การวิเคราะห์การถดถอยด้วยวิธีของริดจ์

ชิดชนก เชิงเชาว์

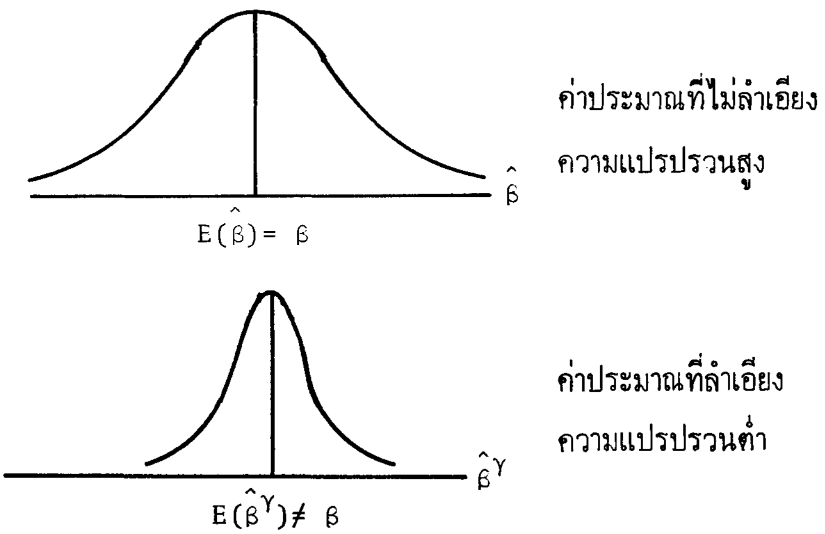
ในการวิเคราะห์การถดถอยโดยทั่ว ๆ ไปแล้ว วิธียกกำลังสองอย่างน้อยที่สุด (least squares method) เป็นวิธีมาตรฐานในการประมาณค่าสัมประสิทธิ์การถดถอย (regression coefficient) ของประชากร เนื่องจากวิธีการดังกล่าวมีคุณสมบัติที่ไม่ลำเอียง (คือค่าเฉลี่ยของค่าประมาณมีค่าเท่ากับค่าของประชากร) และมีค่าความแปรปรวนน้อยที่สุดในกลุ่มค่าประมาณที่ไม่ลำเอียงด้วยกัน

ดังนั้นวิธีการยกกำลังสองอย่างน้อยที่สุดจึงมีคุณสมบัติที่เรียกสั้นๆ ว่า BLUE (Best Linear Unbiased Estimator) แต่ในการวิเคราะห์การถดถอยพหุคูณ (Multiple regression analysis) บางครั้งมีปรากฏการณ์ที่เรียกว่า มัลติคอลลิเนียริตี้ (multicollinearity) เกิดขึ้น กล่าวคือ ตัวแปรอิสระหรือตัวทำนายบางตัวมีลักษณะเกือบจะเป็นผลรวมเชิงเส้นตรง (linear combination) ของตัวทำนายตัวอื่น ๆ แต่ถ้าตัวทำนายบางตัวเป็นผลรวม เชิงเส้นตรงที่สมบูรณ์ของตัวทำนายอื่น ๆ แล้ว ค่าประมาณของสัมประสิทธิ์การถดถอยโดยวิธียกกำลังสองอย่างน้อยที่สุด (ต่อไปนี้จะใช้คำว่า วิธีมาตรฐาน) จะไม่สามารถคำนวณได้ ในกรณีของมัลติคอลลิเนียริตี้นั้น ถึงแม้ว่าจะเป็นไปได้ในการใช้วิธีมาตรฐานคำนวณหาค่าประมาณของสัมประสิทธิ์การถดถอย (ต่อไปนี้จะใช้แต่เพียงว่า ค่าประมาณ) แต่ค่าประมาณโดยวิธีดังกล่าวจะมีความแปรปรวนสูง กล่าวคือ ค่าประมาณโดยวิธีกำลังสองน้อยที่สุดจะมีค่าเปลี่ยนแปลงอย่างมากขึ้นกับกลุ่มตัวอย่างที่เปลี่ยนไป ในกรณีเช่นนี้ตัวประมาณที่ลำเอียงอาจเข้ามาแทนที่วิธีการยกกำลังสองอย่างน้อยที่สุด เนื่องจากตัวประมาณที่ลำเอียงมีความแปรปรวนน้อยกว่าตัวประมาณโดยวิธียกกำลังสองน้อยที่สุดนั่นเอง

การวิเคราะห์การถดถอยโดยวิธีของริดจ์ (Ridge regression analysis) เป็นการหาค่าประมาณของสัมประสิทธิ์การถดถอยที่ลำเอียง ซึ่งได้รับการพัฒนาขึ้นมาเพื่อแก้ไขปัญหาทางด้านมัลติคอลลิเนียริตี้โดยเฉพาะ เนื่องจากตัวประมาณที่ลำเอียงมีความแปรปรวนน้อย ในบางสถานการณ์อาจมีโอกาสที่เข้าใกล้ค่าประชากรที่แท้จริงมากกว่าค่าประมาณที่ไม่ลำเอียง (โปรดสังเกตภาพที่ 1 ประกอบ)

ภาพที่ 1

ค่าประมาณที่ลำเอียงและไม่ลำเอียง



ค่าของความลำเอียง (biased) ของค่าประมาณที่ลำเอียงคำนวณได้จากระยะทางระหว่างค่าเฉลี่ยของค่าประมาณที่ลำเอียงและค่าของประชากรที่แท้จริง $[E(\hat{\beta}^{r}) - \beta]$ นั่นเอง

เกณฑ์ที่นิยมใช้ในการตัดสินคุณภาพของตัวประมาณเรียกว่า ค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง (Mean Squared Error) หรือเรียกย่อ ๆ ว่า MSE ค่าประมาณที่มีค่า MSE น้อย อาจกล่าวได้ว่า มีคุณสมบัติในการเป็นตัวประมาณที่ดีกว่าค่าประมาณที่มีค่า MSE มาก ค่าของ MSE จากตัวประมาณใด ๆ (สมมติแทนค่าด้วย β^*) คำนวณได้จาก

$$\begin{aligned} \text{MSE}\ (\beta^*) &= \text{Var}\ (\beta^*) + [E(\beta^*) - \beta]^2 \\ &= \text{Var}\ (\beta^*) + (\text{bias})^2 \end{aligned}$$

