บทที่ 8

Multicolinearity

8.1 สรุปความรู้เกี่ยวกับ Characteristic Value

เนื่องจากการศึกษาเรื่อง Characteristic Value และ Characteristic Vector ในวิชา Linear Algebra เป็นเรื่องที่นำไปสู่ความรู้เรื่องความเป็นอิสระต่อกันของเวคเตอร์ (Independent หรือ Orthogonal) ขณะเดียวกันกับที่ปัญหา Multicolinearity ก็คือ ปัญหาที่เกิดจากเวคเตอร์ต่าง ๆ ในแมตริกซ์ X ของสมการ $Y = X\beta+U$ มีความข้องเกี่ยวกันอยู่ในบางระดับ การวินิจฉัยปัญหา Multicolinearity จึงต้องอาศัยความรู้เกี่ยวกับ Characteristic Value และ Characteristic Vector เป็นพื้นฐานอยู่ตลอดเวลา และเพื่อให้เกิดความเข้าใจโดยตลอดโดยที่ความรู้พื้นฐานทางพีชคณิตเชิงเส้นจักไม่เป็นอุปสรรคต่อความรู้ความเข้าใจในการศึกษาเรื่องราวของ Multicolinearity ผู้เขียนขอถือโอกาสนี้สรุปความรู้เกี่ยวกับ Characteristic Value ไว้ในที่นี้เสียก่อน โดยจะกล่าวถึงเฉพาะ Characteristic Value ของ Symmetric Matrix เท่านั้น ทั้งนี้เพราะแมตริกซ์ที่เราจะเกี่ยวข้องต่อไปในปัญหา Multicolinearity คือ แมตริกซ์ $X'X$ ซึ่งเป็น Symmetric Matrix

ให้ $Y'AY$ เป็น Quadratic Form โดยเราต้องการ Maximize $Y'AY$ ภายใต้ข้อจำกัดว่า Y เป็น Orthogonal Vector กล่าวคือ $Y'Y = 1$ เมื่อ A เป็น Matrix of Quadratic Form ขนาด $n \times n$

ดังนั้นโดยอาศัย Lagrangian Multplier จะพบว่า

$$F = Y'AY - \lambda(Y'Y-1)$$

$$\frac{\partial F}{\partial Y} = 0 = 2AY - 2\lambda Y = 0$$

$$\text{หรือ} \quad (A - \lambda I_n)Y = 0 \quad \text{หรือ} \quad AY = \lambda Y$$

นั่นคือเวคเตอร์ Y ที่จะมีผลให้ Quadratic Form $Y'AY$ มีค่าสูงสุดภายใต้ข้อจำกัดว่า เวคเตอร์ Y ต้องเป็น Orthogonal Vector คือ เวคเตอร์ Y ที่สอดคล้องกับระบบสมการ $AY = \lambda Y$

และเนื่องจากสมการ $(A-\lambda I_n)Y = 0$ เป็น Linear Homogeneous System ซึ่งจะมี Non-Trivial (Nonnullหรือ Nonzero) ได้ก็เฉพาะเมื่อ $(A-\lambda I_n)$ ไม่มี Full Rank (คือ $r(A-\lambda I_n) < n$) ดังนั้นจึงมีผลให้ $|A-\lambda I_n| = 0$

สมการ $|A-\lambda I_n| = 0$ เรียกว่า Characteristic Polynomial, $f(\lambda)$ รากของสมการ $f(\lambda)$ คือ λ_i, $i = 1,2,...,n$ เรียกว่า Characteristic Root เวคเตอร์ที่สอดคล้องกับรากแต่ละค่าของ $f(\lambda)$ เรียกว่า Characteristic Vector, Y_i; $i = 1,2,...,n$

ตัวอย่างเช่นเมื่อกำหนดให้

$$B = \begin{bmatrix} 2 & 2 & -7 \\ 2 & 1 & -2 \\ 0 & 1 & -3 \end{bmatrix}$$

เราสามารถคำนวณหา Characteristic Value ของ B และ Characteristic Vector ที่สอดคล้องกันได้ดังนี้ (B ไม่เป็น Symmetric Matrix)

จาก $f(\lambda) = |A-\lambda I_3| = 0$ แสดงว่า

$$f(\lambda) = \begin{vmatrix} 2-\lambda & 2 & -7 \\ 2 & 1-\lambda & -2 \\ 0 & 1 & -3-\lambda \end{vmatrix}$$

$$\begin{aligned} f(\lambda) &= -(2-\lambda)(1-\lambda)(3+\lambda)-14+4(3+\lambda)+2(2-\lambda) \\ &= -\lambda^3+13\lambda-12 \end{aligned}$$

เมื่อกำหนดให้ $f(\lambda) = 0$ คือให้

$$-\lambda^3+13\lambda-12 = 0$$

จะได้ $(\lambda-1)(\lambda-3)(\lambda+4) = 0$ หรือ $\lambda = 1,3,-4$

เมื่อ $\lambda_1 = 1$ จะได้ Characteristic Vector จากสมการ $(B-\lambda_1 I_3)\ Y = 0$ ดังนี้

$$\begin{bmatrix} 2-1 & 2 & -7 \\ 2 & 1-1 & -2 \\ 0 & 1 & -3-1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

หรือ

$$y_1+2y_2-7y_3 = 0 \qquad \ldots.(1)$$

$$2y_1 \quad -2y_3 = 0 \qquad \ldots.(2)$$

$$y_2-4y_3 = 0 \qquad \ldots.(3)$$

ข้อสังเกต ขอให้สังเกตว่าสมการทั้ง 3 นี้ไม่เป็นอิสระต่อกัน การหาคำตอบจึงต้องใช้ Parametric Equation

จาก (2) พบว่า $y_1 = y_3$

จาก (3) พบว่า $y_2 = 4y_3$

ถ้าให้ $y_3 = k_1$ จะได้ Characteristic Vector ที่สอดคล้องกับ $\lambda_1 = 1$ จำนวนมากมายหลายชุดดังนี้

$$Y_1 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = k_1 \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$$

$$Y_1 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$$ คือ Characteristic Vector ชุดหนึ่งที่สอดคล้องกับ $\lambda_1 = 1$

สำหรับ $\lambda_2 = 3$, $\lambda_3 = -4$ ก็สามารถหาเวคเตอร์ที่สอดคล้องได้ในทำนองเดียวกัน คือ

$$Y_2 = \begin{bmatrix} 5 \\ 6 \\ 1 \end{bmatrix} \quad , \; Y_3 = \begin{bmatrix} -3 \\ 2 \\ -2 \end{bmatrix}$$

และเราสามารถพิสูจน์ได้โดยง่ายว่า Y_1, Y_2 และ Y_3 เป็นอิสระเชิงเส้นต่อกันเสมอ ดังทฤษฎีต่อไปนี้

ทฤษฎี 1 เมื่อ $\lambda_i \; ; \; i = 1,2,\ldots,n$ เป็น Characteristic Value ที่มีค่าแตกต่างกันของแมตริกซ์ B Characteristic Vector $Y_i \; ; \; i = 1,2,\ldots,n$ ที่สอดคล้องกับ λ_i จะเป็นอิสระเชิงเส้นต่อกันเสมอ และเมื่อ B เป็น Symmetric Matrix เวคเตอร์ Y_i คู่ใด ๆ จะตั้งฉาก (Orthogonal) กันเสมอ

นอกจากนี้ยังมีทฤษฎีเกี่ยวกับ Similar Matrix และ Diagonal Matrix ซึ่งล้วนสืบเนื่องจากปัญหา Characteristic Value ทั้งสิ้น ดังนี้

ทฤษฎี 2 Similar Matrix จะมีดีเทอร์มิแนนท์เท่ากันทั้งให้ค่า Characteristic Value เดียวกัน และมีผลให้มี Characteristic Vector เดียวกัน

หมายเหตุ แมตริกซ์ B จะ Similar กับแมตริกซ์ A ได้ถ้าหากว่าเราสามารถเสาะหาแมตริกซ์ C ที่เป็น Nonsingular มาได้ และมีผลให้ $B = C^{-1}AC$

ทฤษฎี 3 ถ้า λ คือ Characteristic Value ของ A แล้ว λ^k จะเป็น Characteristic Value ของ A^k (โดยที่ k เป็นจำนวนเต็มใด ๆ) ในขณะเดียวกันถ้า Y คือ Characteristic Vector ที่สอดคล้องกับ λ เวคเตอร์ Y ก็จะเป็น Characteristic Vector ที่สอดคล้องกับ λ^k ด้วย

