บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้อง

2.1 วิธีการถดถอยองค์ประกอบหลัก

วิธีการถดถอยองค์ประกอบหลัก เป็นวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ ที่ใช้ในการแก้ปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยทำการแปลงกลุ่มตัวแปรอิสระที่มีอยู่เดิมเป็นกลุ่มตัวแปรใหม่ที่ไม่มีพหุสัมพันธ์กัน เรียกกลุ่มตัวแปรใหม่ว่า “องค์ประกอบหลัก” (principals components) โดยทั่วไปกลุ่มตัวแปรใหม่จะมีจำนวนองค์ประกอบหลักน้อยกว่ากลุ่มตัวแปรเดิม เนื่องจากมีการจำกัดจำนวนองค์ประกอบที่แน่นอนซึ่งมีผลต่อการลดความแปรปรวน แต่ตัวประมาณที่ได้เป็นตัวประมาณที่เอนเอียง

ตัวแบบและวิธีการถดถอยองค์ประกอบหลัก

พิจารณาตัวแบบ

$$Y = X\beta + \varepsilon$$

กำหนดให้ $\lambda_1, \lambda_2, ..., \lambda_p$ เป็นค่าไอเกน (eigenvalues) ของ $X'X$ โดยที่ $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$ และ T เป็นเมทริกซ์เชิงตั้งฉากขนาด $p \times p$ ซึ่งแต่ละคอลัมน์ของเมทริกซ์เป็นเวกเตอร์เฉพาะที่สอดคล้องกับ $\lambda_1, \lambda_2, ..., \lambda_p$ ตามลำดับ และ $TT' = I$

$$Y = X(TT')\beta + \varepsilon$$

$$Y = (XT)(T'\beta) + \varepsilon$$

$$Y = Z\alpha + \varepsilon \qquad .... (2.1.1)$$

โดยที่ $Z = XT$ และ $\alpha = T'\beta$

$$Z'Z = T'X'XT = \Lambda = \text{diag}(\lambda_1, \lambda_2, ...\lambda_p)$$

Z เป็นเมทริกซ์ซึ่งแต่ละสดมภ์คือตัวแปรอิสระชุดใหม่ซึ่งเกิดจากผลรวมเชิงเส้น (linear combination) ของตัวแปรอิสระ แต่ละสดมภ์ประกอบไปด้วย

Z_1 คือ องค์ประกอบหลักที่ 1

Z_2 คือ องค์ประกอบหลักที่ 2

$\vdots$

Z_p คือ องค์ประกอบหลักที่ p

Λ เป็นเมทริกซ์ในแนวทแยงมุม (diagonal matrix) ที่มีขนาด $p \times p$ มีสมาชิกในแนวเส้นทแยงมุมเป็นค่าไอเกนของเมทริกซ์ $X'X$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \cdots & \lambda_p \end{bmatrix}$$

จากสมการ (2.1.1) เราสามารถประมาณค่าเวกเตอร์ของสัมประสิทธิ์การถดถอยแปลงค่า α โดยใช้วิธีกำลังสองน้อยที่สุด

$$\hat{\alpha} = (Z'Z)^{-1} Z'Y$$

$$\mathrm{Var}(\hat{\alpha}_j) = \frac{\sigma^2}{\lambda_j} \qquad ; j = 1,2........, p$$

เกณฑ์ที่ใช้ในการพิจารณาจำนวนองค์ประกอบหลักและการประมาณสัมประสิทธิ์การถดถอยองค์ประกอบหลัก

ในการวิเคราะห์องค์ประกอบหลัก เพื่อนำมาประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ เมื่อมีพหุสัมพันธ์ระหว่างตัวแปรอิสระจะนำองค์ประกอบมาใช้เพียงบางองค์ประกอบเท่านั้น เนื่องจากหลักการและความคิดพื้นฐานของการถดถอยองค์ประกอบหลักเหมือนกับหลักการของกำลังสองน้อยที่สุด คือเป็นการประมาณด้วยวิธีกำลังสองน้อยที่สุดที่ดำเนินการบนองค์ประกอบ

หลัก ดังนั้นถ้าองค์ประกอบหนึ่งถูกตัดทิ้ง ผลลัพธ์ของตัวประมาณของสัมประสิทธิ์การถดถอยของตัวแปรเดิม (X) จะเป็นตัวประมาณที่เอนเอียง ความแปรปรวนที่เกิดจากการสร้างองค์ประกอบนั้นๆ จะถูกตัดทิ้งด้วย ส่งผลให้ความแปรปรวนมีค่าลดลง การพิจารณาจำนวนองค์ประกอบที่นำมาใช้จึงเป็นเรื่องสำคัญ โดยเกณฑ์ที่ใช้ในการตัดสินใจเลือกองค์ประกอบเหล่านั้นมีอยู่หลายเกณฑ์ด้วยกัน ในการวิจัยนี้เราจะใช้ค่าเฉลี่ยของค่าไอเกนทั้งหมด $(\bar{\lambda})$ เป็นเกณฑ์

$$\bar{\lambda} = \frac{1}{p}\sum_{j=1}^{p}\lambda_j$$

โดยจำนวนองค์ประกอบหลักที่เราจะตัดออกจะมีค่าเท่ากับจำนวนของค่าไอเกนที่มีค่าน้อยกว่า $(\bar{\lambda})$ นั้นทิ้งไป จำนวนองค์ประกอบที่เหลือจะนำมาสร้างตัวแบบถดถอยด้วยวิธีกำลังสองน้อยที่สุด

สมมติให้มีค่าไอเกนจำนวน s ค่าสุดท้าย $\lambda_1 > \lambda_2 > \ldots > \lambda_{p-s} > \bar{\lambda} > \lambda_{p-s+1} \cdots > \lambda_p$ ที่มีค่าน้อยกว่าค่าไอเกนเฉลี่ย เราจะได้ค่าสัมประสิทธิ์การถดถอยองค์ประกอบหลัก $\hat{\alpha}_{PC}$ ซึ่งอยู่ในรูป $\hat{\alpha}$ ดังนี้

$$\hat{\alpha}_{PC} = B'\hat{\alpha}$$

$$b_1 = b_2 = \ldots = b_{p-s} = 1 \text{ และ } b_{p-s+1} = b_{p-s+2} = \ldots. = b_p = 0$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{p-s} \\ b_{p-s+1} \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\hat{\alpha}_{PC} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_{p-s} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \Lambda_{p-s}^{-1} T'_{p-s} X'Y$$