จะเห็นได้ว่า ค่า MSE ของ β^* คือผลบวกระหว่างความแปรปรวนของ β^* และค่าความลำเอียงของ β^* ยกกำลังสองนั่นเอง เนื่องจากวิธีการยกกำลังสองอย่างน้อยที่สุดเป็นตัวประมาณที่ไม่ลำเอียง (ค่าความลำเอียง = 0) MSE ของตัวประมาณโดยวิธียกกำลังสองน้อยที่สุด ($\hat{\beta}$) จึงมีค่าเท่ากับความแปรปรวนของตัวมันเองเพียงอย่างเดียว

โมเดลมาตรฐานของการวิเคราะห์การถดถอย

โมเดลมาตรฐานของการถดถอยอาจเขียนได้ในรูป

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

สมมติว่า จำนวนตัวทำนายทั้งหมดมี = p ตัว โดยมีขนาดของกลุ่มตัวอย่าง = n

ดังนั้น Y คือ เวคเตอร์ขนาด $(n \times 1)$ ของตัวแปรตามที่วัดได้จากกลุ่มตัวอย่างขนาด n

X คือ แมตริกซ์ขนาด $(n \times p)$ แสดงค่าตัวทำนายทั้ง p ตัวที่วัดได้จากกลุ่มตัวอย่างขนาด n

ε คือ เวคเตอร์ขนาด $(n \times 1)$ ของความคลาดเคลื่อนของตัวแปรตาม โดยมีคุณสมบัติ

$$E(\varepsilon) = 0 \text{ และ } E(\varepsilon\varepsilon') = \sigma^2$$

β คือ เวคเตอร์ขนาด $(p \times 1)$ ของสัมประสิทธิ์การถดถอยของประชากร

ค่าประมาณและความแปรปรวนของค่าประมาณโดยวิธีมาตรฐาน (วิธียกกำลังสองอย่างน้อยที่สุด)

คำนวณได้จาก

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

$$\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad \text{———————} \quad (2)$$

ในกรณีที่ความคลาดเคลื่อนเป็นอิสระต่อกัน และมีการกระจายเป็นรูปโค้งปกติ ตัวประมาณที่ไม่ลำเอียง ซึ่งคำนวณโดยวิธียกกำลังสองอย่างน้อยที่สุด และวิธีคล้ายคลึงมากที่สุด (maximum Likelihood estimation) จะได้ผลเหมือนกันดังที่ปรากฏในสมการที่ 1 โดยที่ผลรวมของความคลาดเคลื่อนยกกำลังสองจะมีค่าน้อยที่สุด กล่าวคือ

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

เมื่อ $\phi(\hat{\beta})$ คือ ผลรวมของความคลาดเคลื่อนยกกำลังสอง เมื่อใช้ $\hat{\beta}$ เป็นตัวประมาณ

ในสถานการณ์ที่มัลติคอลลิเนียริตี้เกิดขึ้น คือ X ซึ่งอยู่ในรูปของแมตริกซ์สหสัมพันธ์ (correlation matrix) มีความเบี่ยงเบนออกจาก identity matrix ค่าที่ได้รับผลกระทบกระเทือน คือ ระยะทางระหว่างเวคเตอร์ของตัวประมาณ ($\hat{\beta}$) และเวคเตอร์ของค่าประชากร (β) รวมทั้งค่าความแปรปรวนของค่าประมาณด้วย กล่าวคือ

ให้ L = ระยะทางระหว่าง β และ $\hat{\beta}$

$$L^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta)$$

จะได้ว่า

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

และ

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

เมื่อ ε กระจายเป็นรูปโค้งปกติ

$$\text{Var } (L^2) = 2\sigma^2 \text{trace } (X'X)^{-2}$$

ถ้า eigenvalue ของ $X'X$ กำหนดโดย

$$\lambda\text{max} = \lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \ . \ . \ . \geqslant \lambda_p = \lambda\text{min} > 0$$

ค่าเฉลี่ยของระยะทางยกกำลังสองระหว่าง $\hat{\beta}$ และ β กำหนดโดย

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

และ

$$\text{Var}(L^2) = 2\sigma^2 \sum_{i=1}^{p} (1/\lambda_i)^2$$

ดังนั้น ค่าต่ำสุดที่เป็นไปได้ของ $E(L^2)$ และ $\text{Var}(L^2)$ คือ $\sigma^2/\lambda\text{min}$ และ $2\sigma^2/\lambda\text{max}$ ตามลำดับ ในกรณีที่มัลติคอลลิเนียริตี้เกิดขึ้น ค่าของ λmin จะมีค่าต่ำมาก ดังนั้น ระยะทางระหว่าง $\hat{\beta}$ และ β จะห่างกันมากขึ้น เป็นเหตุให้การประมาณค่าของ β โดยวิธีมาตรฐานคลาดเคลื่อนจากความเป็นจริงเป็นอย่างมาก

เฮิร์ลและเคนนาร์ด (Hoerl and Kennard, 1970a) เสนอว่า ในกรณีที่มีมัลติคอลลิเนียริตี้เกิดขึ้น ควรใช้วิธีของริดจ์แทนวิธีมาตรฐาน โดยมีวิธีการคำนวณคือ

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

และ $\mathrm{Var}\,(\hat{\beta}^{r}) = \sigma^{2}\, Z\,(X'X)^{-1}Z'$

เมื่อ Ip = identity matrix ขนาด $(p \times p)$

k = เป็นค่าคงที่ ซึ่ง > 0

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

$\hat{\beta}^{r}$ = ค่าประมาณของสัมประสิทธิ์การถดถอยโดยวิธีของริดจ์นั่นเอง

ความสัมพันธ์ระหว่างค่าประมาณโดยวิธีริดจ์และวิธีมาตรฐาน คือ

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

ดังนั้น จะเห็นได้ว่า ถึงแม้ว่าไม่เกิด มัลติคอลลิเนียริตี้ ค่าประมาณโดยวิธีริดจ์ก็ไม่เท่ากับค่าประมาณโดยวิธีมาตรฐานอยู่ดี

แฟเดน (Faden, 1978) แสดงให้เห็นว่า ในกรณีที่ไม่มีมัลติคอลลิเนียริตี้ ตัวประมาณโดยวิธีริดจ์ คือ

$$\hat{\beta}^{r} = \begin{bmatrix} 1/k' + 1 & & & \\ & 1/k' + 1 & & \\ & & \ddots & \\ & & & 1/k' + 1 \end{bmatrix}\hat{\beta}$$

เมื่อ $k' = k/n$ และสมาชิกของแมตริกซ์ที่ไม่ได้อยู่บนแนวเส้นทแยงมุมเดียวกับ $1/k + 1$ มีค่าเป็น o หมดทุกตัว

