่างตัวแปรอิสระในอัตราที่สูงจะทำให้ $|X'X|$ มีค่าเล็กลงเข้าใกล้ศูนย์ เนื่องจาก $|X'X|$ มีค่าเท่ากับผลคูณของค่าเฉพาะของเมทริกซ์ $X'X$ จึงส่งผลให้ค่าเฉพาะบางค่าต่ำมาก ดังนั้นจากสมการที่ (2.1.9) และ (2.1.10) เราจะเห็นได้ว่า $E(L_1^2)$ และ $Var(L_1^2)$ จึงมีค่าสูงขึ้นตามไปด้วย นอกจากนี้การเกิดความสัมพันธ์ระหว่างตัวแปรอิสระส่งผลให้ความแปรปรวนของค่าประมาณสัมประสิทธิ์การถดถอยมีค่ามากและเกิดความสัมพันธ์กันสูงระหว่างสัมประสิทธิ์การถดถอยที่ใช้ประมาณค่า

2.2 วิธีริดจ์รีเกรสชันที่ใช้ข้อสนเทศโดยหลักเกณฑ์ (Ridge Regression with Prior Information Method)

ในปี ค.ศ. 1970 โฮเอิล (Hoerl) และเคนนาร์ด (Kennard) ได้เสนอการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีริดจ์รีเกรสชันเพื่อแก้ปัญหาในกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์กัน โดยใช้หลักการที่ว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นฟังก์ชันของ $(X'X)^{-1}$ ดังนั้นการที่จะลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองให้ต่ำลงจึงต้องมุ่งไปที่การลด $(X'X)^{-1}$ ให้ต่ำลงหรือหมายถึงการทำให้ $|X'X|$ เพิ่มมากขึ้น ซึ่งทำได้โดยการบวกค่าคงที่ที่มากกว่าศูนย์กับสมาชิกทุกตัวในแนวทแยงมุมของเมทริกซ์ $X'X$ ดังนั้นค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีริดจ์รีเกรสชันอยู่ในรูปของ

$$\underset{\sim}{\hat{\beta}}_R(k) = (X'X + kI)^{-1}(X'\underset{\sim}{y}) \quad ; \; k > 0 \tag{2.2.1}$$

ในปี ค.ศ.1976 สวินเดล (Swindel) ได้เสนอการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีริดจ์รีเกรสชันที่ใช้ข้อสนเทศโดยหลักเกณฑ์เพื่อทำให้การประมาณสัมประสิทธิ์การถดถอยพหุคูณมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีริดจ์รีเกรสชัน ซึ่งตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณอยู่ในรูปแบบของ

$$\begin{aligned}
\underset{\sim}{\hat{\beta}}_R(kI,\underset{\sim}{j}) &= (X'X + kI)^{-1}(X'\underset{\sim}{y} + k\underset{\sim}{j}) \\
&= (X'X + kI)^{-1}X'X(X'X)^{-1}(X'\underset{\sim}{y} + k\underset{\sim}{j}) \\
&= (I + k(X'X)^{-1})^{-1}(\underset{\sim}{\hat{\beta}} - \underset{\sim}{j} + \underset{\sim}{j} + k(X'X)^{-1}\underset{\sim}{j}) \\
&= (I + k(X'X)^{-1})^{-1}(\underset{\sim}{\hat{\beta}} - \underset{\sim}{j}) + \underset{\sim}{j} \qquad (2.2.2)
\end{aligned}$$

จากสมการที่ (2.2.1) และ (2.2.2) เราจะเห็นได้ว่า $\hat{\underset{\sim}{\beta}}_R(kI,\underset{\sim}{j}) = \hat{\underset{\sim}{\beta}}$ เมื่อ k=0 หรือ $\underset{\sim}{j} = \hat{\underset{\sim}{\beta}}$ และ $\hat{\underset{\sim}{\beta}}_R(kI,\underset{\sim}{j}) = \hat{\underset{\sim}{\beta}}_R(k)$ เมื่อ $\underset{\sim}{j} = \underset{\sim}{0}$ ดังนั้นเราจึงกล่าวได้ว่าวิธีกำลังสองน้อยสุดและวิธีริดจ์รีเกรสชันเป็นกรณีเฉพาะของวิธีริดจ์รีเกรสชันที่ใช้ข้อสนเทศโดยหลักเกณฑ์ กำหนดให้ $\underset{\sim}{j}$ มีการแจกแจงแบบปกติและเป็นอิสระกับ $\hat{\underset{\sim}{\beta}}$ กล่าวคือ $\underset{\sim}{j} \sim N(\underset{\sim}{\beta}, V)$ และตัวประมาณนูน (convex estimator) อยู่ในรูปของ