ทฤษฎี 4 ถ้าแมตริกซ์ A Similar กับ Diagonal Matrix D แล้ว D จะสามารถจัดรูปได้ดังนี้

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

เมื่อ $\lambda_i \; ; \; i = 1,2,\ldots,n$ คือ Characteristic Value ของ A

ทฤษฎี 6 ถ้า λ_i คือ Characteristic Vector **ที่สอดคล้องกับ** Characteristic Value **ของ** B **เมื่อ** $B = C^{-1}AC$ **แล้ว** $X_i = CY_i$ **จะเป็น** Characteristic Vector **ของ** A **ที่สอดคล้องกับ** Characteristic Value λ_i **ของ** B **ดังกล่าว**

ทฤษฎี 6 A **จะ** Similar **กับ** Diagonal Matrix D **ก็ต่อเมื่อ** Characteristic Vector **ของ** A **เป็นอิสระเชิงเส้นต่อกัน**

ทฤษฎี 7 ถ้า Characteristic Value **ของ** A **มีค่าแตกต่างกันแล้ว** Characteristic Vector **ของ** A **จะต่างกันและ** A **จะ** Similar **กับ** B

ทฤษฎี 8 Real Symmetric Matrix ทุกรูปจะ Orthogonally Similar กับ D

ประโยชน์จากทฤษฎีเหล่านี้สามารถสรุปได้ดังนี้

สมมุติว่า Y คือแมตริกซ์ที่สดมภ์ของ Y คือ Characteristic Vector $Y_1, Y_2, \ldots, Y_n$ กล่าวคือ $Y = [Y_1, Y_2, \ldots, Y_n]$ และถ้าหากว่า $\lambda_1, \lambda_2, \ldots, \lambda_n$ มีค่าต่างกันทั้งหมด สดมภ์ของ Y จะตั้งได้ฉาก (Orthogonal) กัน คือ $Y_i'Y_i = 1$ มีผลให้ Y เป็น Orthogonal matrix ($Y' = Y^{-1}$ หรือ $Y'Y = I$) และเราสามารถเสนอระบบสมการ $AY_i = \lambda_i Y_i$; $i = 1,2,\ldots,n$ ร่วมกันในรูปแมตริกซ์ได้ดังนี้

$$A[Y_1, Y_2, \ldots, Y_n] = [\lambda_1 Y_1, \lambda_2 Y_2, \ldots, \lambda_n Y_n] \quad \text{หรือ} \quad AY = YD$$

$$D = \begin{vmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{vmatrix} = diag[\lambda_1, \lambda_2, \ldots, \lambda_n]$$

พิจารณาสมการ AY = YD (ขอให้สังเกตว่าแมตริกซ์ Y ในที่นี้คือแมตริกซ์ C ในทฤษฎี 2) จะพบว่า Y เป็น Orthogonal Matrix เมื่อนำ Y^{-1} คูณด้านหน้าสมการ AY = YD จะได้

$$Y^{-1}AY = Y^{-1}YD = D$$

หรือ

$$Y^{-1}AY = D$$

แสดงว่าเมื่อ A มี Characteristic Value ที่แตกต่างกัน เราย่อมสามารถหา Orthogonal Matrix Y ซึ่งประกอบไปด้วย Characteristic Vector $Y_1, Y_2, \ldots, Y_n$ ที่มีผลให้ $Y'AY$ Similar กับ D เมื่อ $D = \text{diag}\ (\lambda_1, \lambda_2, \ldots, \lambda_n)$

หมายเหตุ ความรู้ในส่วนนี้ผู้เขียนได้เคยอ้างอิงไปครั้งหนึ่งแล้วในตอนที่ว่าด้วย GLS เกี่ยวกับการหาแมตริกซ์ P ที่มีผลให้ $V(PU) = \sigma^2 I_n$ ความรู้เหล่านี้จะถูกนำไปใช้อีกครั้งหนึ่งในการวิเคราะห์ปัญหา Multicolinearily และ Ridge Regression

นอกเหนือจากที่กล่าวมาแล้วหากพิจารณาถึง Rank และ Trace เราจะพบรายละเอียดเพิ่มเติมดังนี้

จาก $Y'AY = D$ จะพบว่า

$$|Y'AY| = |D| = \lambda_1 \lambda_2 \ldots \lambda_n$$

เมื่อพิจารณา Orthogonal Matrix Y เราทราบว่า $|Y| = |Y'| = \pm 1$ และทราบว่า $|Y'AY| = |Y'|\,|A|\,|Y|$ ดังนั้น $|A| = |D| = \lambda_1 \lambda_2 \ldots \lambda_n$ หรือเราสามารถกล่าวได้ว่า ดีเทอร์มิแนนต์ของแมตริกซ์ใด ๆ มีค่าเท่ากับผลคูณของ Characteristic Vector

และจาก $Y'AY = D$ จะพบว่า $tr(Y'AY) = tr\,D$ แต่ $tr(Y'AY) = tr(AYY') = tr\,A$ ขณะที่ $tr\ D = \lambda_1 + \lambda_2 + \ldots + \lambda_n$ นั่นคือ $tr\,A = X_1 + X_2 + \ldots + \lambda_n$ หรือกล่าวได้ว่า trace ของแมตริกซ์ใด ๆ มีค่าเท่ากับผลรวมของ Characteristic Value

นอกจากนี้เรายังสามารถพิสูจน์ได้ว่า $r(A) = r(D)$ ขณะเดียวกัน จากสมการ $AY = YD$ ถ้านำ Y^{-1} คูณด้านหลังจะได้ $AYY^{-1} = YDY^{-1}$ หรือ $A = YDY'$ หรือ

$$A = \lambda_1 Y_1 Y_1' + \lambda_2 Y_2 Y_2' + \ldots + \lambda_n Y_n Y_n'$$

และ $A^2 = AA = (YDY')(YDY') = YD(Y'Y)DY' = YDI_nDY' = YD^2Y'$

$$= \lambda_1^2 Y_1Y_1' + \lambda_2^2 Y_2Y_2' + \ldots + \lambda_n^2 Y_nY_n'$$

$$A^r = AA\ldots A = YD^rY'$$

$$= \lambda_1^r Y_1Y_1' + \lambda_2^r Y_2Y_2' + \ldots + \lambda_n^r Y_nY_n'$$

ซึ่งแสดงว่า Characteristic Value ของ A^r ก็คือกำลังที่ r ของ Characteristic Value ของ A นั่นเอง

8.2 ปัญหา Multicolinearity และผลกระทบทางสถิติ

ปัญหา Multicolinearity[1] หมายถึงสถานการณ์ที่ตัวแปรต่าง ๆ มีความสัมพันธ์กันอย่างน้อย 2 ครั้งคือ

ครั้งแรก ตัวแปรสัมพันธ์กันในรูป Linear Model $Y = X\beta + U$

ครั้งที่สอง ตัวแปรอิสระ X's มีความสัมพันธ์เชิงเส้นต่อกัน (Linearly Dependent) คือ $c_1X_1+c_2X_2+c_3X_3+\ldots+c_kX_k = 0$ เมื่อ c_i เป็นค่าคงที่ที่ไม่ใช่ 0 ทั้งหมด

จากความสัมพันธ์ทั้งสองระดับนี้ เมื่อพิจารณาให้ดีแล้วพบว่าความสัมพันธ์ครั้งที่สองคือความสัมพันธ์ที่นำไปสู่ปัญหา Multicolinearity หรือถ้าจะกล่าวให้เฉพาะเจาะจงเราอาจกล่าวได้ว่า ปัญหา Multicolinearity คือ ปัญหาที่เกิดขึ้นจากการที่ตัวแปรอิสระมีความสัมพันธ์ภายในต่อกันในอัตราค่อนข้างสูง ซึ่งความสัมพันธ์ลักษณะดังกล่าวจะมีผลให้แมตริกซ์ $X'X$ เป็น Singular Matrix หรือมีลักษณะของ Near Singularity[2]

---

[1] Multicolinearity = Multi + Colinearity คำว่า Colinear ในทางพีชคณิตเชิงเส้นหมายถึง สถานการณ์ที่เวคเตอร์หนึ่งต้องพึ่งพิงเวคเตอร์อื่น หรือเวคเตอร์หนึ่งทาบหรือทับไปบนเวคเตอร์อื่น ซึ่งแสดงให้เห็นว่าได้เกิด Linearly Dependent ระหว่างเวคเตอร์ขึ้น คำว่า Multicolinearity ถูกนำมาใช้ในงานวิเคราะห์สมการถดถอยเป็นครั้งแรกในปี ค.ศ. 1934 โดย Ragnar Frisch