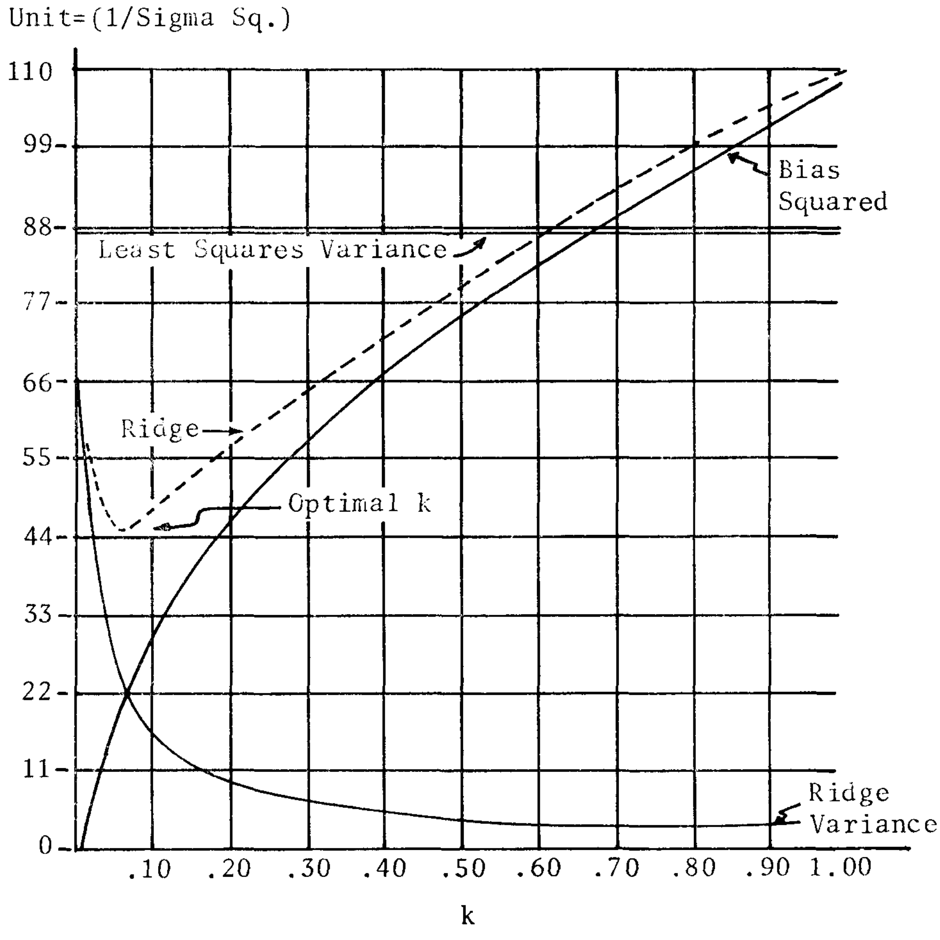
ความสัมพันธ์ระหว่างความลำเอียง ความแปรปรวนของค่าประมาณ ค่า K และ MSE

เฮิร์ลและเคนนาร์ด (Hoerl and Kennard, 1970 a) ได้แสดงให้เห็นความสัมพันธ์ระหว่างความแปรปรวนของค่าประมาณ ความลำเอียง และค่า k โดยใช้ข้อมูลจริงของกอร์แมนและโทแมน (Gorman and Toman, 1966) ความสัมพันธ์ดังกล่าวแสดงได้โดยภาพที่ 2

ภาพที่ 2

ตัวแปรต่าง ๆ ที่มีผลต่อ MSE

(จาก Hoerl and Kennard, 1970 a)



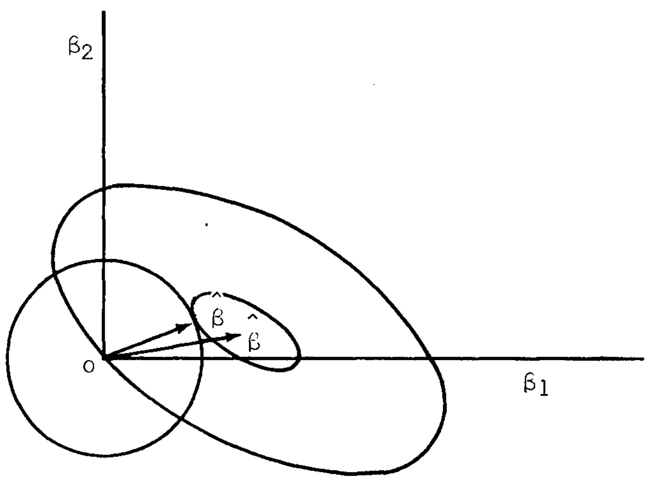
จากภาพที่ 2 แสดงให้เห็นถึงความแปรปรวนของตัวประมาณ โดยวิธีมาตรฐานซึ่งมีค่าคงที่คือ เป็นอิสระจากค่า k นั่นคือ ไม่มีความลำเอียงนั่นเอง ส่วนความแปรปรวนของค่าประมาณโดยวิธีริดจ์มีค่าลดลงเรื่อย ๆ ตามค่าของ k ที่เพิ่มขึ้นเรื่อย ๆ แต่ค่ายกกำลังสองของความลำเอียงโดยวิธีของริดจ์กลับแปรผันตรงกับค่า k กล่าวคือ ถ้าค่า k เพิ่มขึ้น ความลำเอียงจะเพิ่มขึ้นด้วย ส่วน MSE ของตัวประมาณโดยวิธีของริดจ์นั้นมีเท่ากับผลบวกของความแปรปรวนและค่ายก

กำลังสองของความลำเอียง เมื่อ k = O ค่า MSE ของริดจ์มีค่าเท่ากับ MSE ของวิธีมาตรฐาน (ซึ่งเท่ากับความแปรปรวนนั่นเอง) เมื่อ k มีค่าเพิ่มขึ้น MSE จะลดลงเรื่อย ๆ จนถึงจุด ๆ หนึ่ง (สมมุติว่าเป็น k_i) ค่า MSE จะเริ่มเพิ่มขึ้นอีกจนสูงกว่า MSE โดยวิธีมาตรฐาน ดังนั้น ค่า k_i จึงเป็นค่าที่เหมาะสมที่สุดในวิธีการของริดจ์ ดังนั้น อาจสรุปได้ว่าในกรณีที่เราใช้ $k > O$ ใส่ลงไปในสมการการถดถอยวิธีมาตรฐานเพื่อเป็นการลดค่าความแปรปรวน โดยยอมรับความลำเอียงที่เพิ่มขึ้น อาจทำให้สามารถลดค่า MSE ได้ซึ่งจะเป็นผลให้ได้ตัวประมาณใหม่ ซึ่งมีค่าใกล้เคียงกับค่าประชากรมากกว่าตัวประมาณโดยวิธีมาตรฐานได้นั่นเอง