จากนั้นจึงใช้วิธีกำลังสองน้อยที่สุดประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ ได้ตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ โดยวิธีการถดถอยองค์ประกอบหลักดังนี้

$$\begin{aligned} \hat{\beta}_{PC} &= T_{p-s}\hat{\alpha}_{pc} \\ &= T_{p-s}(\Lambda_{p-s}^{-1} T'_{p-s} X'Y) \end{aligned}$$

$$Var(\hat{\beta}_{PC}) = \sigma^2 T_{p-s} \Lambda_{p-s}^{-1} T'_{p-s}$$

โดย $\hat{\beta}_{PC}$ เป็นตัวประมาณที่เป็นมาตรฐาน (Standardized Estimator) สมาชิกแต่ละตัวของ $\hat{\beta}_{PC}$ เรียกว่าสัมประสิทธิ์การถดถอยที่เป็นมาตรฐาน (Standardized Regression Coefficient) นอกจากนี้ $\hat{\beta}_{PC}$ ยังมีคุณสมบัติเป็นตัวประมาณที่เอนเอียง ซึ่งไม่มีการตัดตัวแปร X_j ใด ๆ ออกจากสมการแต่จะมีการตัด Z_j บางตัวออกแทน

ความเอนเอียงของสัมประสิทธิ์การถดถอยองค์ประกอบหลัก

เมื่อพิจารณาองค์ประกอบหลักที่ถูกตัดทิ้งไป s องค์ประกอบ และเหลืออยู่ p-s องค์ประกอบ โดยมีเมทริกซ์ของเวกเตอร์เฉพาะ $T = [T_1, T_2, ..., T_p]$ ซึ่งสามารถแบ่งส่วนออกเป็น

$$T = [T_{p-s} : T_s]$$

ในทำนองเดียวกัน พิจารณา Λ ซึ่งเป็นเมทริกซ์ในแนวทแยงมุมที่มีสมาชิกในแนวเส้นทแยงมุมเป็นค่าไอเกนของ $X'X$ สามารถแบ่งส่วนออกเป็น

$$\Lambda = \begin{bmatrix} \Lambda_{p-s} & 0 \\ 0 & \Lambda_s \end{bmatrix}$$

โดยที่ Λ_s และ Λ_{p-s} เป็นเมทริกซ์ซึ่งประกอบไปด้วยสมาชิกในแนวทแยงมุมเป็นค่าไอเกนที่สัมพันธ์กับองค์ประกอบที่ตัดทิ้ง และค่าไอเกนที่สัมพันธ์กับองค์ประกอบที่เลือกไว้ตามลำดับ

เมื่อเราตัดองค์ประกอบ s องค์ประกอบสุดท้ายทิ้งไป เราสามารถเขียนตัวประมาณกำลังสองน้อยที่สุดของสัมประสิทธิ์การถดถอยสำหรับพารามิเตอร์ p ตัวได้เป็น

$$\hat{\beta}_{PC} = \left[T_1\, T_2 \ldots T_{p-s}\right] \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_{p-s} \end{bmatrix}$$

หรือ $\hat{\beta}_{PC} = T_{p-s}\hat{\alpha}_{p-s}$

ดังนั้น $E\left[\hat{\beta}_{PC}\right] = T_{p-s}\alpha_{p-s} = T_{p-s}T'_{p-s}\beta$

เนื่องจาก $TT' = I = T_sT'_s + T_{p-s}T'_{p-s}$

จะได้ว่า

$$\begin{aligned} E\left[\hat{\beta}_{PC}\right] &= \left[I - T_sT'_s\right]\beta \\ &= \beta - T_sT'_s\beta \\ &= \beta - T_s\alpha_s \end{aligned}$$

ดังนั้นตัวประมาณสัมประสิทธิ์การถดถอยมีความเอนเอียงเท่ากับ $T_s\alpha_s$ โดยที่ α_s เป็นเมทริกซ์ขององค์ประกอบหลักที่ถูกตัดทิ้งไป

ความแปรปรวนของสัมประสิทธิ์การถดถอยองค์ประกอบหลัก

การตัดองค์ประกอบเป็นผลให้ความแปรปรวนของสัมประสิทธิ์การถดถอย $\hat{\beta}_{pc}$ มีค่าลดลงโดยสามารถพิจารณาได้จาก

$$\begin{aligned} \mathrm{Var}(\hat{\beta}) &= \sigma^2 (X'X)^{-1} \\ &= \sigma^2\, T\Lambda^{-1} T' \\ &= \sigma^2 (T_s\Lambda_s^{-1} T'_s + T_{p-s}\Lambda_{p-s}^{-1} T'_{p-s}) \end{aligned}$$

และ $\mathrm{Var}(\hat{\beta}_{PC}) = \sigma^2\, T_{p-s}\, \Lambda_{p-s}^{-1}\, T'_{p-s}$

จะสังเกตได้ว่า ความแตกต่างของเมทริกซ์ความแปรปรวนสำหรับตัวประมาณกำลังสองน้อยที่สุดและตัวประมาณองค์ประกอบหลักคือ $T_s\Lambda_s^{-1}T'_s$ โดยเมื่อตัดองค์ประกอบที่ไม่ต้องการออก จะทำให้ได้ความแปรปรวนที่มีค่าลดลง