[2] ถ้า Column Vector ที่ประกอบกันขึ้นเป็นแมตริกซ์ X มีลักษณะของ Linearly Dependent แมตริกซ์ X จะไม่มี Full Rank ย่อมหมายความว่า $X'X$ เป็น Singular Matrix คือ $|X'X| = 0$ อันจะมีผลให้ $(X'X)^{-1}$ ไม่ปรากฏค่า แต่ถ้า Column Vector ไม่ถึงกับมี Linearly Dependent แต่มีลักษณะที่ผันแปรร่วมทิศทางกัน $X'X$ จะเป็นเพียง Near Singularity ที่ $|X'X|$ มีค่าใกล้ 0 ซึ่งเรายังคงหา $(X'X)^{-1}$ ได้ แต่สมาชิกของ $(X'X)^{-1}$ จะมีค่าสูง

ในทางปฏิบัติเรามักจะพบว่าตัวแปรอิสระมีความสัมพันธ์กันในลักษณะของการผันแปรร่วมไปในทิศทางเดียวกันเท่านั้น คือไม่มีความสัมพันธ์สูงถึงระดับที่ตัวแปรอิสระบางส่วนมีค่าเท่ากันหรือเป็นตัวแปรเดียวกัน ปัญหาที่ปรากฏพบจึงปรากฏในลักษณะที่ $X'X$ มีธรรมชาติของ Near Singularity เท่านั้น ไม่ถึงกับเป็น Singular Matrix แต่ถึงแม้ว่า $X'X$ มีธรรมชาติของ Near Singularity ที่มีผลให้เรายังคงประมาณค่า β ได้จากสูตรเดิมคือ $\hat{\beta} = (X'X)^{-1}X'Y$ แต่สิ่งที่เป็นปัญหาก็ยังคงปรากฏอยู่ซึ่งจะได้สรุปเป็นข้อ ๆ ดังนี้

1. เมื่อเกิดปัญหา Multicolinearity ขึ้น นักวิจัยจะประสบความยุ่งยากในการจำแนกอิทธิพลของตัวแปรอิสระ ออกมาให้เห็นอย่างชัดแจ้งว่า ตัวแปรอิสระหนึ่ง ๆ มีอิทธิพลต่อ Y มากน้อยเพียงใด ที่เป็นเช่นนี้เพราะตัวแปรอิสระบางส่วนผูกพันกันอยู่ หรือมีทิศทางของการผันแปรร่วมกัน ตัวอย่างเช่น งานวิเคราะห์ทางเศรษฐกิจเราพบว่าในจังหวะที่สถานการณ์ทางเศรษฐกิจเฟื่องฟู (Boom) ตัวแปรรายได้ การบริโภค การออม การลงทุน ราคาสินค้า ตลอดถึงการว่าจ้าง (หรือการมีงานทำ) จะขยายตัว หรือเพิ่มค่าไปในทางบวกร่วมกัน ขณะเดียวกันที่เมื่อจังหวะที่สถานการณ์ทางเศรษฐกิจฟุบตัวลง ปัจจัยเหล่านี้ก็ลดตัวลง หรือลดค่าลง คือเปลี่ยนแปลงไปในทางลบร่วมกัน (Negative Covariation) หากนักวิจัยนำตัวแปรเหล่านี้เข้าเป็นตัวแปรสำหรับวิเคราะห์สมการถดถอย $Y = f(X's)$ ปัญหาที่เกิดขึ้นก็คือ $X'X$ มีลักษณะของ Near Singularity และนักวิจัยเองก็ไม่อาจแยกได้ว่าปัจจัยใดเป็นสาเหตุให้ปัจจัย Y ผันแปรไป จะขจัดปัจจัยใดทิ้งก็ไม่อาจแน่ใจว่าจะผิดพลาดหรือไม่

สมมุติว่าแบบจำลองที่กำหนดขึ้นคือ

$$Y = \beta_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+u$$

และเราทราบว่า X_3, X_4 และ X_5 มีลักษณะของการผันแปรร่วมไปในทิศทางเดียวกัน (มี Covariation หรือ move together) ซึ่งเป็นสถานการณ์ที่เราไม่อาจแยกประมาณค่า β_3, β_4 และ β_5 ได้ ทางออกที่พอเป็นไปได้คือถือว่า $X_3 = X_4 = X_5$ แบบจำลองข้างต้นจึงเปลี่ยนรูปเป็น

$$Y = \beta_1+\beta_2X_2+(\beta_3+\beta_4+\beta_5)X_3+u$$

ต้องประมาณในรูปผลรวมคือ $(\beta_3+\beta_4+\beta_5)$ ไม่อาจแยกประมาณได้ ปัญหาก็คือ เราหา BLUE ของ $(\beta_3+\beta_4+\beta_5)$ ได้หรือไม่ หรือนัยหนึ่งค่าประมาณของ $(\beta_3+\beta_4+\beta_5)$ เป็น BLUE ของ $(\beta_3+\beta_4+\beta_5)$ หรือไม่

2. เมื่อเกิดปัญหา Multicolinearity ค่าประมาณของ β's จะไม่แม่นยำและไม่มั่นคง หมายความว่าเมื่อเพิ่มหรือลดขนาดตัวอย่างเพียงเล็กน้อย ค่าประมาณของ β_j's จะเปลี่ยนแปลงไปอย่างใหญ่หลวง แต่ $\hat{V}(\hat{\beta}_j)$ อาจไม่ได้รับความกระทบกระเทือนจากปัญหา Multicolinearity กล่าวคือจากการศึกษาในวาระต่าง ๆ กันพบว่าบางครั้งเมื่อเพิ่ม Colinear Variable เข้าสู่สมการจะพบว่า $\hat{V}(\hat{\beta}_j)$ เพิ่มสูงขึ้นอย่างรวดเร็ว แต่ในบางครั้งกลับพบว่า $\hat{V}(\hat{\beta}_j)$ มิได้เปลี่ยนแปลงไปแต่ประการใด อย่างไรก็ตามแม้อิทธิพลของ Multicolinearity ที่มีต่อ $\hat{V}(\hat{\beta}_j)$ จะยังไม่แจ้งชัดเพราะมีประเด็นพอให้ท้วงติงได้ แต่โดยทั่วไปแล้ว $\hat{V}(\hat{\beta}_j)$ จะสูงขึ้นเมื่อมีปัญหา Multicolinearity เหตุที่ $\hat{V}(\hat{\beta}_j)$ มีค่าสูงย่อมมีผลให้ t_j มีค่าต่ำเพราะ $t_j = \hat{\beta}_j / \sqrt{\hat{V}(\hat{\beta}_j)}$ ผลที่ติดตามมาคือ β_j ไม่มีนัยสำคัญ และทำให้นักวิจัยตัดตัวแปรอิสระ X_j ทิ้งไปทั้งที่ X_j อาจเป็นตัวแปรที่ทรงความสำคัญต่องานเป็นอย่างยิ่งก็ได้

3. เมื่อเกิดปัญหา Multicolinearity ค่าประมาณของ β_j บางตัวจะมีเครื่องหมายผิดไปจากความเป็นจริง แต่ถ้าเครื่องหมายถูกต้องขนาดของ $\hat{\beta}_j$ ก็ผิดแผกไปจากที่ควรจะเป็น หรือมีค่าที่ไม่ก่อให้เกิดประโยชน์ในเชิงอภิปรายผล

4. เมื่อเกิดปัญหา Multicolinearity จะพบว่า $R^2_{1.23..k} < R^2_{j.23...,j-1,j+1,..k}$; $j = 2,3,..,k$

5. เมื่อเกิดปัญหา Multicolinearity ฟังก์ชันของ $\beta_1,\beta_2,..,\beta_k$ คือ

$$W'\beta = W_1\beta_1+W_2\beta_2+\ldots+W_k\beta_k$$

จะไม่เป็น Linear Unbiased Estimator

สำหรับผลกระทบที่ 5 นี้เป็นผลกระทบที่นับว่ารุนแรงที่สุดของปัญหา Multicolinearity เพราะถึงแม้จะเกิดปัญหา Multicolinearity ขึ้น $\hat{\beta}$ ก็ยังคงเป็น Unbiased Estimator ของ β และเราพอจะแก้ปัญหาการประมาณค่า $\hat{\beta}_j$ ได้ด้วยการประมาณผลรวมของ β_j (ดูข้อ 1) แต่สิ่งที่นับว่าเป็นปัญหาก็คือค่าประมาณของผลรวมดังกล่าวไม่เป็น Linear Unbiased Estimator และในบางครั้งผลรวมหรือฟังก์ชั่นดังกล่าวก็เป็นสิ่งที่ไม่อาจประมาณค่าได้