ภาพที่ 3

รูปทรงทางเรขาคณิตของตัวประมาณโดยวิธีของริดจ์

(จาก Marquardt and Snee, 1975)



รูปทรงทางเรขาคณิตของค่าประมาณโดยวิธีริดจ์

ภาพที่ 3 แสดงรูปทรงทางเรขาคณิตของค่าประมาณริดจ์ ในกรณีที่มีค่าประชากร 2 ตัว คือ β_1 และ β_2 จุดกลางของรูปเอลลิปส์ เป็นตำแหน่งซึ่งผลรวมของความคลาดเคลื่อนยกกำลังสองมีค่าน้อยที่สุด ดังนั้นจึงเป็นตำแหน่งของค่าประมาณโดยวิธียกกำลังสองน้อยที่สุด (แทนด้วยเวคเตอร์ $\hat{\beta}$) ส่วนรูปเอลลิปส์เล็กคือ โลกัสของค่าประมาณโดยวิธีอื่น ๆ ซึ่งทำให้ค่าผลรวมของความคลาดเคลื่อนยกกำลังสองน้อยที่สุดมีค่าคงที่ค่าใดค่าหนึ่ง ซึ่งสูงกว่าค่าที่ได้จากวิธียกกำลังสอง

น้อยที่สุดเสมอ ส่วนรูปวงกลมรอบจุดตัดระหว่างแกน x และ y จะสัมผัสกับรูปเอลลิปส์เล็กที่จุดซึ่งเป็นตำแหน่งของเวคเตอร์ของค่าประมาณโดยวิธีของริดจ์ จากภาพดังกล่าวอาจสรุปง่าย ๆ ได้ว่าถ้ากำหนดให้ค่าผลรวมของความคลาดเคลื่อนยกกำลังสอง มีค่าเท่ากับค่าคงที่ใด ๆ (ซึ่งสูงกว่าค่าต่ำสุดที่ได้จากวิธียกกำลังสองน้อยที่สุด) แล้ว วิธีการของริดจ์จะได้ค่าประมาณของสัมประสิทธิ์การถดถอยที่มีค่าต่ำที่สุดนั่นเอง

โมเดลของริดจ์

ค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง (Mean squared error หรือ MSE) ของค่าประมาณโดยวิธีของริดจ์อยู่ในรูปแบบต่อไปนี้

กำหนดให้ p เป็น orthogonal matrix ซึ่งทำให้

$$P'X'XP = D$$

เมื่อ D เป็น diagonal matrix ซึ่งค่าสมาชิกที่ i มีค่าเท่ากับ λ_i ค่า MSE ของ $\hat{\beta}^r$ จึงถูกกำหนดโดย

$$\text{MSE}\ (\hat{\beta}^r) = E[(\hat{\beta}^r - \beta)'\ (\hat{\beta}^r - \beta)] \qquad (3)$$

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

เมื่อ $\alpha = p\beta$ หรืออาจกล่าวว่า α คือ β ที่อยู่ในรูปแบบของคานอนิคอล (canonical form) นั่นเอง

เทอมแรกของสมการที่ (3), $\sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2}$ คือค่าความแปรปรวนของ $\hat{\beta}^r$ ซึ่งจะมีค่าลดลง ถ้าค่า k เพิ่มขึ้น ส่วนเทอมที่สอง, $k^2 \sum_{i=1}^{p} \frac{\alpha_i^2}{\alpha_i+k}$ คือค่ายกกำลังสองของความลำเอียงซึ่งจะมีค่าเพิ่มขึ้นถ้าค่า k มีค่าลดลง ดังนั้นค่า MSE ของ $\hat{\beta}^r$ จะมีค่าน้อยกว่า MSE ของ $\hat{\beta}$ ถ้าสามารถหาจุดสมดุลระหว่างสองเทอมดังกล่าวได้

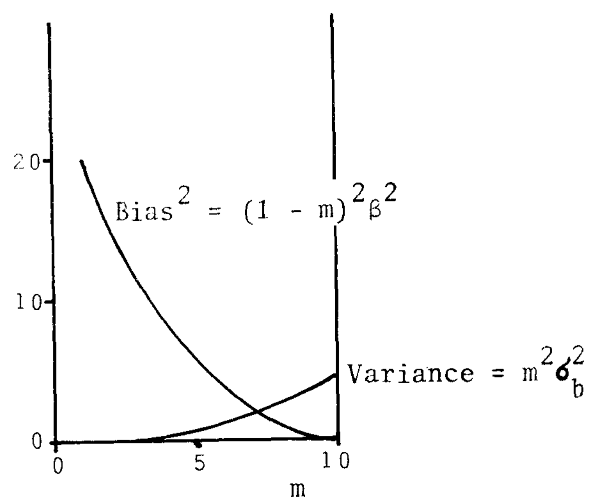
ดาร์ลิงตัน (Darlington, 1978) กล่าวว่า ตัวประมาณที่ลำเอียงสามารถคำนวณได้จากการคูณตัวประมาณที่ไม่ลำเอียงด้วยค่าคงที่ ในกรณีของการวิเคราะห์การถดถอยอย่างง่าย สมมติให้ b เป็นค่าตัวประมาณที่ไม่ลำเอียงของ β และให้ m เป็นค่าคงที่ใด ๆ ดังนั้น mb จะเป็นค่าประมาณที่ไม่ลำเอียงของ β ได้ก็ต่อเมื่อ m = 1.000 เท่านั้น อย่างไรก็ตามค่า MSE ของ mb จะ

มีค่าน้อยลง ถ้า $m < 1.000$ โดยจะมีค่าน้อยที่สุดถ้า $m = \beta^2/\beta^2 + \sigma_b^2$ เมื่อ σ_b^2 เป็นค่าความแปรปรวนของ b ดังนั้น เมื่อค่า $0 < m < 1.0$ จะทำให้ $0 \leqslant \left| b^r = mb \right| \leqslant b$ ค่าประมาณโดยวิธีของริดจ์จึงเป็นการลดค่าของค่าประมาณที่ไม่ลำเอียงเข้าหา o โดยลดลงมากหรือน้อยเพียงไรขึ้นกับค่าของ σ_b^2 ในกรณีที่มีตัวทำนายมากกว่า 1 ตัว ค่า m ของตัวประมาณแต่ละตัวจะต้องคำนวณด้วยวิธีซึ่งซับซ้อนมากขึ้น เนื่องจากมีค่าสหสัมพันธ์ระหว่างตัวประมาณต่าง ๆ เข้ามาเกี่ยวข้องด้วยนั่นเอง