2.2 วิธีริดจ์ รีเกรสชันที่มีค่าเบื้องต้น

วิธีริดจ์ รีเกรสชันที่มีค่าเบื้องต้น เป็นวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ กรณีตัวแปรอิสระมีพหุสัมพันธ์กัน โดยพัฒนามาจากวิธีริดจ์ รีเกรสชัน ที่เสนอโดยโฮเอร์ลและเคนนาร์ด เพื่อทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุมีค่าน้อยกว่าวิธีกำลังสองน้อยที่สุด หลักการของวิธีริดจ์ รีเกรสชันคือ เมื่อพบว่า $|X'X|$ มีค่าเข้าใกล้ศูนย์ จะมีผลให้สมาชิกใน $(X'X)^{-1}$ มีค่ามาก ทำให้ค่าประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุของวิธีกำลังสองน้อยที่สุดมีค่าสูงเกินความเป็นจริง ดังนั้นจึงพยายามปรับรูปเมทริกซ์ $X'X$ โดยการบวกค่าคงที่ที่มีค่ามากกว่าศูนย์กับสมาชิกทุกตัวในแนวทแยงมุมหลักของเมทริกซ์ $X'X$ จึงทำให้ค่าประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุวิธีริดจ์ รีเกรสชัน อยู่ในรูปของ

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y \qquad ;k > 0 \qquad \text{..............(2.2.1)}$$

การคำนวณหาค่า k ที่เหมาะสม ในสมการ (2.2.1) เป็นจุดอ่อนของวิธีริดจ์ รีเกรสชัน กล่าวคือ ค่า k ที่ใช้มีรูปแบบที่ไม่แน่นอน ดังนั้นในปี ค.ศ. 1995 ครอส, จิน และ ฮานูมารา ได้พัฒนาวิธีริดจ์ รีเกรสชัน ที่เสนอโดยโฮเอร์ลและเคนนาร์ด เรียกว่า วิธีริดจ์ รีเกรสชันที่มีค่าเบื้องต้น ซึ่งวิธีนี้ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์การถดถอย มีค่าน้อยกว่าวิธีกำลังสองน้อยที่สุด และตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุมีคุณสมบัติไม่เอนเอียง

ตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุของวิธีการนี้ คือ

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

เมื่อ J คือเวกเตอร์ของค่าประมาณเบื้องต้นของค่าเฉลี่ยสัมประสิทธิ์การถดถอยพหุจากวิธี

กำลังสองน้อยที่สุดคือ $J = \left(\sum_{j=1}^{p} \frac{\hat{\beta}_{LS_j}}{p} \right) 1_{p\times 1}$

ค่าประมาณ k แยกเป็น 2 กรณี คือ

กรณีทราบค่า σ^2

$$k = \begin{cases} \dfrac{p\sigma^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \sigma^2 tr(X'X)^{-1}} & \text{เมื่อ } (\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \sigma^2 tr(X'X)^{-1} > 0 \\ \dfrac{p\sigma^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J)} & \text{อื่นๆ} \end{cases}$$

กรณีไม่ทราบค่า σ^2

$$k = \begin{cases} \dfrac{p\hat{\sigma}^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \hat{\sigma}^2 tr(X'X)^{-1}} & \text{เมื่อ } (\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \hat{\sigma}^2 tr(X'X)^{-1} > 0 \\ \dfrac{p\hat{\sigma}^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J)} & \text{อื่น ๆ} \end{cases}$$

ในทางปฏิบัติจะใช้ค่า J ที่เป็นเวกเตอร์คงที่ (Fixed Vector) ไม่ใช้ J ที่เป็นเวกเตอร์สุ่ม (Random Vector) เนื่องจากขึ้นกับค่า k ซึ่งเป็นพารามิเตอร์ที่ต้องการประมาณ และค่า J ที่ครอส จินและฮานูมาราเสนอคือ $J = \left(\sum_{j=1}^{p} \dfrac{\hat{\beta}_{LS_j}}{p} \right) 1_{p\times 1}$ เหตุผลที่ใช้ค่าเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอยพหุจากวิธีกำลังสองน้อยที่สุด เนื่องจากจะให้เครื่องหมายของค่าประมาณสัมประสิทธิ์ที่ถูกต้อง

ตัวประมาณสัมประสิทธิ์การถดถอยพหุมีคุณสมบัติไม่เอนเอียง พิสูจน์ได้ดังนี้

จาก

$$\begin{aligned} \hat{\beta}_{RJ} &= (X'X + kI)^{-1}(X'Y + kJ) \\ &= (X'X + kI)^{-1}X'X(X'X)^{-1}(X'Y + kJ) \\ &= \left[I + k(X'X)^{-1} \right]^{-1}\left[\hat{\beta}_{LS} - J + J + k(X'X)^{-1}J \right] \\ &= \left[I + k(X'X)^{-1} \right]^{-1}\left[(\hat{\beta}_{LS} - J) + (J + k(X'X)^{-1}J) \right] \\ &= \left[I + k(X'X)^{-1} \right]^{-1} (\hat{\beta}_{LS} - J) + J \\ E(\hat{\beta}_{RJ}) &= \left[I + k(X'X)^{-1} \right]^{-1} E(\hat{\beta}_{LS} - J) + E(J) \end{aligned}$$

เนื่องจาก $E(J) = \beta \quad ; J \sim (\beta, \dfrac{\sigma^2 I}{k})$

ดังนั้น $E(\hat{\beta}_{RJ}) = \beta$

แผนภาพที่ 2.1 ขั้นตอนการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุของวิธีริดจ์ รีเกรสชันที่มีค่าเบื้องต้น

$$Y = X\beta + \varepsilon$$

$$\hat{\beta}_{LS} = (X'X)^{-1}X'Y, \hat{\sigma}^2 = \frac{Y'Y - \hat{\beta}_{LS}'X'Y}{n-p}$$

$$J = \left(\sum_{j=1}^{p} \frac{\hat{\beta}_{LS_j}}{p} \right) 1_{p\times 1}$$

$$k = \begin{cases} \dfrac{p\hat{\sigma}^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \hat{\sigma}^2 tr(X'X)^{-1}} & \text{if} \quad (\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J) - \hat{\sigma}^2 tr(X'X)^{-1} > 0 \\ \dfrac{p\hat{\sigma}^2}{(\hat{\beta}_{LS} - J)'(\hat{\beta}_{LS} - J)} & \text{other} \end{cases}$$