หมายเหตุ จากข้อ 1 ผลรวมของพารามิเตอร์คือ $\beta_3+\beta_4+\beta_5$ สามารถเสนอในรูปได้ดังนี้คือ

$$\beta_3+\beta_4+\beta_5 = 0\beta_1+0\beta_2+\beta_3+\beta_4+\beta_5 = [0,0,1,1,1]\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_5\end{bmatrix} = W'\beta$$

สำหรับผลกระทบที่ 5 นี้มีทฤษฎีสำคัญที่เป็นการแสดงให้เห็นว่าเมื่อเกิดปัญหา Multicolinearity ขึ้นเราสามารถหา Linear Unbiased Estimator ของ $W'\beta$ได้ โดยอาศัย Characteristic Vector ดังนี้

ทฤษฎี 8.1 สมการ $Y = X\beta+U$ ที่ X ไม่มี Full Column Rank นั้นเราสามารถหา - Unbiased Estimator ของ $W'\beta$ได้ถ้าหากว่าสามารถจัดให้ W อยู่ในรูปการประกอบกันเชิงเส้นของแถวต่าง ๆ ของ X ซึ่งถ้าหากกระทำได้ BLUE ของ $W'\beta$ คือ $W'\hat{\beta}$เมื่อ $\hat{\beta}$ คือ Solution จากระบบสมการ $X'X\hat{\beta} = X'Y$ โดย $W'\hat{\beta}$จะเป็น Unique Estimator ของ $W'\beta$เสมอ

ทฤษฎี 8.2 เราสามารถประมาณค่าฟังก์ชัน $W'\beta$ได้ก็ต่อเมื่อ W เป็น Linear Combination ของ Characteristic Vector ที่สอดคล้องกับ Characteristic Value λ_i ที่ไม่ใช่ 0 ของ $X'X$

พิสูจน์ (ทฤษฎี 8.2)

จากสมการ $Y = X\beta+U$

เราสามารถหาแมตริกซ์ P ที่แต่ละสดมภ์ของ P คือ Normalized Characteristic Vector ของแมตริกซ์ $X'X$ ที่มีผลให้ P เป็น Orthogonal Matrix ได้ ($P' = P^{-1}$ หรือ $P'P = PP' = I_k$)

ดังนั้นสมการ $Y = X\beta+U$ จึงจัดรูปเสียใหม่ได้ดังนี้คือ

$$
\begin{aligned}
Y &= X(PP')\beta+U \\
&= (XP)(P'\beta)+U \\
&= Z\theta+U \quad \text{เมื่อ} \quad Z = XP \text{ และ } \theta = P'\beta
\end{aligned}
$$

เมื่อพิจารณาแมตริกซ์ Z จะพบว่า Z เป็น Orthogonal Matrix และ $X'X$ เป็น Orthogonal Matrix เพราะ

$$Z'Z = (XP)'(XP) = P'(X'X)P = D = \text{diag}(\lambda_1,\lambda_2,\ldots,\lambda_k)$$

และถ้าแมตริกซ์ X มี Rank เท่ากับ k - j (หมายความว่า X ไม่มี Full Rank) แมตริกซ์ $X'X$ จะมี Characteristic Value อยู่ *j ตัวที่มีค่าเป็น 0 มีผลให้ j สดมภ์สุดท้ายของ* XP เป็น Zero Vector ซึ่งย่อมทำให้เวคเตอร์ θ มีสมาชิกเพียง k - j ตัวหรือ $\theta_{k-j+1}, \theta_{k-j+2},\ldots,\theta_k$ เป็นพารามิเตอร์ที่ถูกต้องถูกขจัดออกไปจากสมการเสียก่อน เพราะไม่อาจประมาณค่าได้จากข้อมูลในแมตริกซ์ X ในทำนองกลับกันพารามิเตอร์ $\theta_1,\theta_2,\ldots,\theta_{k-j}$ จะเป็นสิ่งที่สามารถประมาณค่าได้จากข้อมูลในแมตริกซ์ X

ผลข้างต้นยืนยันว่าฟังก์ชั่น $W'\beta$ จะเป็นสิ่งที่เราสามารถประมาณค่าได้ก็ต่อเมื่อ $W'\beta$ เป็นฟังก์ชั่นที่แปลงรูปให้เป็น Linear Combination ของ $\theta_1,\theta_2,\ldots,\theta_{k-j}$ เท่านั้น กล่าวคือ

$$W'\beta = W'(PP')\beta = (P'W)'(P\beta) = (P'W)'\theta$$

แต่ $\theta_{k-j+1},\theta_{k-j+2},\ldots,\theta_k$ เป็นสิ่งที่ประมาณค่าไม่ได้ ดังนั้น $W'\beta$ จะประมาณได้ก็ต่อเมื่อสมาชิก j ตัวสุดท้ายของเวคเตอร์ $P'W$ ต้องเป็น 0

ผลการพิสูจน์ข้างต้นแสดงให้เห็นว่าเราควรแปลงรูปเวคเตอร์ W อย่างไรจึงจะมีผลให้ $W'\beta$ เป็นฟังก์ชั่นที่สามารถหาค่าประมาณได้ นอกจากนี้ผลพลอยได้ที่สำคัญอีกประการหนึ่งก็คือ การแปลงรูปเวคเตอร์ W ยังสามารถชี้ให้เห็นว่าฟังก์ชั่นใดพึงมีค่าประมาณที่แม่นยำอีกด้วย (ความจริงส่วนนี้จะเชื่อมโยงไปถึง Ridge Regression ที่จะได้กล่าวถึงต่อไป)

ถ้า p_1 เป็น Normalized Characteristic Vector ที่สอดคล้องกับราก λ_1 ของแมตริกซ์ $X'X$ ดังนั้นโดยอาศัยทฤษฎี 8.2 $W'\beta = p_1'\beta$ จะเป็นฟังก์ชั่นที่สามารถประมาณค่าได้ โดยที่ค่าประมาณของ $p_1'\beta$ ก็คือ $p_1'\hat{\beta}$ เมื่อ $\hat{\beta}$ คือรากของสมการ $X'X\hat{\beta} = X'Y$ และ $p_1'\hat{\beta}$ จะเป็น BLUE ของ $p_1'\beta$ (ทฤษฎี 8.1) โดยที่ $V(p_1'\hat{\beta}) = \frac{\sigma^2}{\lambda_1}$ ซึ่งสามารถพิสูจน์ให้เห็นได้ดังนี้

จากสมการนอร์มอล $X'X\hat{\beta} = X'Y$

ดังนั้น $p_1'X'X\hat{\beta} = p_1'X'Y$

แต่จากความรู้เรื่อง Characteristic Value เราทราบว่า $AX = \lambda X$ หรือ $X'A' = \lambda X'$ เมื่อ X คือ Characteristic Vector ที่สอดคล้องกับราก λ

ดังนั้นโดยอาศัยความรู้ดังกล่าวเราย่อมสรุปได้ว่า $p_1'(X'X) = \lambda_1 p_1'$ และ $p_1'(X'X)\hat{\beta}$

$= \lambda_1 p_1'\hat{\beta}$

นั่นคือ $p_1'X'X\hat{\beta} = p_1'X'Y = \lambda_1 p_1'\hat{\beta}$

สิ่งที่เราต้องการทราบคือ $V(p_1'\hat{\beta}) = ?$ ซึ่งจากสมการข้างบนนี้ทำให้เราทราบว่า

$$V(\lambda_1 p_1'\hat{\beta}) = V(p_1'X'X\hat{\beta}) = V(p_1'X'Y)$$

นั่นคือ

$$\lambda_1^2 V(p_1'\hat{\beta}) = V(p_1'X'Y) = (p_1'X')[V(Y)](p_1'X')'$$

$$= \sigma^2 p_1'(X'X)p_1$$

แต่ $p_1'(X'X)p_1 = \lambda_1$ หรือแมตริกซ์ $X'X$ Similar กับแมตริกซ์ $D = [\lambda_1]_{1x1}$