ภาพที่ 4

อิทธิพลของ m ที่มีต่อความแปรปรวนและความลำเอียงของ b

(จาก Darlington, 1978)



ภาพที่ 4 แสดงให้เห็นถึงค่า m ที่สามารถลดค่า MSE ให้ลดลงเหลือน้อยที่สุด ในกรณีนี้ขนาดของกลุ่มตัวอย่างเป็น 20 ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม = 0.50 และตัวแปรทั้งสองอยู่ในรูปแบบมาตรฐานซึ่งทำให้ $\sigma x = \sigma y = 1$ ค่า m ที่สามารถลด MSE ให้เหลือน้อยที่สุด คือ ค่า m ตำแหน่งซึ่งทำให้ผลบวกของเส้นโค้งทั้งสองมีค่าน้อยที่สุด ซึ่งจะเห็นได้ว่า ผลบวกของเส้นโค้งทั้งสองที่น้อยที่สุดจะอยู่ในเขต $m < 1.000$ เสมอ

วิธีการหาค่าคงที่ k สำหรับการคำนวณค่าประมาณโดยวิธีของริดจ์

ในที่นี้จะกล่าวเพียง 2 วิธีเท่านั้น คือ

1. ริดจ์เทรซ (Ridge trace)

วิธีนี้นับได้ว่าเป็นวิธีการซึ่งนิยมใช้กันอย่างแพร่หลายในการหาค่าประมาณโดยวิธีริดจ์ คือ ดูจากกราฟแสดงความสัมพันธ์ระหว่างค่าประมาณของริดจ์และค่า k ($0 < k < \alpha$) ซึ่งเรียกกราฟดังกล่าวว่า ริดจ์เทรซ อย่างไรก็ดีเพื่อความสะดวกในการกำหนดแกนของริดจ์เทรซ เฮิร์ลและเคนนาร์ดกล่าวว่า ค่า k ควรมีค่าตั้งแต่ 0 ถึง 1 ก็เป็นการเพียงพอแล้ว และค่า k ที่เหมาะสมที่สุดคือค่าซึ่งมีผลต่อค่าประมาณโดยวิธีของริดจ์ดังต่อไปนี้

1) ค่าประมาณของริดจ์ทุกตัวเริ่มมีค่าคงที่

2) ค่าประมาณที่มีเครื่องหมายผิดจากที่ควรจะเป็นควรเปลี่ยนเป็นเครื่องหมายที่ถูกต้อง

3) ค่าประมาณที่มีค่าสูงผิดปกติ เมื่อ $k = 0$ ควรเปลี่ยนเป็นค่าซึ่งสมเหตุสมผลมากขึ้น

4) ค่าประมาณที่ได้ ไม่ควรทำให้ค่าผลรวมของความคลาดเคลื่อนยกกำลังสองมีค่าสูงเกินไปเมื่อเปรียบเทียบกับขนาดของความแปรปรวนที่ลดลง (กล่าวคือ ถ้าค่าริดจ์ลดค่าความแปรปรวนได้เพียงเล็กน้อย แต่ทำให้ค่าผลรวมความคลาดเคลื่อนยกกำลังสองสูงขึ้นมากก็ถือว่า ไม่คุ้มค่าในการใช้วิธีการของริดจ์)

5) ค่าประมาณที่มีค่าเปลี่ยนแปลงอย่างมากเมื่อค่า k เพิ่มเพียงเล็กน้อย หรือ มีค่าเข้าหา 0 เมื่อค่า k เพิ่มขึ้น อาจตัดออกจากโมเดลได้

ตัวอย่างของริดจ์เทรซได้แสดงไว้ในตอนท้ายของบทความนี้ด้วยแล้ว

2. วิธีการริดจ์ทั่วไป (Generalized ridge method)

วิธีนี้เสนอโดยเฮิร์ลและเคนนาร์ดเช่นกัน วิธีนี้กำหนดให้ค่า β และ X เปลี่ยนไปอยู่ในรูปคานอนิคอล กล่าวคือ α และ X^* ตามลำดับ นั่นคือ กำหนดให้

$$\alpha = T\beta$$

และ
$$X^* = XT'$$

ดังนั้น ค่าประมาณของริดจ์ในรูปคานอนิคอลคำนวณได้จาก

$$\hat{\alpha}^{r} = \left[(X^{*})'(X^{*}) + K\right]^{-1} (X^{*})'Y$$

เมื่อ K คือ orthogonal matrix ซึ่งมีค่า k_i $(i = 1, 2, 3, \ldots, p)$ ตามแนวทแยงมุมของแมตริกซ์ โดยวิธีนี้ค่า k_i แต่ละตัวจะมีค่าไม่เท่ากัน จึงต้องคำนวณใหม่ p ครั้ง

เฮิร์ลและเคนนาร์ด แสดงให้เห็นว่า โดยวิธีริดจ์ทั่วไป ค่า MSE จะสามารถทำให้น้อยที่สุดได้ เมื่อ $k_i = \sigma^2/\alpha_i^2$ ในทางปฏิบัติค่า k_i จะคำนวณจากกลุ่มตัวอย่าง ดังนั้น $k_i = \hat{\sigma}^2/\hat{\alpha}_i^2$ เมื่อคำนวณค่า k_i ได้แล้ว ค่า $\hat{\beta}^r$ ก็สามารถคำนวณได้เช่นเดียวกัน แต่ค่า k_i ดังกล่าวจะมีค่าที่ค่อนข้างต่ำกว่าปกติ เนื่องจากค่า $\hat{\alpha}_i$ จะมีแนวโน้มที่จะมีค่าสูงกว่าปกติอันเป็นผลมาจากมัลติคอลลิเนียริตี้ ดังนั้นค่า k_i จึงต้องคำนวณใหม่โดย $k_i = \hat{\sigma}^2/\hat{\alpha}_i^{r^2}$ โดย $\hat{\alpha}^r$ เป็นค่าซึ่งคำนวณจาก k_i ที่ได้จากการคำนวณครั้งแรก ค่า k_i จะทำการคำนวณใหม่ไปเรื่อย ๆ โดยวิธีดังกล่าว จนกระทั่งค่า k_i เริ่มมีความเสถียรภาพ คือแทบไม่มีการเปลี่ยนแปลงอีกต่อไปแล้ว จึงใช้ค่า k_i ที่ได้จากการคำนวณครั้งสุดท้ายในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์โดยวิธีริดจ์ อย่างไรก็ตาม วิธีนี้อาจนำไปสู่ค่า k_i ซึ่งเปลี่ยนแปลงไปได้เรื่อย ๆ โดยไม่มีเสถียรภาพได้เช่นกัน