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

2.3 วิธีมูนิชและไคเบรียริดจ์ รีเกรสชัน

วิธีมูนิชและไคเบรียริดจ์ รีเกรสชัน เป็นวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ กรณีตัวแปรอิสระมีพหุสัมพันธ์กัน โดยพัฒนามาจากวิธีริดจ์ รีเกรสชัน ที่เสนอโดยโฮเอร์ลและเคนนาร์ด เพื่อทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุมีค่าน้อยกว่าวิธีกำลังสองน้อยที่สุด หลักการของวิธีริดจ์ รีเกรสชันคือ เมื่อพบว่า $|X'X|$ มีค่าเข้าใกล้ศูนย์ จะมีผลให้สมาชิกใน $(X'X)^{-1}$ มีค่ามาก ทำให้ค่าประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุของวิธีกำลังสองน้อยที่สุดมีค่าสูงเกินความเป็นจริง ดังนั้นจึงพยายามปรับรูปเมทริกซ์ $X'X$ โดยการบวกค่าคงที่ที่มีค่ามากกว่าศูนย์กับสมาชิกทุกตัวในแนวทแยงมุมหลักของเมทริกซ์ $X'X$ จึงทำให้ค่าประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุวิธีริดจ์ รีเกรสชัน อยู่ในรูปทั่วไป

$$\hat{\beta}_R = (X'X + kI)^{-1} X'Y \qquad ;k > 0$$

จากสมการ $Y = X\beta + \varepsilon$ นำมาเขียนในรูปคาโนนิคอล (Canonical form) ได้ดังนี้

$Y = X(TT')\beta + \varepsilon$

$Y = (XT)(T'\beta) + \varepsilon$

$Y = Z\alpha + \varepsilon$

โดย $Z = XT$ และ $\alpha = T'\beta$

จะได้ว่า $Z'Z = T'X'XT$

ตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุวิธีมูนิชและไคเบรียริดจ์ รีเกรสชัน คำนวณได้ดังนี้

จากสมการ $Y = Z\alpha + \varepsilon$ นำมาหาตัวประมาณค่าสัมประสิทธิ์การถดถอยได้ดังนี้

$$\hat{\alpha}(k) = (Z'Z + k_M I)^{-1} Z'Y$$

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}'Z'Y}{n-p}$$

โดย $\hat{\alpha} = (Z'Z)^{-1}Z'Y = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_6 \end{bmatrix}$

$t_{max} = \max\limits_{1 \le j \le p} (\lambda_j)$; $\lambda_1, \lambda_2, \ldots, \lambda_p$ คือ ค่าไอเกนของ $X'X$

และค่าประมาณ k_M คือ

$$k_M = \underset{1 \le j \le 6}{\text{median}} \left(\frac{1}{\sqrt{\dfrac{t_{max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_{max}\hat{\alpha}_j^2}}} \right)$$

จาก $\hat{\alpha}(k) = T'\hat{\beta}_{RM}$

ดังนั้น $\hat{\beta}_{RM} = T\hat{\alpha}(k)$

ตัวสถิติที่ใช้เป็นเกณฑ์ในการเปรียบเทียบ

เกณฑ์ที่ใช้ในการเปรียบเทียบว่าตัวประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุวิธีใดมีประสิทธิภาพมากที่สุด คือค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (Average mean square error : AMSE) และค่าความเอนเอียงของตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ โดยที่

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย AMSE $(\hat{\beta})$ มีรูปแบบดังนี้

$$MSE(\hat{\beta}_j) = \frac{1}{1000}\sum_{a=1}^{1000}(\hat{\beta}_{aj} - \beta_j)^2$$

$$AMSE(\hat{\beta}) = \frac{1}{7}\sum_{j=0}^{6} MSE(\hat{\beta}_j)$$

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยของสัมประสิทธิ์ของตัวแปรที่มีพหุสัมพันธ์กัน AMSE $(\hat{\beta}_{mcl})$ มีรูปแบบดังนี้

$$AMSE(\hat{\beta}_{mcl}) = \frac{1}{3}\sum_{j=0}^{3} MSE(\hat{\beta}_j) \qquad ;\ j = 1, 2, 3$$

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยของสัมประสิทธิ์ของตัวแปรที่ไม่มีพหุสัมพันธ์ AMSE $(\hat{\beta}_{ind})$ มีรูปแบบดังนี้

$$AMSE(\hat{\beta}_{ind}) = \frac{1}{3}\sum_{j=4}^{6} MSE(\hat{\beta}_j) \qquad ;\ j = 4, 5, 6$$

เมื่อ β แทนเวกเตอร์สัมประสิทธิ์การถดถอยเชิงเส้นพหุ

$\hat{\beta}$ แทนเวกเตอร์ของค่าประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุ

p แทนจำนวนตัวแปรอิสระ

ในการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุทั้ง 6 วิธีจะเลือกวิธีที่ให้ค่า AMSE $(\hat{\beta})$ ต่ำสุดในแต่ละกรณี

ค่าเฉลี่ยความเอนเอียงของตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ ABIAS $(\hat{\beta})$ มีรูปแบบดังนี้

$$\mathrm{BIAS}(\hat{\beta}_j) = \frac{1}{1000}\sum_{a=1}^{1000}(\hat{\beta}_{aj} - \beta_j)$$

$$\mathrm{ABIAS}(\hat{\beta}) = \frac{1}{7}\sum_{j=0}^{6}\left|\mathrm{BIAS}(\hat{\beta}_j)\right|$$

ค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ของตัวแปรที่มีพหุสัมพันธ์กัน ABIAS $(\hat{\beta}_{mcl})$ มีรูปแบบดังนี้

$$\mathrm{ABIAS}(\hat{\beta}_{mcl}) = \frac{1}{3}\sum_{j=0}^{3}\left|\mathrm{BIAS}(\hat{\beta}_j)\right| \qquad ; \ j = 1, 2, 3$$