ดังนั้น
$$\lambda_1^2 V(p_1'\hat{\beta}) = \sigma^2 \lambda_1$$
$$V(p_1'\hat{\beta}) = \frac{\sigma^2}{\lambda_1}$$

---

[1] จากทฤษฎีใน Multivariate Normal Density ที่ $Y \sim N(\mu, V)$ ถ้า $Z = BY$ เราสามารถพิสูจน์ได้ว่า $Z \sim N(B\mu, BVB')$

นอกจากนี้เรายังสามารถพิสูจน์ได้ว่า $p_i'\hat{\beta}$ และ $p_j'\hat{\beta}$ เป็นอิสระต่อกัน คือ $Cov(p_i'\hat{\beta}, p_j'\hat{\beta}) = 0$ ในทุกค่า i, j ที่ $i \neq j$

และจากความรู้เหล่านี้ ถ้า $p_1, p_2, \ldots, p_{k-j}$ เป็น Characteristic Vector ที่สอดคล้องกับ $\lambda_1, \lambda_2, \ldots, \lambda_{k-j}$ ที่ไม่ใช่ 0 ของแมตริกซ์ $X'X$ และถ้าเราจัดฟังก์ชั่น $W'\beta$ อยู่ในรูป

$$W = K_1p_1+K_2p_2+\ldots+K_{k-j}p_{k-j}$$

เราย่อมสามารถพิสูจน์ได้ว่า

$$V(W'\hat{\beta}) = \sigma^2 \left(\frac{K_1^2}{\lambda_1} + \frac{K_2^2}{\lambda_2} + \ldots + \frac{K_{k-j}^2}{\lambda_{k-j}} \right)$$

จึงเห็นได้ว่า ถ้า K_j มีค่าสูง และ λ_j มีค่าต่ำ $W'\hat{\beta}$ จะขาดความแม่นยำ และถ้า λ_j มีค่าสูงเพียงใด $W'\hat{\beta}$ ก็จะยิ่งมีความแม่นยำสูงขึ้นเพียงนั้น

หมายเหตุ เมื่อ $W = K_1p_1+K_2p_2+\ldots+K_kp_k$ ดังนั้น

$$\begin{aligned} V(W'\hat{\beta}) &= W'V(\hat{\beta})(W')' \\ &= W'[\sigma^2(X'X)^{-1}]W \\ &= \sigma^2W'[\sum_{j=1}^{k} \frac{1}{\lambda_j} p_j p_j']W \\ &= \sigma^2\sum_{j=1}^{k} \frac{K_j^2}{\lambda_j} \end{aligned}$$

แสดงว่าปัญหา Multicolinearity มิได้มีผลให้สมการ $Y = X\beta+U$ ขาดลักษณะที่ดีที่พึงนำไปใช้ในงานพยากรณ์ ปัญหาที่เป็นผลกระทบจึงปรากฏในรูปอื่นที่กล่าวมาแล้ว

8.3 สาเหตุของปัญหา Multicolinearity

ในสถานการณ์ทางปฏิบัติที่มิใช่งานทดลองในห้องทดลองเรามักพบว่าตัวแปรอิสระต่าง ๆ มีความโน้มเอียงที่จะผันแปรไปในทิศทางเดียวกัน โดยเฉพาะในกรณีของอนุกรมเวลา ดังนั้นการตั้งข้อตกลงว่าตัวแปรอิสระทั้งหลายจะต้องเป็นอิสระต่อกัน จึงค่อนข้างจะฝืนสถานการณ์อยู่บ้าง[1] แต่ก็จำเป็นต้องตกลงไว้เช่นนั้นเพื่อผลในทางทฤษฎี (หากพบสถานการณ์ที่ขัดแย้งก็ให้พยายามหาทางปรับแก้วิธีประมาณค่า β ให้เหมาะสมต่อไปโดยไม่ทำลายข้อตกลงเดิม)

เราอาจกล่าวได้ว่าปัญหา Multicolinearity เกิดจากสาเหตุดังต่อไปนี้

1. ตัวแปรมีธรรมชาติที่จะผันแปรร่วมทิศทางเดียวกันอยู่แล้ว และยิ่งมีปรากฏการณ์ภายนอกเช่น ภัยธรรมชาติ สงคราม การนัดหยุดงาน ความผันผวนทางการเมือง ฯลฯ ตัวแปรจำนวนมากก็จะมีความโน้มเอียงที่จะผันแปรร่วมทิศทางเดียวกันมากยิ่งขึ้น เข้มข้นขึ้น บางครั้งถึงกับมีค่าเท่ากันในบางช่วงเวลา

2. การใช้ตัวแปรดัมมี่ไว้เท่ากับจำนวนกลุ่มย่อย (Category, Subsample) ของกลุ่มตัวอย่างทั้ง ๆ ที่แบบจำลองนั้นมีเทอมคงที่อยู่ (Intercept Term)

3. การใช้ Lagged Variable ในสมการถดถอย การใช้ Lagged Variable แม้จะมีส่วนช่วยให้แบบจำลองของเราสามารถเคลื่อนไหวรับได้กับความเปลี่ยนแปลงที่เกิดขึ้นตามกาลเวลา (ทำให้ Linear Model มีธรรมชาติของ Dynamic Model) แต่ค่าของ Lagged Variable ทั้งหลายจะเป็นค่าเดียวกัน แม้ในวาระเดียวกันจะมีค่าต่างกันแต่ค่าของ Lagged Variable ก็จะแสดงให้เห็นถึงการผันแปรค่าร่วมทิศทางกันเด่นชัดเป็นพิเศษ ตัวอย่างเช่นเราใช้ x_t ซึ่งมีค่าเท่ากับ 2, 3, 4, 5, 6, 7, 8, 9, 10 เป็นตัวแปรอิสระ ถ้าหากเพิ่ม x_{t-1} และ x_{t-2} ร่วมเป็นตัวแปรอิสระ จะพบว่าค่าของ x_t, x_{t-1} และ x_{t-2} ปรากฏดังนี้

---

[1] ข้อตกลงสมการถดถอยที่ตกลงว่าแมตริกซ์ X มี Full Column Rank แปลว่าสดมภ์ต่าง ๆ ของ X ต้องเป็นอิสระเชิงเส้นต่อกัน การตั้งข้อตกลงเช่นนี้มีเจตนาเพื่อบังคับให้แมตริกซ์ $(X'X)^{-1}$ มีค่าปรากฏ

วาระที่ (period)	X_t	X_{t-1}	X_{t-2}
3	4	3	2
4	5	4	3
5	6	5	4
6	7	6	5
7	8	7	6
8	9	8	7
9	10	9	8

ขอให้สังเกตว่า X_t, X_{t-1} และ X_{t-2} มีลักษณะของการผันแปรร่วมทาง (Covariation) กันอย่างเด่นชัดแม้ในวาระเดียวกัน จะให้ค่าต่างกัน แต่ก็มีค่าเดียวกันในต่างวาระ

ปัญหา Multicolinearity มักเกิดขึ้นกับข้อมูลอนุกรมเวลามากกว่าข้อมูลสำรวจเพราะตัวแปรมีธรรมชาติที่จะผันแปรร่วมทางกันตามกาลเวลา แม้จะมีผู้กล่าวว่าปัญหานี้มิใช่ปัญหาสำคัญในงานที่มุ่งเพื่อการพยากรณ์ (Klein, L.R., 1967) แต่ปัญหาอันเป็นผลกระทบทางสถิติ (ตอน 8.2) ก็มิใช่สิ่งที่ไม่มีความสำคัญ เรากล้าพอที่จะเฉยเมยต่อผลกระทบเหล่านี้หรือ...?