สำหรับวิธีการคำนวณค่า k โดยวิธีอื่น ๆ ได้แก่ ค่าประมาณริดจ์โดยวิธีตรง (Directed ridge estimator) โดยกิลคีย์ และเมอร์ฟี (Guilkey and Murphy, 1975) วิธีของลอว์เลสและหวัง (Lewless and Wang, 1976) วิธีของเฮิร์ล เคนนาร์ด และบอลด์วิน (Hoerl, Kennard and Baldwin, 1975)

วิธีดังกล่าวได้รับการประเมินด้วยวิธีการซิมมูเลชั่น (Simulation study) ซึ่งได้ผลว่าตัวประมาณของริดจ์มีคุณสมบัติที่ดีกว่าตัวประมาณมาตรฐาน โดยเฉพาะอย่างยิ่งเมื่อใช้ค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง (MSE) เป็นเกณฑ์ในการตัดสิน อย่างไรก็ดีมีนักสถิติอีกมากที่คัดค้านว่า การออกแบบการทดลองของการศึกษาแบบซิมมูเลชั่นมีความลำเอียง ซึ่งทำให้ค่าประมาณโดยวิธีของริดจ์มีแนวโน้มที่จะมีคุณสมบัติเหนือกว่าวิธีมาตรฐานอยู่แล้วนั่นเอง (Draper and Nostrand, 1979 and Pagel, 1981)

ตัวอย่างการประยุกต์ใช้ของค่าประมาณโดยวิธีของริดจ์

ลุยและบี (Liu and Bee, 1984) ทดลองใช้การวิเคราะห์การถดถอยตามวิธีของริดจ์เพื่อศึกษาความสัมพันธ์ระหว่างจำนวนคดีอาชญากรรม (ตัวแปรตาม) และตัวทำนาย 3 กลุ่ม ได้แก่ สถานภาพทางเศรษฐกิจ สถิติการจับกุม และช่วงเวลาในรอบปี เช่น ช่วงต้นปี กลางปี หรือปลายปี

ข้อมูลทั้งหมดได้จากเมืองยังสทาวน์ รัฐโอไฮโอ สหรัฐอเมริกา โดยเก็บตั้งแต่ปี 1972 ถึง 1980 รายละเอียดของตัวแปรทั้งหมดมีดังต่อไปนี้

X1 = รายได้ของพลเมือง

X2 = อัตราการว่างงาน

X3 = จำนวนโรงงานที่ปิดกิจการ

X4 = ค่าใช้จ่ายของกรมตำรวจ

X5 = จำนวนการจับกุมคดี

X6 = ตัวแปรหุ่น แสดงช่วงเวลาเดือนมกราคม–มีนาคม = 1, ช่วงเวลาอื่น = 0

X7 = ตัวแปรหุ่น แสดงช่วงเวลาเมษายน–มิถุนายน = 1, ช่วงเวลาอื่น = 0

X8 = ตัวแปรหุ่น แสดงช่วงเวลากรกฎาคม–กันยายน = 1, ช่วงเวลาอื่น = 0

Y = จำนวนคดีอาชญากรรมในช่วง 3 เดือน

ค่าสหสัมพันธ์ของตัวทำนายทั้งแปดแสดงได้ดังต่อไปนี้

	1	2	3	4	5	6	7
2	−.78*						
3	.11	−.06					
4	−.14	−.09	−.33				
5	.06	−.40*	−.28	.90*			
6	−.48*	.31	.04	.02	.002		
7	.06	−.04	−.07	.04	.05	−.35*	
8	.19	−.06	−.07	−.01	.01	−.35*	−.32

*P < .05

จากค่าสหสัมพันธ์ทั้งสิ้น 28 ค่า ค่าสหสัมพันธ์ที่มีนัยสำคัญมีจำนวนทั้งสิ้น 6 ค่า ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระด้วยกันที่ค่อนข้างสูง อาจชี้ให้เห็นถึงสถานการณ์มัลติคอลลิเนียริตี้ได้ อย่างไรก็ดี ดัชนีที่มีค่าเที่ยงตรงกว่าในการศึกษามัลติคอลลิเนียริตี้ ได้แก่ ค่าไอเกน (eigenvalues) ซึ่งได้จากแมตริกซ์สหสัมพันธ์ของตัวทำนายนั่นเอง จากตัวอย่างที่กล่าวมาค่าไอเกนทั้งแปดตัว ได้แก่

$\lambda 1 = 2.298 \quad \lambda 2 = 2.037 \quad \lambda 3 = 1.317 \quad \lambda 4 = 1.125$

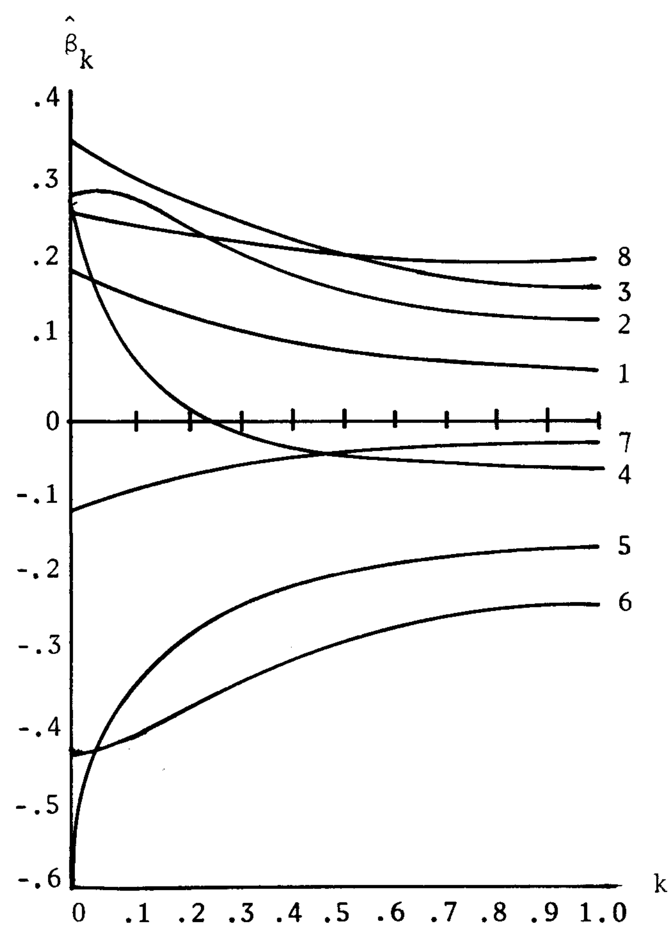
$\lambda 5 = 0.719 \quad \lambda 6 = 0.301 \quad \lambda 7 = 0.170 \quad \lambda 8 = 0.032$