$\underset{\sim}{\beta}(C,\underset{\sim}{j}) = C\hat{\underset{\sim}{\beta}} + (I-C)\underset{\sim}{j}$ ภายใต้สมมุติฐานคือ $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0},\sigma^2 I_n)$ และ $\hat{\underset{\sim}{\beta}} \sim N(\underset{\sim}{\beta},\sigma^2(X'X)^{-1})$

โดยที่ V เป็นเมทริกซ์ความแปรปรวนร่วมขนาด p×p และ V มีค่าลำดับชั้น = p

และ C เป็นเมทริกซ์ขนาด p×p

เราสามารถเขียนค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณนูน $\underset{\sim}{\beta}(C,\underset{\sim}{j})$ ในรูปของ

$$MSE(\underset{\sim}{\beta}(C,\underset{\sim}{j})) = trace(\sigma^2 C(X'X)^{-1}C' + CVC' - CV - VC' + V)$$

ดังนั้น $\frac{\partial}{\partial C} MSE(\underset{\sim}{\beta}(C,\underset{\sim}{j})) = 2C(\sigma^2(X'X)^{-1} + V) - 2V = 0$

ซึ่งความคลาดเคลื่อนกำลังสองจะมีค่าต่ำสุดเมื่อ $C = V(\sigma^2(X'X)^{-1} + V)^{-1}$

เนื่องจาก $\hat{\underset{\sim}{\beta}}$ มีการแจกแจงใด ๆที่มีค่าเฉลี่ย $\underset{\sim}{\beta}$ และเมทริกซ์ความแปรปรวนร่วมเป็น $\sigma^2(X'X)^{-1}$ กล่าวคือ $\hat{\underset{\sim}{\beta}} \sim M(\underset{\sim}{\beta},\sigma^2(X'X)^{-1})$ และ $\underset{\sim}{j} \sim M\left(\underset{\sim}{\beta},\left(\frac{\sigma^2}{k}\right)I_n\right)$; $k>0$ เมื่อกำหนดตัวประมาณนูนอยู่ในรูปของ $\underset{\sim}{\beta}(C,\underset{\sim}{j}) = C\hat{\underset{\sim}{\beta}} + (I-C)\underset{\sim}{j}$ ดังนั้นค่าเหมาะสมที่สุดของ C ที่ทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุด คือ

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

นอกจากนี้ $\underset{\sim}{\beta}(C,\underset{\sim}{j})$ มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงของ $\underset{\sim}{\beta}$

สำหรับปัญหาหนึ่งของวิธีริดจ์รีเกรสชันที่ใช้ข้อสนเทศโดยหลักเกณฑ์ คือการคำนวนหาค่า k ที่เหมาะสม โดยสวินเดลไม่ได้เสนอวิธีการประมาณค่า k แต่ใช้วิธีการลองแทนค่า k ไปเรื่อย ๆ จนได้ค่า k ที่เหมาะสม ดังนั้นในปี ค.ศ. 1995 Robert H. Crouse และ Chun Jin จึงได้เสนอการประมาณ $\frac{1}{k}$ แทนที่จะประมาณค่า k โดยตรงดังนี้

เนื่องจาก $\hat{\underset{\sim}{\beta}} - \underset{\sim}{j}$ มีการแจกแจงใด ๆ ที่มีค่าเฉลี่ยเป็น $\underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วมเป็น $\left(\frac{\sigma^2}{k}\right)I_a + \sigma^2(X'X)^{-1}$ กล่าวคือ

$$\hat{\underset{\sim}{\beta}} - \underset{\sim}{j} \sim M\left(\underset{\sim}{0}, \left(\frac{\sigma^2}{k}\right)I_a + \sigma^2(X'X)^{-1}\right)$$