8.4 การตรวจสอบ

ในทางปฏิบัติเรามีเครื่องมือ และยุทธวิธีที่จะใช้ตรวจสอบว่าได้เกิดปัญหา Multicolinearity ขึ้นหรือไม่ รุนแรงเพียงใดได้หลายวิธีด้วยกัน แต่ในแต่ละวิธีก็ยังมีจุดอ่อนให้ท้วงติงได้ว่า ยังไม่ดีพอ หรือ ยังมีจุดอ่อนอยู่ ที่เป็นเช่นนี้เพราะเราไม่อาจควบคุมให้แมตริกซ์ X และ X'X มีลักษณะต่าง ๆ ได้ตามต้องการ เพราะแมตริกซ์ X และ X'X มีธรรมชาติที่ผันผวนไปได้มากมายตามธรรมชาติของตัวแปร X's และค่าของตัวแปรเหล่านั้น รวมตลอดถึงจังหวะเวลาที่ใช้ในการบันทึกข้อมูล ด้วยเหตุนี้นักวิจัยพึงใช้ยุทธวิธีในการตรวจจับปัญหา Multicolinearity ที่จะกล่าวถึงต่อไปหลาย ๆ วิธีผสมผสานกันไม่ควรใช้วิธีหนึ่งเป็นการเฉพาะเพราะอาจไม่ได้ผลเท่าที่ควร และในที่นี้จะกล่าวถึงเฉพาะกรณีที่ถือว่าตัวแปรอิสระเป็น Mathematical Variable เท่านั้น ผู้ที่สนใจกรณีของ Random Variable ให้ติดตามศึกษาได้จากเอกสารอ้างอิง

8.4.1 สหสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ

วิธีนี้กระทำโดยการวัดค่าสหสัมพันธ์ระหว่างตัวแปรอิสระคือ

$$r_{st} = \frac{\Sigma(x_{s_i}-\bar{x}_s)(x_{t_i}-\bar{x}_t)}{\sqrt{\Sigma(x_{s_i}-\bar{x}_s)^2\Sigma(x_{t_i}-\bar{x}_t)^2}} \quad ; s,t = 2,3,..,k \; ; s \neq t$$

วิธีนี้มีแนวคิดเดิมดังนี้คือ ถ้า $x_s = cx_t$ ซึ่งแสดงว่าเวคเตอร์ x_s และ x_t มี Linearly Dependent เราสามารถพิสูจน์ได้โดยง่ายได้ว่า $r_{st} = 1$ โดยนัยกลับกันถ้า $r_{st} = 1$ ย่อมแสดงว่าเวคเตอร์ x_s และ x_t มี Perfect Colinearity เหตุที่ $x_s = cx_t$ ผลสะท้อนก็คือแมตริกซ์ X ไม่มี Full Rank

ดังนั้น สหสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระจึงนับว่าเป็นดัชนีที่พอจะสามารถชี้ให้เห็นว่าได้เกิดปัญหา Multicolinearity ขึ้นหรือไม่ ในทางปฏิบัติเราจะถือว่า ถ้า $r_{st} > 0.80$ หรือ 0.90 คือสัญญาณอันตรายของงานเพราะถือว่ากำลังเกิดปัญหา Multicolinearity ที่รุนแรง นอกเหนือจากใช้เฉพาะ r_{st} เป็นเกณฑ์เราอาจใช้วิธีเปรียบเทียบค่า r_{st} กับ R^2 ซึ่งก็ยังคงอยู่ในข่ายของการใช้ข้อสนเทศจากสหสัมพันธ์เชิงเส้น ตรวจสอบปัญหา Multicolinearity อยู่เพียงแต่ปรับปรุงให้ดีขึ้น กติกาใหม่นี้เราจะถือว่าเมื่อใดก็ตามที่ $r_{st} > R^2$ แสดงว่าได้เกิดปัญหา Multicolinearity ที่รุนแรงขึ้นแล้วให้เร่งแก้ไขเป็นการด่วน

วิธีการเหล่านี้แม้จะมีข้อดีอยู่ในแง่ที่ช่วยชี้ให้เห็น Linearity Dependent หรือ Near Dependent ระหว่างเวคเตอร์ แต่ถ้าตัวแปรหลายคู่ให้ค่าสหสัมพันธ์สูงกว่า 0.80 หรือ 0.90 หรือสูงกว่า R^2 นักวิจัยจะเริ่มสับสน และไม่อาจตัดสินใจได้ว่าตัวแปรใดกันแน่ที่พึ่งพิงตัวแปรอื่น หรือถ้าจะกล่าวในแง่มุมของผู้มองโลกในแง่ร้ายก็คือ "วิธีนี้ดีแต่ชี้ให้เห็นปัญหา แต่ไม่ยอมชี้ทางแก้ปัญหา"

8.4.2 Partial Regression

คำว่า Partial Regression หมายถึงการวิเคราะห์สมการถดถอย Y = f(X's) ทั้งสิ้น k - 1 ชุด แต่ละชุดใช้ตัวแปรอิสระ k - 2 ตัว โดยสับเปลี่ยนตัวแปรอิสระออกจากสมการเดิมคราวละ 1 ตัว แล้วคำนวณหา R^2 ของแต่ละสมการแล้วนำไปเปรียบเทียบกับ R^2 ของสมการเดิม (เต็มรูป) เช่น

$$Y = \beta_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+u$$

ซึ่งเป็นสมการเต็มรูป สมมุติคำนวณหา Multiple Correlation ได้เท่ากับ R^2 จากสมการเต็มรูปนี้ให้สับเปลี่ยนเอาตัวแปรอิสระออกจากสมการแล้ววิเคราะห์สมการถดถอยระหว่าง Y กับ X's ที่เหลืออยู่แล้วคำนวณหา R^2 ในกรณีนี้จะพบว่ามี Partial Regression อยู่ 4 ชุดคือ

$Y = \beta_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+u$ ได้ R^2 : (คัด X_5 ทิ้ง)

$Y = \beta_1+\beta_2X_2+\beta_3X_3+\beta_5X_5+u$ ได้ R^2 : (คัด X_4 ทิ้ง)

$Y = \beta_1+\beta_2X_2+\beta_4X_4+\beta_5X_5+u$ ได้ R^2 : (คัด X_3 ทิ้ง)

$Y = \beta_1+\beta_3X_3+\beta_4X_4+\beta_5X_5+u$ ได้ R^2 : (คัด X_2 ทิ้ง)

การตัดสินใจให้กระทำดังนี้คือ ถ้า R^2 จากสมการเต็มรูปมีค่าใกล้เคียงกับ R^2 จาก Partial Regression แสดงว่าเกิดปัญหา Multicolinearity ที่รุนแรง

ความจริงแล้วแนวคิดของวิธี Partial Regression พัฒนามาจากเทคนิคการหา Best subset ของตัวแปรอิสระนั้นเอง (ขอให้ย้อนไปศึกษาเรื่อง SW, FS, BE ในตอน 4.5.2.2) กล่าวคือ ถ้า Sub-Model ใดให้ค่า R^2 สูงอยู่แล้ว เมื่อเพิ่มตัวแปรอิสระอื่นเข้าไปให้เป็น Full-Model กลับมิได้ช่วยให้ R^2 ของ Full-Model สูงกว่าที่เป็นอยู่เดิมก็แสดงว่าตัวแปรที่เพิ่มเข้าไปใหม่นั้นเป็นตัวแปรที่ไม่มีประโยชน์ในเชิงสร้างสรรค์แต่ประการใด เพราะอาจเป็นไปได้ที่ตัวแปรที่เพิ่มเข้าสู่สมการดังกล่าวมีธรรมชาติที่คล้ายคลึงกับตัวแปรอิสระตัวใดตัวหนึ่งที่มีอยู่เดิม

สำหรับวิธีนี้ ถ้ามองกันโดยผิวเผินก็อาจจะเห็นว่าเป็นวิธีที่ดี แต่ถ้าจะพิจารณาให้ละเอียดตามย่อหน้าข้างบนนี้จะพบว่าการที่ R^2 ของ Full-Model มิได้สูงขึ้นกว่าที่เคยเป็นอยู่เดิมใน Sub-Model นั้นมิได้หมายความว่าตัวแปรอิสระตัวใหม่ที่เพิ่มเข้าสู่สมการมีความสัมพันธ์เชิงเส้นกับตัวแปรที่มีอยู่เดิมเสมอไป แต่ตัวแปรตัวใหม่ที่กล่าวถึงนั้นอาจมิใช่ปัจจัยที่มีความหมายหรือความสำคัญในการควบคุมความเคลื่อนไหวของ Y ก็ได้ และถ้าหากว่า R^2 จาก Sub-Model หลายชุดมีค่าใกล้เคียงกับ R^2 ของ Full-Model นักวิจัยจะเริ่มรวนเรและตัดสินใจไม่ได้ วิธีนี้จึงมีจุดอ่อนอยู่มาก

8.4.3 F - test และ t - test

ในการวิเคราะห์สมการถดถอย $Y = \beta_1+\beta_2X_2+\ldots+\beta_kX_k+u$ นั้นเรามีความจำเป็นต้องตรวจสอบนัยสำคัญอย่างน้อย 2 ระยะ คือ