ค่าเฉลี่ยของความเอนเอียงของสัมประสิทธิ์ของตัวแปรที่ไม่มีพหุสัมพันธ์ ABIAS $(\hat{\beta}_{ind})$ มีรูปแบบดังนี้

$$\mathrm{ABIAS}(\hat{\beta}_{ind}) = \frac{1}{3}\sum_{j=4}^{6}\left|\mathrm{BIAS}(\hat{\beta}_j)\right| \qquad ; \ j = 4, 5, 6$$

เมื่อ β_j แทนสัมประสิทธิ์การถดถอยเชิงเส้นพหุตัวที่ j

$\hat{\beta}_{aj}$ แทนค่าประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุตัวที่ j จากการประมาณครั้งที่ a

ในการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุทั้ง 6 วิธีจะเลือกวิธีที่ให้ค่า ABIAS $(\hat{\beta})$ ต่ำสุดในแต่ละกรณี

งานวิจัยที่เกี่ยวข้อง

โฮเอร์ลและเคนนาร์ค (Hoerl and Kennard, 1970) ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ เมื่อเกิดปัญหาตัวแปรอิสระมีความสัมพันธ์กัน $\left|X'X\right|$ จะมีค่าเข้าใกล้ศูนย์ซึ่งจะส่งผลให้ $(X'X)^{-1}$ มีค่ามากทำให้ค่าประมาณสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดมีค่าสูงเกินความเป็นจริง ดังนั้นถ้าบวกค่าคงที่ k ที่มากกว่าศูนย์เข้ากับสมาชิกทุกตัวบนเส้นทแยงมุมหลักของเมทริกซ์ $X'X$ จะทำให้ $|X'X+kI|$ มีค่าสูงขึ้น และทำให้ค่าประมาณสัมประสิทธิ์การถดถอยถูกต้องมากขึ้น และมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยที่สุด แต่ตัวประมาณค่าประมาณสัมประสิทธิ์การถดถอยที่ได้จะเอนเอียง

โฮเอร์ลและเคนนาร์ค (Hoerl and Kennard, 1970) ได้เสนอวิธีการประมาณค่าพารามิเตอร์ k ซึ่งวิธีการนี้ต้องทดลองกำหนดค่า k ที่เพิ่มจากศูนย์ทีละน้อย จนกว่าจะได้ค่า k ที่เหมาะสมที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์การถดถอยมีค่าน้อยกว่าวิธีกำลังสองน้อยที่สุด และทดลองเพิ่มค่า k ทีละน้อยอีกจนกว่าจะได้ค่า k ที่เหมาะสมที่สุด ซึ่งจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์การถดถอยมีค่าต่ำสุด และค่า k ที่เหมาะสมอยู่ในช่วง $0 \leq k \leq 1$

ครอส จิน และ ฮานูมารา(Crouse, Jin and Hanumara, 1995) ได้พัฒนาวิธีการริดจ์ รีเกรสชัน ที่ใช้ร่วมกับค่าเบื้องต้น ของสวินเดลให้มีประสิทธิภาพ โดยเสนอการประมาณค่า k ที่สามารถหาค่า k ที่แน่นอนได้ และค่าประมาณที่ได้มีความแกร่ง ซึ่งนำไปใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุในกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์กัน นอกจากนี้ยังได้เปรียบเทียบวิธีการประมาณค่าระหว่าง วิธีกำลังสองน้อยที่สุด, วิธีริดจ์ รีเกรสชัน ที่ประมาณค่า k โดย โฮเอร์ล เคนนาร์ค และบาร์ดวิน วิธีริดจ์ รีเกรสชัน ที่ใช้ร่วมกับค่าเบื้องต้น (J) เมื่อ J=0 และ $J=\left[\sum_{i=1}^{p}\frac{\hat{\beta}_{ls_i}}{p}\right]1_{p*1}$ พบว่าวิธีริดจ์ รีเกรสชัน จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยมีค่าน้อยกว่าวิธีกำลังสองน้อยที่สุด และเมื่อใช้ค่า $J=\left[\sum_{i=1}^{p}\frac{\hat{\beta}_{ls_i}}{p}\right]1_{p*1}$ จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์การถดถอยมีค่าต่ำที่สุด และตัวประมาณค่าที่ได้มีคุณสมบัติไม่เอนเอียง

ดุษฎีพรรณ วายุพักต์ (2524) ได้ศึกษา การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชัน โดยพิจารณาจากเกณฑ์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลจริง 5 ชุด เลือกเอาข้อมูลชุดที่ตัวแปรอิสระมีค่าสหสัมพันธ์ระหว่าง 0.7-0.99 เป็นส่วนมาก โดยผลการศึกษาเป็นดังนี้ วิธีริดจ์ รีเกรสชัน ให้ค่าประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุที่เอนเอียงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดเมื่อเปรียบเทียบกับวิธีกำลังสองน้อยที่สุด วิธีการของริดจ์ รีเกรสชัน เป็นวิธีการที่ยุ่งยากและเสียเวลาในการคำนวณค่าสถิติต่าง ๆ มาก แต่ถ้าข้อมูลที่นำมาศึกษาตัวแปรมีพหุสัมพันธ์กัน ควรใช้วิธีริดจ์รีเกรสชัน ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุดีกว่าวิธีกำลังสองน้อยที่สุด