และ $E\left[(\hat{\underset{\sim}{\beta}} - \underset{\sim}{j})'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{j})\right] = \left(\frac{p\sigma^2}{k}\right) + \sigma^2 \operatorname{trace}(X'X)^{-1}$ ดังนั้นการประมาณที่ไม่เอนเอียงของ $\frac{1}{k}$ เมื่อทราบค่า σ^2 คือ

$$\frac{1}{\hat{k}} = \frac{(\hat{\underset{\sim}{\beta}} - \underset{\sim}{j})'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{j}) - \sigma^2 \operatorname{trace}(X'X)^{-1}}{p\sigma^2} \qquad (2.2.3)$$

เนื่องจากเราสามารถเขียนตัวประมาณริดจ์รีเกรสชันที่ใช้ข้อสนเทศโดยหลักเกณฑ์ในรูปของ

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

โดยที่ $\hat{k} = \dfrac{p\sigma^2}{(\hat{\underset{\sim}{\beta}} - \underset{\sim}{j})'(\hat{\underset{\sim}{\beta}} - \underset{\sim}{j}) - \sigma^2 \operatorname{trace}(X'X)^{-1}}$

เราจะได้ว่า $E[\hat{\underset{\sim}{\beta}}(\hat{k}I, \underset{\sim}{j})] = E[\underset{\sim}{j}] = \underset{\sim}{\beta}$ กล่าวคือ $\hat{\underset{\sim}{\beta}}(\hat{k}I, \underset{\sim}{j})$ เป็นตัวประมาณที่ไม่เอนเอียงของ $\underset{\sim}{\beta}$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ส่วนกรณีที่ไม่ทราบค่าของ σ^2 เราสามารถประมาณด้วยตัวประมาณที่ไม่เอนเอียงของ σ^2 คือ $s^2 = \frac{1}{n-p}(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})$ ดังนั้นค่าประมาณของ $\frac{1}{k}$ คือ

$$\frac{1}{\hat{k}} = \frac{(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j}) - s^2\,\mathrm{trace}\,(X'X)^{-1}}{ps^2}$$

เราจะได้ว่า $\hat{k} = \dfrac{ps^2}{(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j}) - s^2\,\mathrm{trace}(X'X)^{-1}}$ และ $E[\hat{\underset{\sim}{\beta}}(\hat{k}I,\underset{\sim}{j})] = E[\underset{\sim}{j}] = \underset{\sim}{\beta}$

ในปี ค.ศ. 1987 พลิสกิน (Pliskin) ได้แสดงให้เห็นว่าสำหรับเวกเตอร์คงที่ใด ๆ คือ $\underset{\sim}{j}$ เมื่อใช้ค่า k ที่เหมือนกันแทนใน $\hat{\underset{\sim}{\beta}}(kI,\underset{\sim}{j})$ และ $\hat{\underset{\sim}{\beta}}(kI,\underset{\sim}{0})$ แล้ว

$$E\left[(\hat{\underset{\sim}{\beta}}(kI,\underset{\sim}{j})-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}(kI,\underset{\sim}{j})-\underset{\sim}{\beta})\right] \le E\left[(\hat{\underset{\sim}{\beta}}(kI,\underset{\sim}{0})-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}(kI,\underset{\sim}{0})-\underset{\sim}{\beta})\right]$$

จากสมการที่ (2.2.3) ค่าประมาณของ $\frac{1}{k}$ มีโอกาสเป็นลบได้ แต่ $\underset{\sim}{j}$ มีการแจกแจง คือ $M(\underset{\sim}{\beta},\left(\frac{\sigma^2}{k}\right)I_n)$ ซึ่งสวินเดลได้แสดงให้เห็นแล้วว่ามีค่า k ที่ทำให้ $\hat{\underset{\sim}{\beta}}(\hat{k}I,\underset{\sim}{j})$ มี MSE ต่ำกว่า $\hat{\underset{\sim}{\beta}}$ ของการประมาณด้วยวิธีกำลังสองน้อยสุด ดังนั้นเมื่อ $(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j}) < \sigma^2\mathrm{trace}(X'X)^{-1}$ ซึ่งทำให้ $\frac{1}{\hat{k}}$ เป็นลบได้ ผู้วิจัยจึงควรใช้ค่าประมาณของ k คือ

$$\hat{k} = \frac{ps^2}{(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})}$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ผู้วิจัยขอสรุปการเลือกใช้ค่าประมาณของ k ดังนี้