ค่าไอเกนที่น้อยที่สุดจะเป็นเครื่องชี้สถานการณ์มัลติคอลลิเนียริตี้ได้ดีพอสมควร กล่าวคือค่าไอเกนที่น้อยที่สุดยิ่งมีค่าน้อยเท่าใด สถานการณ์มัลติคอลลิเนียริตี้ยิ่งรุนแรงมากขึ้นเท่านั้น ส่วนอัตราส่วนระหว่างค่าไอเกนที่มากที่สุดและน้อยที่สุด (λ max/λ min) หรือที่เรียกกันว่าตัวเลขเงื่อนไข (condition number) ก็สามารถชี้ถึงสถานการณ์มัลติคอลลิเนียริตี้ได้เช่นกัน กล่าวคือ ถ้าตัวเลขเงื่อนไขมีค่าสูงมากเท่าไรก็แสดงว่า สถานการณ์มีความรุนแรงมากขึ้นเท่านั้น สำหรับในตัวอย่างนี้ ค่าของตัวเลขเงื่อนไข = 70 ซึ่งถือว่าระดับของมัลติคอลลิเนียริตี้ อยู่ในระดับปานกลาง (ในกรณีที่ถือว่ามัลติคอลลิเนียริตี้มีค่าสูงมากนั้น ค่าของตัวเลขเงื่อนไขควรมากกว่า 100 ขึ้นไป)

ลุยและบีได้ใช้วิธีการวิเคราะห์การถดถอยโดยวิธีของริดจ์ในการหาตัวทำนายที่มีความคงเส้นคงวาในการทำนายจำนวนคดีอาชญากรรม วิธีการริดจ์เทรซจึงได้นำมาใช้ตามภาพที่ 5

ภาพที่ 5

ริดจ์เทรซจากการใช้ตัวทำนายเพียง 8 ตัว

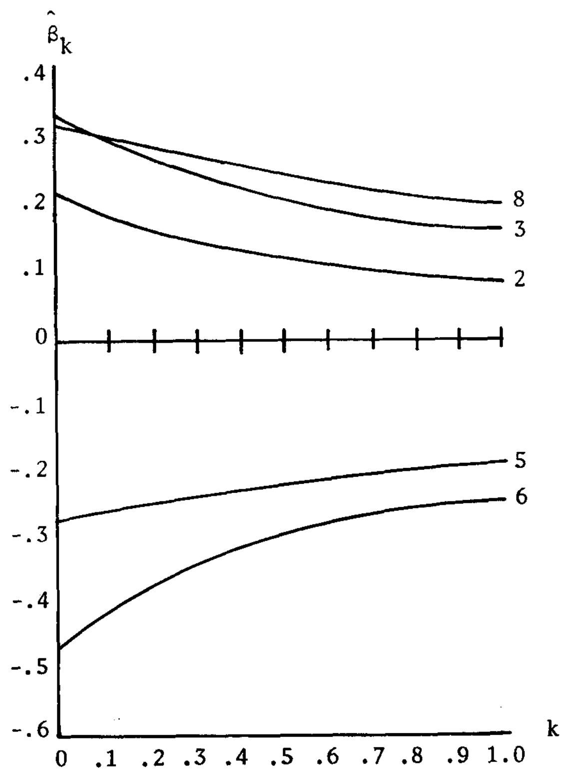


จากภาพที่ 5 จะเห็นว่า ค่าประมาณของสัมประสิทธิ์การถดถอยโดยวิธีมาตรฐานส่วนมากมีสูงกว่าที่ควรเป็นและขาดความเสถียรภาพ ซึ่งสังเกตได้จาก เมื่อค่า k เพิ่มขึ้นเพียงเล็กน้อยค่าประมาณเหล่านี้มีค่าลดลงอย่างมาก โดยเฉพาะอย่างยิ่ง X5 (อัตราการจับกุม) นอกจากนี้ค่าประมาณของ X4 (ค่าใช้จ่ายของตำรวจ) เปลี่ยนเครื่องหมายจากบวกไปเป็นลบเมื่อค่า k เพิ่มขึ้น นั่นคือ โดยวิธีมาตรฐานชี้ว่ายิ่งเพิ่มค่าใช้จ่ายให้ฝ่ายตำรวจมากขึ้น คดีอาชญากรรมกลับจะยิ่งสูงขึ้น ในทำนองกลับกัน ค่าประมาณโดยวิธีริดจ์ชี้ว่า ยิ่งเพิ่มค่าใช้จ่ายของตำรวจมากขึ้นจำนวนคดีกลับลดลงนั่นเอง

จากริดจ์เทรซจะเห็นได้ชัดว่า ตัวทำนายหมายเลข 1, 4 และ 7 มีอำนาจในการทำนายจำนวนคดีอาชญากรรมค่อนข้างต่ำ รวมทั้งขาดความเสถียรภาพ จากการยึดหลักการของเฮิร์ลและเคนนาร์ด ในการใช้ริดจ์เทรซเป็นเครื่องมือลดจำนวนตัวทำนาย ลุยและบีได้ตัดตัวแปรทั้ง 3 ดังกล่าวทิ้ง แล้วสร้างริดจ์เทรซขึ้นใหม่ โดยมีตัวทำนายเพียง 5 ตัว คือ X_2, X_3, X_5, X_6 และ X_8 ริดจ์เทรซที่มีตัวทำนายเพียง 5 ตัวดังกล่าว แสดงไว้ในภาพที่ 6

ภาพที่ 6

ริดจ์เทรซเมื่อใช้ตัวทำนาย 5 ตัว



จากภาพที่ 6 จะเห็นว่าระบบของตัวประมาณทั้งหมดเริ่มมีเสถียรภาพที่ k = .75 ค่าประมาณของสัมประสิทธิ์สหสัมพันธ์โดยวิธีมาตรฐานและวิธีของริดจ์ เมื่อตัวทำนายทั้งหมดคงอยู่และเมื่อตัวทำนายถูกตัดออก 3 ตัว ได้แสดงไว้ในตารางต่อไปนี้

ตารางที่ 1

ตารางแสดงค่าประมาณโดยวิธีมาตรฐานและวิธีของริดจ์

ตัวทำนายที่	เมื่อใช้ตัวทำนายทุกตัว k = 0	เมื่อใช้ตัวทำนายทุกตัว k = .75	ใช้ตัวทำนาย 5 ตัว k = 0	ใช้ตัวทำนาย 5 ตัว k = .75
1	.191	.067		
2	.271	.131	.217	.97
3	.344	.183	.342	.194
4	.331	− .064		
5	− .570	− .178	− .288	− .215
6	− .441	− .254	− .451	− .256
7	− .109	− .045		
8	.207	.230	.330	.217
R^2	.698	.573	.662	.564