ดวงพร ชูรักษ์ (2529) ได้ศึกษา การเปรียบเทียบการประมาณค่าในการวิเคราะห์การถดถอยเชิงเส้นพหุโดยวิธีริดจ์ รีเกรสชัน วิธีการถดถอยองค์ประกอบหลักและวิธีกำลังสองน้อยที่สุด ในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ผลการศึกษาเป็นดังนี้ ข้อมูลที่ใช้ในการศึกษาได้จากวิธีจำลองของ วินเคอร์น และ เชอร์ชิลล์ (Wichern and Churchill) ซึ่งจะได้ตัวแปรอิสระที่มีความสัมพันธ์กันในระดับต่าง ๆ เมื่อข้อมูลมีความแปรปรวนเป็น 0.01 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองโดยวิธีริดจ์ รีเกรสชัน และ วิธีการถดถอยองค์ประกอบหลัก จะมีค่ามากกว่าวิธีกำลังสองน้อยที่สุดเสมอไม่ว่าตัวแปรอิสระจะมีพหุสัมพันธ์กันมากหรือน้อย เมื่อความแปรปรวนของข้อมูลมีค่าเท่ากับ 1 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองโดยวิธีริดจ์ รีเกรสชัน วิธีการถดถอยองค์ประกอบหลัก จะมีค่าน้อยกว่าวิธีกำลังสองน้อยที่สุด ในกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์กันมาก แต่เมื่อความแปรปรวนของข้อมูลเท่ากับ 25 และ 100 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองโดยวิธีริดจ์รีเกรสชัน วิธีการถดถอยองค์ประกอบหลัก จะมีค่าน้อยกว่าวิธีกำลังสองน้อยที่สุดไม่ว่าตัวแปรอิสระจะมีพหุสัมพันธ์กันมากหรือน้อย

โอมาน (Oman, 1991) ได้เปรียบเทียบค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีกำลังสองน้อยที่สุด วิธีการถดถอยองค์ประกอบหลัก และวิธีสไตน์ โดยใช้เกณฑ์การเปรียบเทียบคือค่าพยากรณ์ของความคลาดเคลื่อนกำลังสอง ผลที่ได้คือการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุด้วยวิธีสไตน์ให้ค่าพยากรณ์ของความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยที่สุดในทุกกรณี ตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุโดยวิธีการถดถอยองค์ประกอบหลัก ให้ค่าพยากรณ์ของความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยที่สุด แต่เมื่อเลือกจำนวนองค์ประกอบ (Component) ไม่เหมาะสมกล่าวคือเลือกจำนวนองค์ประกอบน้อยเกินไปก็ทำให้ ค่าพยากรณ์ของความคลาดเคลื่อนกำลังสองมีค่ามากกว่าวิธีกำลังสองน้อยที่สุดได้

ธันยากร ต้นชลขันธ์ (2538) ได้ศึกษา การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชัน ซึ่งประมาณค่า k โดยวิธีค้นหาแบบลำดับ (Sequential Search) และวิธีที่ใช้หลักการของริดจ์และสไตน์ ในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ เมื่อการแจกแจงของความคลาดเคลื่อนเป็นแบบปกติ แบบปกติปลอมปน ในกรณีการแจกแจงแบบลอกนอร์มอล ที่ระดับความสัมพันธ์ 3 ระดับ คือ ระดับต่ำ ระดับปานกลาง และระดับสูง เกณฑ์ที่ใช้ในการเปรียบเทียบคือเปอร์เซ็นต์อัตราส่วนของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ผลการศึกษาเป็นดังนี้ วิธีกำลังสองน้อยที่สุด สามารถใช้ได้เมื่อระดับความสัมพันธ์ของตัวแปรอิสระอยู่ในระดับต่ำและปานกลาง เมื่อขนาดตัวอย่างมีขนาดใหญ่ และการกระจายของความคลาดเคลื่อนมีค่าน้อย วิธีริดจ์ รีเกรสชัน จะให้ผลดีเป็นส่วนใหญ่ และให้ผลดีมากขึ้นเมื่อระดับความสัมพันธ์ของตัวแปรอิสระเพิ่มขึ้น และจะเห็นผลชัดเจนเมื่อระดับความสัมพันธ์เท่ากับ 0.99 สำหรับทุกการแจกแจงแต่จะให้ผลดีลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีที่ใช้หลักการของริดจ์และสไตน์ จะให้ผลดีเมื่อขนาดตัวอย่างมาก และการกระจายของความคลาดเคลื่อนน้อย

กาญจนา ภัทราวิวัฒน์ (2539) ศึกษาการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชัน ที่มีค่าเบื้องต้นและวิธีเจเนรัลไลซ์ ลู กรณีเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยศึกษาภายใต้เกณฑ์ AMSE$(\hat{\theta})$ และ Bias$(\hat{\theta})$ และศึกษาลักษณะการแจกแจงของตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุทั้ง 3 วิธี ให้ความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติ และแบบดับเบิลเอ็กซ์โปเนนเชียล ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 0.1 และ 1.0 ผลการศึกษาภายใต้เกณฑ์ AMSE$(\hat{\theta})$ สรุปได้ว่า วิธีกำลังสองน้อยที่สุด มีประสิทธิภาพสูงสุดเมื่อระดับพหุสัมพันธ์ 0.3-0.7 ที่ขนาดตัวอย่างมากกว่าหรือเท่ากับ 25 ในทุกการแจกแจง ที่มีความแปรปรวนเป็น 1 หรือที่ทุกระดับพหุสัมพันธ์ เมื่อความแปรปรวนเป็น 0.1 วิธีริดจ์ รีเกรสชัน ที่มีค่าเบื้องต้น มีประสิทธิภาพสูงสุดเมื่อระดับพหุสัมพันธ์ 0.9 ที่มีความแปรปรวนเป็น 1 ขนาดตัวอย่างมากกว่าหรือเท่ากับ 20 หรือที่ความแปรปรวนเป็น 5 ในทุกขนาดตัวอย่าง วิธีเจเนรัลไลซ์ ลู มีประสิทธิภาพสูงสุดเมื่อระดับพหุสัมพันธ์ 0.3-0.7 ที่มีความแปรปรวนเป็น 1 ขนาดตัวอย่างมากกว่าหรือเท่ากับ 20 หรือที่ความแปรปรวนเป็น 5 ที่ขนาดตัวอย่างมากกว่าหรือเท่ากับ 20

สมพล จารุธนศักดิ์กูร (2539) ได้ศึกษา การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างวิธีกำลังสองน้อยที่สุด และวิธีริดจ์ รีเกรสชันที่ใช้ข้อสนเทศโดยหลักเกณฑ์ และวิธีลิวคีเจียนทั่วไป เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ เกณฑ์การเปรียบเทียบคือ อัตราส่วนของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ขนาดของตัวอย่างที่ใช้ในการวิจัยเท่ากับ 12,30,50 และ 100