<u>กรณีทราบค่า</u> σ^2

$$\hat{k} = \begin{cases} \dfrac{p\sigma^2}{(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})-\sigma^2\,\mathrm{trace}(X'X)^{-1}} & ;\quad (\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})-\sigma^2\,\mathrm{trace}(X'X)^{-1}>0 \\[2ex] \dfrac{p\sigma^2}{(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})} & ;\quad \text{อื่น ๆ} \end{cases}$$

<u>กรณีไม่ทราบค่า</u> σ^2

$$\hat{k} = \begin{cases} \dfrac{ps^2}{(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})-s^2\,\mathrm{trace}(X'X)^{-1}} & ;\quad (\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})-s^2\,\mathrm{trace}(X'X)^{-1}>0 \\[2ex] \dfrac{ps^2}{(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{j})} & ;\quad \text{อื่น ๆ} \end{cases}$$

โดยที่ $s^2 = \dfrac{1}{n-p}(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}})$

2.3 วิธีลิว คีเจียนทั่วไป (Generalized Liu Kejian Method)

ในปี ค.ศ. 1993 ลิว คีเจียน ได้เสนอวิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระโดยนำข้อดีวิธีริดจ์รีเกรสชันและวิธีของสไตน์มาผสมผสานกัน ซึ่งค่าประมาณของสัมประสิทธิ์การถดถอยพหุคูณวิธีของลิว คีเจียนอยู่ในรูปแบบดังนี้

$$(2.3.1) \qquad \hat{\underset{\sim}{\beta}}_L(d) = (X'X+I)^{-1}(X'\underset{\sim}{y}+d\hat{\underset{\sim}{\beta}}) \quad ; \quad 0<d<1$$

เราสามารถเขียนตัวประมาณในสมการที่ (2.3.1) อยู่ในเทอมของค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุดดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\hat{\underset{\sim}{\beta}}_L(d) = (X'X+I)^{-1}(X'\underset{\sim}{y}+d\hat{\underset{\sim}{\beta}}) \quad ; \quad 0<d<1$$

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

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

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

โดยที่ $Z = (X'X+I)^{-1}(X'X+dI)$

ค่าคาดหวังของตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีของลิว ดีเจียนคือ

$$E[\hat{\underset{\sim}{\beta}}_L(d)] = E(Z\hat{\underset{\sim}{\beta}})$$

$$= ZE(\hat{\underset{\sim}{\beta}})$$

$$= Z\underset{\sim}{\beta}$$

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

เนื่องจากตัวประมาณ $\hat{\underset{\sim}{\beta}}_L(d)$ สามารถเขียนได้ในอีกรูปแบบหนึ่งดังนี้

$$\hat{\underset{\sim}{\beta}}_L(d) = \left[I-(X'X+I)^{-1}(I-dI)\right]\hat{\underset{\sim}{\beta}}$$

ดังนั้น $\hat{\underset{\sim}{\beta}}_L(d)$ มีความเอนเอียงเท่ากับ $-(X'X+I)^{-1}(I-dI)\underset{\sim}{\beta}$

เราจะเห็นได้ว่าความเอนเอียงที่เกิดขึ้นเป็นฟังก์ชันลดลง โดยความเอนเอียงจะลดลงเมื่อค่าของ d มีค่าสูงขึ้นและมีค่าเท่ากับศูนย์เมื่อ d=1 ซึ่งเป็นการประมาณสัมประสิทธิ์ความถดถอยพหุด้วยวิธีกำลังสองน้อยสุดนั่นเอง

ตัวแบบทั่วไปในสมการที่ (2.1.1) สามารถเขียนในรูปแบบใหม่โดยใช้เมทริกซ์เชิงตั้งฉากได้ดังนี้

(2.3.2) $$\underset{\sim}{y} = Z\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon}$$

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

เมื่อ $\underset{\sim}{y}$ คือ เวกเตอร์ของตัวแปรตามขนาด $n\times1$

X คือ เมทริกซ์ของตัวแปรอิสระขนาด $n\times p$ และ X มีค่าลำดับชั้นเท่ากับ p

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยขนาด $p\times1$

$\underset{\sim}{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนขนาด $n\times1$ โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และ $Cov(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