1. ทดสอบ $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ vs $H_1 : \beta_j$ not all equal; $j = 2, 3, \ldots, k$

โดยใช้ตัวสถิติ F คือ $F_c = \dfrac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - \hat{\beta}'X'Y}$ เป็นเกณฑ์ในการตัดสินใจ วิธีนี้เรียกว่า Overall F-test โดยจะมีเจตนาที่จะตรวจดูภาพกว้าง ๆ ของงานว่าสมการถดถอยมีวิสัยที่อยู่ในเกณฑ์ที่พอจะยอมรับได้หรือไม่ หรืออีกนัยหนึ่งก็คือตัวแปรอิสระ X's ทั้งหลายที่กำหนดขึ้นนั้นร่วมกันควบคุมความเคลื่อนไหวของ Y ได้หรือไม่

2. ทดสอบ $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$; $i = 1, 2, \ldots, k$ โดยใช้ตัวสถิติ $t_j = \dfrac{\hat{\beta}_j}{\sqrt{\hat{V}(\hat{\beta})}}$ เป็นเกณฑ์ในการตัดสินใจ การทดสอบวิธีนี้มีเจตนาที่จะวิเคราะห์ในรายละเอียดว่า ตัวแปรอิสระแต่ละตัวมีอิทธิพลต่อ Y แท้จริงหรือไม่ (มีนัยสำคัญหรือไม่) มากน้อยเพียงไร (ระดับนัยสำคัญ α มีค่าเท่าไร)

ในสถานการณ์ทางปฏิบัตินั้น หากพบว่าสมการ $Y = X\beta + U$ ให้ค่า R^2 ค่อนข้างสูง F มีนัยสำคัญ (คือยอมรับ $H_1 : \beta_j$ not all zero) แต่กลับพบว่า t_j ไม่มีนัยสำคัญ (คือยอมรับ $H_0 : \beta_j = 0$) บางส่วนหรือทั้งหมด จะถือว่าได้เกิดปัญหา Multicolinearity ขึ้นแล้ว

วิธีนี้แม้จะเป็นที่ยอมรับว่าพอจะใช้ตรวจสอบปัญหา Multicolinearity ได้แต่ในบางครั้งแม้ว่า F จะมีนัยสำคัญขณะที่ t ไม่มีนัยสำคัญก็มิได้ยืนยันว่าตัวแปรอิสระจะพึ่งพิงเชิงเส้นต่อกันเสมอได้ และวิธีนี้มิได้ชี้ให้เห็นว่าเกิด Linearly Dependent หรือ Near Dependency ที่จุดใด

8.4.4 Auxilary Regression

แนวคิดของ Auxilary Regression พัฒนามาจากความรู้เรื่อง Linearly Dependent และ Basis ของ Vector Space ดังนี้

"ถ้า $\alpha_1, \alpha_2, \ldots, \alpha_n$ เป็น Basis ของ Vector Space $\mathcal{V}$ แล้วเวคเตอร์ V ใด ๆ ใน $\mathcal{V}$ จะเกิดขึ้นจากการประกอบกันเชิงเส้นของเวคเตอร์ใน Basis เสมอ กล่าวคือ $V = \alpha_1 V_1 + \alpha_2 V_2 + \ldots + \alpha_n V_n$ เมื่อ $\alpha_1, \alpha_2, \ldots, \alpha_n$ เป็นสมาชิกของ Field $\mathcal{F}$ และ α_i ; $i = 1, 2, \ldots, n$ มิใช่ 0 ทั้งหมด" และ "ถ้า $\mathcal{S}$ เป็นอนุเซทของ $\mathcal{V}$ โดยที่สมาชิกของ $\mathcal{S}$ มี Linearly Dependent แล้ว เวคเตอร์ V ย่อมเกิดขึ้นจากการประกอบกันของเวคเตอร์ใน $\mathcal{S}$ ได้"[1]

---

[1] มนตรี พิริยะกุล, **คณิตศาสตร์สำหรับนักสถิติ (พีชคณิตเชิงเส้น)**, (สำนักพิมพ์กราฟิคอาร์ต, กรุงเทพฯ, 2525), หน้า 204 อ่านเพิ่มเติมบทที่ 4 Vector Space

โดยอาศัยความรู้ดังกล่าวเราจึงสามารถวัดดูว่าตัวแปรอิสระมีความสัมพันธ์เชิงเส้น (Linearly Dependent) ต่อกันหรือไม่...? มากน้อยเพียงใด...? ด้วยการวิเคราะห์สมการถดถอยระหว่างตัวแปรอิสระ X_j กับตัวแปรอิสระอื่น ๆ คือ $X_j = f(X_2, X_3, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k)$; $j = 2, 3, \ldots, k$ แล้วหาค่าสหสัมพันธ์พหุคูณ $R^2_{j.23\ldots j-1, j+1\ldots k}$ (ต่อไปจะใช้เขียนในรูปย่อคือ R_j^2) ถ้า R_j^2 มีค่าสูงแสดงว่าเกิดปัญหา Multicolinearity ขึ้น เช่นจากความสัมพันธ์เดิมคือ $Y = f(X_2, X_3, X_4, X_5)$ ให้นักวิจัยคำนวณหาค่า R_2^2, R_3^2, R_4^2 และ R_5^2 จากสมการ $X_2 = f(X_3, X_4, X_5)$, $X_3 = f(X_2, X_4, X_5)$, $X_4 = f(X_2, X_3, X_5)$ และ $X_5 = f(X_2, X_3, X_4)$ ตามลำดับ หากพบว่า R_2^2 มีค่าสูง แสดงว่า X_2 เกิดจากการประกอบกันเชิงเส้นของเวคเตอร์ X_3, X_4 และ X_5 ถ้า R_5^2 มีค่าสูงแสดงว่า X_5 เกิดจากการประกอบกันเชิงเส้นของเวคเตอร์ X_2, X_3 และ X_4 ดังนี้เป็นต้น

การตรวจสอบโดยวิธีนี้ชี้ให้เห็นเค้าโครงความสัมพันธ์ระหว่างเวคเตอร์ได้ใกล้เคียงกับความเป็นจริงมาก แต่ถ้า R_j^2 หลายค่ามีค่าสูงเราจะพบปัญหาเพราะไม่อาจจำแนกได้ว่าความสัมพันธ์รูปใดคือความสัมพันธ์ที่แท้จริงระหว่างตัวแปรอิสระ

8.4.5 Variance Inflation Factor (VIF)

วิธีนี้เกี่ยวเนื่องอยู่กับวิธี 8.4.4 โดยเราจะศึกษาโครงสร้างของแมตริกซ์ $\sigma^2(X'X)^{-1}$ *โดยอาศัยความรู้เกี่ยวกับ* R_j^2 *กล่าวคือเราสามารถพิสูจน์ได้ว่าสมาชิกใน Main Diagonal ของแมตริกซ์* $(X'X)^{-1}$ คือ $\frac{1}{(1-R_j^2)}$ เมื่อ R_j^2 ; $j = 2,3,\ldots,k$ คือสัมประสิทธิ์ R^2 ในข้อ 8.4.4 นั่นคือ

$$c_{jj} = \frac{1}{(1-R_j^2)} \quad ; \ j = 2,3,\ldots,k$$ [1]

เมื่อ c_{jj} คือสมาชิก ณ.ตำแหน่ง (j, j) ของแมตริกซ์ $(X'X)^{-1}$

จาก $c_{jj} = \frac{1}{(1-R_j^2)}$; $j = 2,3,\ldots,k$ จะพบว่าถ้า $R_j^2 \to 1$ แล้ว c_{jj} จะมีค่าสูงมีผลให้ $V(\hat{\beta}_j) = \sigma^2 c_{jj}$ มีค่าสูงซึ่งแสดงว่า X_j คือตัวแปรที่เกิดจากการประกอบกันเชิงเส้นของตัวแปรอิสระอื่น ๆ (ขอให้สังเกตว่าวิธี 8.4.4 และ 8.4.5 ให้ผลสรุปเดียวกัน) สิ่งที่เป็นปัญหา

---

[1] อ่านตอน 4.4.2 ซึ่งกล่าวว่า $\hat{V}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\Sigma x_{ji}^2 (1-R_j^2)}$ แสดงว่า $c_{jj} = \frac{1}{(1-R_j^2)} = \frac{\hat{V}(\hat{\beta}_j)\Sigma x_{ji}^2}{\hat{\sigma}^2}$