ความคลาดเคลื่อนเป็นกลุ่มตัวอย่างที่สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ การแจกแจงแบบปกติปลอมปน และการแจกแจงแบบลอกนอร์มอล ระดับความสัมพันธ์ระหว่างตัวแปรอิสระเท่ากับ 0.30,0.60,0.90 และ 0.99 และจำนวนตัวแปรอิสระที่ศึกษาคือ 3 และ 5 ผลการวิจัยปรากฏว่า วิธีริดจ์ รีเกรสชัน ที่ใช้ข้อสนเทศโดยหลักเกณฑ์ มีประสิทธิภาพดีที่สุดทุกการแจกแจงของความคลาดเคลื่อนสุ่มยกเว้นการแจกแจงแบบลอกนอร์มอลในกรณีที่ระดับความสัมพันธ์เป็น 0.30 (ทุกขนาดตัวอย่าง) และ 0.60 (ขาดตัวอย่างเท่ากับ 30,50,100) เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 และระดับความสัมพันธ์เท่ากับ (0.30,0.30) (ขนาดตัวอย่างเท่ากับ 50 และ 100) เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 ปรากฏว่าวิธีกำลังสองน้อยที่สุดมีประสิทธิภาพดีที่สุด

กรรณิการ์ หิรัญกสิ (2540) ได้ศึกษา การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างวิธีกำลังสองน้อยที่สุดแบ่งส่วน วิธีการถดถอยองค์ประกอบหลัก และวิธีกำลังสองน้อยที่สุด เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ เกณฑ์การเปรียบเทียบคือ อัตราส่วนของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ของการพยากรณ์ จำนวนตัวแปรอิสระที่ใช้ในการวิจัยเท่ากับ 5,8 และ 12 การแจกแจงของความคลาดเคลื่อนเป็นแบบปกติซึ่งมีค่าเฉลี่ยเป็น 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.0, 3.0, 5.0, 7.0 และ 10.0 ระดับความสัมพันธ์มีค่าเท่ากับ 0.60,0.70,0.80,0.95,0.975 และ 0.99 ผลการวิจัยได้ข้อสรุปในแต่ละวิธีดังนี้ วิธีกำลังสองน้อยที่สุดมีประสิทธิภาพดีที่สุดในกรณีที่จำนวนตัวแปรอิสระมีความสัมพันธ์กันไม่มาก หรือไม่เกินครึ่งหนึ่งของจำนวนตัวแปรอิสระทั้งหมด ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 และในทุกระดับความสัมพันธ์ วิธีการถดถอยองค์ประกอบหลักมีประสิทธิภาพดีที่สุดในกรณีที่ตัวแปรอิสระทุกตัวมีความสัมพันธ์กันไม่ว่าจะเป็นพหุสัมพันธ์ 1 กลุ่มหรือ 2 กลุ่ม และระดับความสัมพันธ์ระหว่างตัวแปรอิสระมีค่าสูง ส่วนวิธีกำลังสองน้อยที่สุดแบ่งส่วนมีประสิทธิภาพดีที่สุดในกรณีที่ตัวแปรอิสระที่มีความสัมพันธ์กันมีจำนวนเกินครึ่งหนึ่งของจำนวนตัวแปรอิสระทั้งหมด ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 7.0 และ 10.0 และในบางกรณีซึ่งระดับความสัมพันธ์มีค่าเท่ากับ 0.95,0.975 และ 0.99 โดยที่ความแปรปรวนของความคลาดเคลื่อนเท่ากับ 5.0 ปัจจัยที่มีผลต่อค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์ ของวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในแต่ละวิธีเรียงตามลำดับจากมากไปน้อยคือ จำนวนตัวแปรอิสระและลักษณะความสัมพันธ์ของตัวแปรอิสระ ความแปรปรวนของความคลาดเคลื่อน และระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดยที่ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ของการพยากรณ์ ของแต่ละวิธีจะแปรผันตามหรือแปรผกผันกับระดับความสัมพันธ์ระหว่างตัวแปรอิสระในบางกรณี ทั้งนี้ขึ้นอยู่กับจำนวนตัวแปรอิสระที่มีความสัมพันธ์กันและลักษณะของความสัมพันธ์ของตัวแปรอิสระ

จีราพร ไทยถานันดร์ (2541) ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ 4 วิธี คือ วิธีกำลังสองน้อยที่สุด วิธีริดจ์รีเกรสชันที่มีค่าเบื้องต้น วิธีออลโมส อันไบแอส เจเนรัลไลซ์ ลู และวิธีมัลติเปิล ชริงค์เกจ ปรินซิเปิล คอมโปเนนท์ รีเกรสชัน เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าเฉลี่ยของค่าพยากรณ์ของความคลาดเคลื่อนกำลังสอง (APMSE) และค่าสัมบูรณ์ของความเอนเอียงของตัวประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ $\left|\text{Bias}(\hat{\theta}_i)\right|$ โดยจำนวนตัวแปรอิสระที่ใช้เท่ากับ 5 การแจกแจงของความคลาดเคลื่อนสุ่มเป็นแบบปกติ และแบบโลจิสติก ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 0.05 , 0.1 , 1 , 2 , 3 , 5 , 6 และ 10 ขนาดตัวอย่างเท่ากับ 20, 30 และ 50 ระดับความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละคู่มีระดับความสัมพันธ์ต่ำ ระดับความสัมพันธ์ปานกลาง ระดับความสัมพันธ์สูง ผลการวิจัยเมื่อพิจารณาเกณฑ์ APMSE เมื่อความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติ ในกรณีที่ความแปรปรวนของความคลาดเคลื่อนสุ่มมีค่าเป็น 2 , 3 , 5 , 6 และ 10 ความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละคู่มีระดับความสัมพันธ์ต่ำ ระดับความสัมพันธ์ปานกลาง ระดับความสัมพันธ์สูง ทุกขนาดตัวอย่างที่ศึกษา พบว่าวิธีริดจ์ รีเกรสชันที่มีค่าเบื้องต้น ให้ค่า APMSE ต่ำสุด เมื่อความคลาดเคลื่อนสุ่มมีการแจกแจงแบบโลจิสติก วิธีริดจ์รีเกรสชันที่มีค่าเบื้องต้น วิธีมัลติเปิล ชริงค์เกจ ปรินซิเปิล คอมโปเนนท์ รีเกรสชัน ให้ค่า APMSE ต่ำสุดในกรณีเดียวกันกับการที่ความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติ เมื่อพิจารณาเกณฑ์ $\left|\text{Bias}(\hat{\theta}_i)\right|$ พบว่าวิธีกำลังสองน้อยที่สุดให้ค่า $\left|\text{Bias}(\hat{\theta}_i)\right|$ ต่ำสุดเป็นส่วนใหญ่