Q คือ เมทริกซ์เชิงตั้งฉาก (orthogonal matrix) ขนาด $p \times p$ โดยแต่ละแนวตั้งของ Q คือเวกเตอร์เฉพาะ (eigenvector) ที่สอดคล้องกับค่าเฉพาะของ $X'X$ กล่าวคือ $Q = (q_1, q_2, \ldots\ldots, q_p)$ เมื่อ $q_1, q_2, \ldots\ldots, q_p$ เป็นเวกเตอร์เฉพาะที่สอดคล้องกับค่าเฉพาะ $\lambda_1, \lambda_2, \ldots\ldots, \lambda_p$ ตามลำดับ

Z คือ เมทริกซ์ที่แต่ละคอลัมน์ของ Z เป็นตัวแปรอิสระชุดใหม่ซึ่งเป็นผลรวมเชิงเส้น (linear combination) ของตัวแปรอิสระ

และ Λ คือ เมทริกซ์ขนาด $p \times p$ ซึ่งสมาชิกในแนวทแยงมุมเป็นค่าเฉพาะของเมทริกซ์ $X'X$ และสมาชิกนอกแนวทแยงมุมเป็น 0 กล่าวคือ $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots\ldots, \lambda_p)$

จากสมการที่ (2.3.2) เราจะได้ค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีกำลังสองน้อยสุดคือ

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

ค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีริดจ์รีเกรสชันคือ

$$\begin{aligned}\hat{\underset{\sim}{\alpha}}_R(k) &= (Z'Z+kI)^{-1}Z'\underset{\sim}{y} \\ &= (\Lambda+kI)^{-1}Z'\underset{\sim}{y} \\ &= \left[I+k(\Lambda)^{-1}\right]^{-1}\hat{\underset{\sim}{\alpha}} \qquad (2.3.3)\end{aligned}$$

และค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีลิว ทีเจียนคือ

$$\begin{aligned}\hat{\underset{\sim}{\alpha}}_L(d) &= (Z'Z+I)^{-1}(Z'\underset{\sim}{y}+d\hat{\underset{\sim}{\alpha}}) \\ &= (\Lambda+I)^{-1}(Z'\underset{\sim}{y}+d\hat{\underset{\sim}{\alpha}}) \\ &= \left[I-(\Lambda+I)^{-1}(I-dI)\right]\hat{\underset{\sim}{\alpha}} \qquad (2.3.4)\end{aligned}$$

เนื่องจาก $\hat{\underset{\sim}{\beta}}$, $\hat{\underset{\sim}{\beta}}_R(k)$ และ $\hat{\underset{\sim}{\beta}}_L(d)$ เป็นผลมาจากการนำเมทริกซ์เชิงตั้งฉากคูณกับ $\hat{\underset{\sim}{\alpha}}$, $\hat{\underset{\sim}{\alpha}}_R(k)$ และ $\hat{\underset{\sim}{\alpha}}_L(d)$ ตามลำดับ กล่าวคือ $\hat{\underset{\sim}{\beta}} = Q\hat{\underset{\sim}{\alpha}}$, $\hat{\underset{\sim}{\beta}}_R(k) = Q\hat{\underset{\sim}{\alpha}}_R(k)$ และ $\hat{\underset{\sim}{\beta}}_L(d) = Q\hat{\underset{\sim}{\alpha}}_L(d)$ ดังนั้นการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณจึงสามารถใช้ $\hat{\underset{\sim}{\alpha}}$, $\hat{\underset{\sim}{\alpha}}_R(k)$ และ $\hat{\underset{\sim}{\alpha}}_L(d)$ สำหรับการเปรียบเทียบได้เนื่องจากการเขียนอยู่ในเทอมของค่าเฉพาะทำให้สะดวกในการทำความเข้าใจและยังให้ผลลัพธ์สอดคล้องกับการเปรียบเทียบด้วย $\hat{\underset{\sim}{\beta}}$, $\hat{\underset{\sim}{\beta}}_R(k)$ และ $\hat{\underset{\sim}{\beta}}_L(d)$

เราสามารถเขียนสมาชิกตำแหน่งที่ i ของตัวประมาณในสมการที่ (2.3.3) และ (2.3.4) อยู่ในพจน์ของค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด ($\hat{\alpha}_i$) ดังนี้

(2.3.5) $$\hat{\alpha}_{R_i}(k) = \left(\frac{\lambda_i}{\lambda_i+k}\right)\hat{\alpha}_i$$