จากตารางที่ 1 จะเห็นได้ว่าเมื่อใช้ตัวทำนายทั้งเจ็ดในการทำนายสถิติคดีอาชญากรรม จะพบว่า X1 (อัตราค่าแรง) X1 (อัตราการว่างงาน) X3 (จำนวนโรงงานที่ปิด) และ X8 (ตัวแปรหุ่น แสดงถึงควอเตอร์ที่ 3 ของปี) มีความสัมพันธ์ในทางบวกกับอัตราการก่อคดีอาชญา-กรรม ส่วน X4 (ค่าใช้จ่ายของกรมตำรวจ) X5 (อัตราการจับกุม) X6 (ตัวแปรหุ่นแสดงถึงควอเตอร์แรกของปี) และ X7 (ตัวแปรหุ่นแสดงถึงควอเตอร์ที่สองของปี) ต่างมีความสัมพันธ์ในทางลบกับอัตราการก่อคดีอาชญากรรม เมื่อตัวทำนายถูกลดเหลือเพียง 5 ตัวความสัมพันธ์ดังกล่าวไม่เปลี่ยนแปลง เมื่อพิจารณาในด้านทิศทางของความสัมพันธ์ (เครื่องหมายของตัวประมาณ) ยิ่งกว่านั้นค่า R^2 โดยวิธีของริดจ์ลดจาก .573 เมื่อใช้ตัวทำนายทั้งแปดเหลือ .564 เมื่อใช้ตัวทำนายเพียง 5 ตัว อาจเห็นได้ว่า เมื่อวิธีการของริดจ์สามารถลดตัวทำนายลงได้โดยค่า R^2 มีค่าลดลงบ้างเพียงเล็กน้อยเท่านั้น

เมื่อยึดถือรูปแบบที่ใช้ตัวทำนายเพียง 5 ตัว อาจสรุปผลการวิจัยได้ว่าเมื่อควบคุมตัวแปรอื่นให้คงที่ อัตราการจับกุมมีผลต่อคดีอาชญากรรมในทางลบ กล่าวคือ ถ้ามีการจับกุมมากขึ้น

คดีอาชญากรรมจะลดลง แต่ถ้ามีการจับกุมน้อยลง คดีอาชญากรรมกลับจะสูงขึ้น ในทำนองเดียวกัน ตัวแปรที่ชี้ให้เห็นสภาพทางเศรษฐกิจ เช่น อัตราการว่างงาน หรือจำนวนโรงงานที่ปิดกิจการ เป็นเครื่องชี้ถึงอัตราคดีอาชญากรรมได้อย่างดี กล่าวคือ ถ้าอัตราการว่างงานหรือจำนวนโรงงานที่ปิดกิจการมีมากขึ้น ก็จะทำให้จำนวนคดีอาชญากรรมมีมากขึ้นด้วย นอกจากนี้ช่วงเวลาในรอบปีก็มีส่วนช่วยในการชี้ให้เห็นจำนวนคดีอาชญากรรมเช่นกัน กล่าวคือ จำนวนคดีอาชญา-กรรมมีแนวโน้มที่จะมีมากในควอเตอร์ที่สาม (ฤดูร้อน) ของปี แต่มีแนวโน้มที่จะลดลงในควอเตอร์ที่หนึ่ง (ฤดูหนาว) ของปีนั่นเอง

จากตัวอย่างดังกล่าวจะเห็นได้ว่า ผู้วิจัยได้ใช้วิธีการของริดจ์ในการทำให้ตัวประมาณแต่ตัวมีความเสถียรภาพสูงขึ้น กล่าวคือ เมื่อใช้ริดจ์เทรซเป็นหลักในการคัดตัวทำนายบางตัวออก ตัวประมาณที่เหลือจะมีความเสถียรภาพสูงขึ้นอย่างมาก ดังจะเห็นได้ว่า ตัวทำนายทั้งห้ามีค่าเกือบคงที่ไม่ว่าจะใช้วิธีของริดจ์หรือวิธีมาตรฐานก็ดี ผู้อ่านอาจจะเห็นว่า ค่า R^2 อาจมีค่าลดลงบ้างเมื่อใช้วิธีของริดจ์ อย่างไรก็ดี วิธีของริดจ์มักใช้เพื่อต้องการหาเสถียรภาพของตัวทำนายเป็นจุดมุ่งหมายหลัก โดยยอมให้ค่า R^2 ลดลงบ้างเป็นการแลกเปลี่ยนนั่นเอง

บรรณานุกรม

Churngchow, C. (1987). *Ridge regression : Application to educational data.* Unpublished doctoral dissertation. The Florida State University.

Darlington, R.B. (1978). Reduced-variance regression. *Psychological Bulletin, 85,* 1238-1255.

Draper, N.R. and Nostrand, R.C. (1979). Ridge regression and James Stein estimators : Review and comments. *Technometrics, 21,* 451-466.

Faden, V.B. (1978). *Shrinkage in ridge regression and ordinary least squares multiple regression estimators.* Unpublished doctoral dissertation University of Maryland.

Guilkey, D.K. and Murphy, J.L. (1975). Directed ridge regression technigues in case of multicollinearity. *Journal of the American Statistical Association, 70,* 767-775.

Gorman, J.W. and Toman, R. (1966). Selection of variables for fitting equation to data. *Technometrics, 8,* 27-51.

Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression : Biased estimation for non-orthogonal problems. *Technometrics, 12,* 55-65.

Hoerl, A.E. and Kennard, R.W. (1970b). Ridge Regression : Applications to non-orthogonal problems, *Technometrice, 12,* 69-82.

Hoerl, A.E and Kennard, R.W. (1976). Kidge Regression : Iterative estimation of the biasing parameter. *Communications in Statistics, 5,* 77-88.

Hoerl, A.E. Kennard, R.W. and Baldwin, K.F. (1975). Rkdge regression : Some Simulations. *Communications in Statistics, 4,* 105-123.

Lawless. J.F. and Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in statistics, A5 (4),* 307-323.

Liu, Y, and Bee, R.H. (1984). Ridge regression : An application in criminology. *Communication in statistics. Theory and methods. 13,* 263-271.

Marquardt, D.W. and Snee, R.D. (1975). Ridge regression in practice. *The American Statistician. 29,* 3-20.

Pagel, M.D. (1981). Comment on Hoerl and Kennard's ridge regression simulation methodology. *Communication in statistics. Theory and methods. 13,* 263-271.