วราภรณ์ บุญยไพศาลเจริญ (2546) ได้ศึกษา การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยการเปรียบเทียบวิธีการถดถอยองค์ประกอบหลัก(PC) วิธีการถดถอยแบบรากแฝง(LR) และวิธีการประมาณของลิวเมื่อมีข้อจำกัด(RL) ซึ่งเกณฑ์การเปรียบเทียบคือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยของสัมประสิทธิ์การถดถอยพหุคูณ ขนาดตัวอย่างที่ใช้ในการวิจัยเท่ากับ 15, 20, 30, 40, 50, 60, 70, 80, 90 และ 100 ความคลาดเคลื่อนมีการแจกแจงแบบปกติ ค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.0, 5.0 และ 10.0 ตามลำดับ ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ เท่ากับ 0.10, 0.30, 0.50, 0.70, 0.80 และ 0.90 ผลการวิจัยปรากฏว่าระดับความสัมพันธ์ ขนาดตัวอย่าง จำนวนตัวแปรอิสระ และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนต่างมีผลต่อการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณของทั้งสามวิธี โดยค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยของสัมประสิทธิ์การถดถอยพหุคูณมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น แต่มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์ และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน มีค่ามากขึ้น หรือ

จำนวนตัวแปรอิสระมีจำนวนมากขึ้น กรณีข้อมูลสอดคล้องกับข้อจำกัด ในทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง วิธีการประมาณของลิวเมื่อมีข้อจำกัดจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด ยกเว้นกรณีที่ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีขนาดเล็ก(1.0) วิธีการถดถอยองค์ประกอบหลัก วิธีการถดถอยแบบรากแฝงและวิธีการประมาณของลิวเมื่อมีข้อจำกัด จะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยไม่แตกต่างกัน กรณีข้อมูลไม่สอดคล้องกับข้อจำกัด ในทุกกรณีของระดับความสัมพันธ์ ขนาดตัวอย่าง ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน และจำนวนตัวแปรอิสระ วิธีการถดถอยองค์ประกอบหลักเป็นวิธีที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยของสัมประสิทธิ์การถดถอยพหุคูณต่ำที่สุด

เปรมวดี ชูไสว (2548) ได้ศึกษาปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ เมื่อเกิดพหุสัมพันธ์ ระหว่างตัวแปรอิสระ วิธีการเปรียบเทียบที่นำมาพิจารณาคือ วิธีกำลังสองน้อยสุด วิธีริดจ์ รีเกรสชัน วิธีการถดถอยองค์ประกอบหลัก โดยเกณฑ์เปรียบเทียบคือ อัตราส่วนของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอย โดยจำนวนตัวแปรอิสระที่ใช้เท่ากับ 3,6 และ 9 มีส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนเท่ากับ 1,5 และ 10 ขนาดตัวอย่างเท่ากับ 5p, 10p, 15p, 20p, 25p และ 30p เมื่อ p คือจำนวนตัวแปรอิสระโดยแบ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระ เป็น น้อย (0.15-0.30) ปานกลาง (0.31-0.5,0.51-0.65) และมาก (0.66-0.85,0.86-0.99) ผลที่ได้จากการวิจัย พบว่า มากกว่า 99% ของสถานการณ์สามารถสรุปได้ว่าในระดับความสัมพันธ์มาก ช่วง 0.86-0.99 นั้นจะทำให้เกิดปัญหาพหุสัมพันธ์สูงจนส่งผลกระทบต่อการประมาณค่าด้วยตัวประมาณแบบวิธีกำลังสองน้อยสุด ดังนั้นวิธีการถดถอยองค์ประกอบหลักจึงเป็นวิธีที่มีประสิทธิภาพที่สุด ในทุกกรณี ในส่วนระดับความสัมพันธ์อื่น นั้นจะต้องพิจารณาค่าส่วนเบี่ยงเบนของค่าคลาดเคลื่อนประกอบด้วย ถ้าส่วนเบี่ยงเบนของค่าคลาดเคลื่อน เท่ากับ 1 วิธีกำลังสองน้อยสุดจะมีประสิทธิภาพมากสุด และในส่วนของการเบี่ยงเบนค่าคลาดแคลื่อนเท่ากับ 5 และ 10 วิธีกำลังสองน้อยสุด หลังจากที่ทำการแก้ไขให้ x เป็นแกนตั้งฉากซึ่งกันและกันแล้ว จะเป็นวิธีที่ดีที่สุด ปัจจัยที่มีผลต่อความคลาดเคลื่อนนั้น พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น AMSE ที่ได้จะให้ค่าลดลง ตรงกันข้าม ถ้าระดับความสัมพันระหว่างตัวแปรอิสระ และส่วนเบี่ยงเบนของค่าคลาดเคลื่อนเพิ่มขึ้น AMSE ที่ได้จะมีค่าเพิ่มขึ้น ยกเว้นกรณีวิธีการถดถอยองค์ประกอบหลัก ที่ เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระช่วง 0.86-0.99 จะน้อยกว่า ค่า AMSE ในระดับอื่น