(2.3.6) $$\hat{\alpha}_{L_i}(d) = \left(\frac{\lambda_i+d}{\lambda_i+1}\right)\hat{\alpha}_i$$

จากสมการที่ (2.3.5) เราจะเห็นได้ว่า $\hat{\alpha}_{R_i}(0) = \hat{\alpha}_i$ และ $\hat{\alpha}_{R_i}(1) = \left(\frac{\lambda_i}{\lambda_i+1}\right)\hat{\alpha}_i$ และสมการที่ (2.3.6) จะได้ว่า $\hat{\alpha}_{L_i}(1) = \hat{\alpha}_i$ และ $\hat{\alpha}_{L_i}(0) = \left(\frac{\lambda_i}{\lambda_i+1}\right)\hat{\alpha}_i$ ดังนั้นทุก ๆ ค่าของ k โดยที่ $0<k<1$ เราสามารถหาทุก ๆ ค่าของ d โดยที่ $0<d<1$ ซึ่งทำให้ $\hat{\alpha}_{R_i}(k) = \hat{\alpha}_{L_i}(d)$ สำหรับในทางปฏิบัติแล้วการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีริดจ์รีเกรสชันจะได้ค่า k ที่เหมาะสมมีค่าอยู่ใกล้ 0 ซึ่งทำให้การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีลิว คีเจียนสามารถหาค่า d ที่ทำให้ $\hat{\alpha}_{R_i}(k)=\hat{\alpha}_{L_i}(d)$ แต่วิธีของลิว คีเจียนมีข้อได้เปรียบกว่าวิธีริดจ์รีเกรสชันกล่าวคือ $\hat{\alpha}_{L_i}(d)$ เป็นฟังก์ชันเชิงเส้นของ d ดังนั้นการคำนวนหาค่า d ที่เหมาะสมจึงสะดวกกว่าการคำนวนหาค่า k ที่เหมาะสมจากวิธีริดจ์รีเกรสชัน

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\hat{\underset{\sim}{\alpha}}_L(d)$ มีค่าเท่ากับ

$$MSE(\hat{\underset{\sim}{\alpha}}_L) = trace(Cov(\hat{\underset{\sim}{\alpha}}_L)) + (bias)^2$$

โดยที่ $Cov(\hat{\underset{\sim}{\alpha}}_L) = \sigma^2(\Lambda+I)^{-1}(\Lambda+dI)\Lambda^{-1}(\Lambda+dI)(\Lambda+I)^{-1}$

และ $bias = E(\hat{\underset{\sim}{\alpha}}_L)-\underset{\sim}{\alpha} = \left[(\Lambda+I)^{-1}(\Lambda+dI)\underset{\sim}{\alpha}-\underset{\sim}{\alpha}\right]$

(2.3.7) $$MSE(\hat{\underset{\sim}{\alpha}}_L) = \sigma^2\sum_{i=1}^{p}\frac{(\lambda_i+d)^2}{\lambda_i(\lambda_i+1)^2} + (d-1)^2\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+1)^2} = g(d)$$

$$g'(d) = 2\sigma^2\sum_{i=1}^{p}\frac{\lambda_i+d}{\lambda_i(\lambda_i+1)^2} + 2(d-1)\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+1)^2}$$

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

เราจึงสรุปได้ว่ามีค่าของ d โดยที่ $0<d<1$ ที่ทำให้ $g(d) < g(1)$ หรือกล่าวอีกนัยหนึ่งได้ว่า $MSE(\hat{\underset{\sim}{\alpha}}_L) < MSE(\hat{\underset{\sim}{\alpha}})$; $0<d<1$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากสมการที่ (2.3.7) เราจะเห็นได้ว่า MSE($\hat{\underset{\sim}{\alpha}}_L$) จะมีค่าต่ำสุดที่

$$d = \sum_{i=1}^{p} \frac{\alpha_i^2 - \sigma^2}{(\lambda_i+1)^2} \Bigg/ \sum_{i=1}^{p} \frac{\sigma^2 + \lambda_i \alpha_i^2}{\lambda_i(\lambda_i+1)^2}$$

เมื่อแทนค่าตัวประมาณที่ไม่เอนเอียงของ α_i^2 และ σ^2 ซึ่งก็คือ $\hat{\alpha}_i^2 - \frac{\hat{\sigma}^2}{\lambda_i}$ และ $\hat{\sigma}^2$ ตามลำดับ จะได้ค่าประมาณของ d อยู่ในรูปของ

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

ในปี ค.ศ. 1995 Fikri Akdeniz และ Selahattin Kaciranlar ได้เสนอวิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเป็นวิธีของลิว ดีเจียนที่อยู่ในรูปทั่วไปและตรวจสอบคุณสมบัติทางสถิติพร้อมทั้งการคำนวนหาค่า d_i ; $i=1,2,\ldots\ldots,p$ ที่เหมาะสม ซึ่งค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีลิว ดีเจียนทั่วไปอยู่ในรูปแบบดังนี้

(2.3.8) $$\hat{\underset{\sim}{\beta}}_L(D) = (X'X+I)^{-1}(X'\underset{\sim}{y}+D\hat{\underset{\sim}{\beta}})$$

โดยที่ D เป็นเมทริกซ์ขนาด p×p ซึ่งสมาชิกในแนวทแยงมุมเป็นค่าคงที่และสมาชิกนอกแนวทแยงมุมเป็น 0 กล่าวคือ $D = \mathrm{diag}(d_1, d_2, \ldots\ldots, d_p)$; $0<d_i<1$

จากสมการที่ (2.3.8) เราจะได้สัมประสิทธิ์การถดถอยพหุคูณวิธีลิว ดีเจียนทั่วไปอยู่ในรูปของ

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

ดังนั้น $\hat{\underset{\sim}{\beta}}_L(D)$ มีความเอนเอียงเท่ากับ $-(X'X+I)^{-1}(I-D)\underset{\sim}{\beta}$

เราจะเห็นได้ว่าความเอนเอียงจะลดลงเมื่อสมาชิกในแนวทแยงมุมของ D มีค่าสูงขึ้นและจะมีค่าเท่ากับศูนย์เมื่อ D=I โดยสมาชิกตำแหน่งที่ i ของ $\hat{\underset{\sim}{\beta}}_L(D)$ คือ

$$\hat{\beta}_{L_i} = \left[\frac{\lambda_i + d_i}{\lambda_i + 1} \right] \hat{\beta}_i = (1-\delta_i)\hat{\beta}_i$$

เมื่อ $$\delta_i = \frac{1-d_i}{\lambda_i+1} \quad ; \quad 0<\delta_i<1$$

โดยที่ λ_i คือ ค่าเฉพาะตัวที่ i ของเมทริกซ์ $X'X$

$\hat{\beta}_i$ คือ สมาชิกตำแหน่งที่ i ของสัมประสิทธิ์การถดถอยพหุคูณที่ได้จากวิธีกำลังสองน้อยสุด ($\hat{\underset{\sim}{\beta}}$)

และ $\hat{\beta}_{L_i}$ มีความเอนเอียงคือ $bias(\hat{\beta}_{L_i}) = -\delta_i \beta_i$

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\hat{\beta}_{L_i}$ จะอยู่ในรูปของ

$$MSE\left(\hat{\beta}_{L_i}\right) = \frac{\sigma^2(\lambda_i + d_i)^2 + \lambda_i(1-d_i)^2\beta_i^2}{\lambda_i(\lambda_i+1)^2}$$

$$= \frac{\sigma^2}{\lambda_i}(1-\delta_i)^2 + \delta_i^2\beta_i^2$$

ดังนั้น $MSE\left(\hat{\beta}_{L_i}\right)$ มีค่าต่ำสุดที่ $d_i = \dfrac{\lambda_i(\beta_i^2 - \sigma^2)}{\lambda_i\beta_i^2 + \sigma^2}$; $i=1,2,\ldots\ldots\ldots,p$

เนื่องจากในทางปฏิบัติแล้ว β_i และ σ^2 ไม่ทราบค่า เราจึงแทนพารามิเตอร์ทั้งสองด้วยตัวประมาณที่ไม่เอนเอียงคือ $\hat{\beta}_i$ และ $\hat{\sigma}^2$ ตามลำดับ ดังนั้นค่าประมาณของ d_i ที่เหมาะสมซึ่งทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุด คือ

$$\hat{d}_{i(opt)} = \frac{\lambda_i(\hat{\beta}_i^2 - \hat{\sigma}^2)}{\lambda_i\hat{\beta}_i^2 + \hat{\sigma}^2} \quad ; \quad i=1,2,\ldots\ldots,p$$