ก็คือ ถ้า c_{jj} หลายค่ามีค่าสูงจะมีปัญหาในการตรวจสอบปัญหา Multicolinearity เช่นเดียวกับกรณี 8.4.4 วิธีนี้พัฒนาโดย D.E. Marquardt และ R.D. Snee ในปี 1970, 1973 และ 1975 โดยทั้ง Marquardt และ Snee ถือว่าวิธีนี้คือวิธีเดียวที่ดีที่สุด[1]

8.4.6 การวิเคราะห์ค่าดีเทอร์มิแนนท์ของแมตริกซ์ $X'X$

จากการศึกษาในตอน 4.4 เราทราบว่าแมตริกซ์ $X'X$ สำหรับสมการถดถอยต่อไปนี้คือ

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + u_i - \bar{u} \quad ; i = 1,2,\ldots,n$$

เมื่อ $y_i = Y_i - \bar{Y}$ และ $x_{ji} = X_{ji} - \bar{X}_j$; $i = 1,2,\ldots,n$; $j = 2,3,\ldots,k$

จะสามารถแปลงรูปเป็น Correlation Matrix ได้โดยง่าย และจากสมการแปลงรูปข้างต้น ถ้าแปลงรูปตัวแปรทุกตัวด้วยการถ่วงน้ำหนักด้วย Standard Deviation คือ

$$\frac{y_i}{s_1} = \beta_2 \frac{x_{2i}}{s_2} + \beta_3 \frac{x_{3i}}{s_3} + \ldots + \beta_k \frac{x_{ki}}{s_k} + \frac{u_i - \bar{u}}{s_u} \; ; \; i = 1,2,\ldots,n$$

จะพบว่าแมตริกซ์ $(X'X)$ คือ Correlation Matrix R_X ที่สมาชิกของแมตริกซ์ดังกล่าวคือสหสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ

ดังเหตุผลดังกล่าวการวิเคราะห์ค่าดีเทอร์มิแนนท์ของแมตริกซ์ $(X'X)$ และการวิเคราะห์ค่าดีเทอร์มิแนนท์ของแมตริกซ์ R_X จึงมีผลลัพธ์ตรงกันกล่าวคือ

---

[1] Judge, G.G. et. al., **Op.Cit.,** p.475 และ Webster, J.T., Gunst, R.F. and Mason, R.L., "Recent Development in Stepwise Regression Procedures" (The Univ. of Kentucky Conference on Regression with a Large Number of Predictor Variables, Lexington, Kentucky, Oct. 11-12, 1973). p.37

$$X'X = \begin{vmatrix} \frac{x_{21}}{s_2} & \frac{x_{22}}{s_2} & \frac{x_{23}}{s_2} & \cdots & \frac{x_{2n}}{s_2} \\ \frac{x_{31}}{s_3} & \frac{x_{32}}{s_3} & \frac{x_{33}}{s_3} & \cdots & \frac{x_{3n}}{s_3} \\ \\ \frac{x_{k1}}{s_k} & \frac{x_{k2}}{s_k} & \frac{x_{k3}}{s_k} & \cdots & \frac{x_{kn}}{s_k} \end{vmatrix} \begin{vmatrix} \frac{x_{21}}{s_2} & \frac{x_{31}}{s_3} & \cdots & \frac{x_{k1}}{s_k} \\ \frac{x_{22}}{s_2} & \frac{x_{32}}{s_3} & \cdots & \frac{x_{k2}}{s_k} \\ \frac{x_{23}}{s_2} & \frac{x_{33}}{s_3} & \cdots & \frac{x_{k3}}{s_k} \\ \vdots & \vdots & & \\ \frac{x_{2n}}{s_2} & \frac{x_{3n}}{s_3} & \cdots & \frac{x_{kn}}{s_k} \end{vmatrix}$$

$$= \begin{vmatrix} \Sigma \frac{x_{2i}^2}{s_2^2} & \Sigma \frac{x_{2i}x_{3i}}{s_2 s_3} & \cdots & \Sigma \frac{x_{2i}x_{ki}}{s_2 s_k} \\ \Sigma \frac{x_{2i}x_{3i}}{s_2 s_3} & \Sigma \frac{x_{3i}^2}{s_3^2} & \cdots & \Sigma \frac{x_{3i}x_{ki}}{s_3 s_k} \\ \\ \Sigma \frac{x_{2i}x_{ki}}{s_2 s_k} & \Sigma \frac{x_{3i}x_{ki}}{s_3 s_k} & \cdots & \Sigma \frac{x_{ki}^2}{s_k^2} \end{vmatrix}$$

$$= \begin{vmatrix} 1 & r_{23} & r_{24} & \cdots & r_{2k} \\ r_{23} & 1 & r_{34} & \cdots & r_{3k} \\ r_{24} & r_{34} & 1 & & r_{4k} \\ \\ r_{2k} & r_{3k} & r_{4k} & \cdots & 1 \end{vmatrix}$$

$$= R_x$$

การวิเคราะห์ดีเทอร์มิแนนท์ของแมตริกซ์ $X'X$ สามารถกระทำได้ 2 แนวทางดังนี้

แนวทางที่ 1 การวิเคราะห์เฉพาะ $|X'X|$

จากการบรรยายโครงสร้างของแมตริกซ์ $X'X$ เมื่อตัวแปรได้รับการจัดรูปสู่รูปมาตรฐาน (Standardized Variable) จะพบว่า $|X'X|$ มีค่าปรากฏระหว่าง 0 ถึง 1 กล่าวคือ ถ้าตัวแปรอิสระ X_i และ X_j ทุกคู่มี Perfect Correlation สมาชิกของ $X'X$ จะเป็น 1 ทุกตัว มีผลให้ $|X'X| = 0$ ในทางตรงกันข้ามถ้าตัวแปรอิสระ X_i และ X_j ไม่มีความสัมพันธ์กันเลย คือ $r_{ij} = 0$ แมตริกซ์ $X'X$ จะเป็น Identity Matrix ซึ่งมีผลให้ $|X'X| = 1$ ด้วยเหตุนี้หากนักวิจัยตรวจสอบ $|X'X|$ แล้วพบว่า $|X'X|$ เท่ากับ 0 หรือใกล้ 0 แสดงว่าตัวแปรอิสระบางคู่มีความสัมพันธ์เชิงเส้นต่อกัน

วิธีนี้แม้จะพอชี้ให้เห็นปัญหา Multicolinearity ได้แต่ก็ยังมีจุดอ่อนหลายประการดังนี้

(1) $|X'X| \to 0$ ชี้ให้เห็นว่าได้เกิดปัญหา Multicolinearity ขึ้นแล้วเท่านั้น แต่มิได้ชี้ให้เห็นว่า Multicolinearity เกิดขึ้นเพราะเหตุใด? ตัวแปรอิสระใดบ้างเกิดขึ้นจากการประกอบกันเชิงเส้นของตัวแปรอื่น? มีตัวแปรอิสระใดบ้างทำหน้าที่เสมือน Basis ของ Column Space ของแมตริกซ์ X

(2) โดยอาศัยความรู้เรื่อง Characteristic Vector ในตอน 8.1 เราสามารถพิสูจน์ได้โดยง่ายว่า

$$|X'X| = \lambda_1 \lambda_2 \cdots \lambda_k = \prod_{i}^{k} \lambda_i$$

แสดงว่าในกรณีที่ $|X'X| \to 0$ อาจเนื่องมาจากสาเหตุอื่นคือ Characteristic Root ของ $X'X$ บางตัวมีค่าต่ำ การสรุปว่ามีปัญหา Multicolinearity โดยยึดถือเอาจากสถานการณ์ที่ $|X'X| \to 0$ จึงยังมีข้อโต้แย้ง

แนวทางที่ 2 Farrar - Glauber Test

ฟาร์ราร์ และ กลอเบอร์ (Farrar, D.E. and Glauber, R.R., 1967)[1] ได้พยายามหาหนทางสร้างเกณฑ์การทดสอบ Multicolinearity โดยการเปรียบเทียบค่า $|X'X|$ กับค่าวิกฤตที่เหมาะสม มิใช่เพียงแต่ตรวจดูว่า $|X'X|$ มีค่าใกล้ 0 หรือใกล้ 1 ดังแนวทางที่ 1 ซึ่งโน้มเอียงไปในเชิงอัตตนิยมมากเกินไป โดยฟาร์ราร์และกลอเบอร์มุ่งตรวจดูว่าแมตริกซ์ X เบนห่างจากธรรมชาติของ Orthogonality หรือไม่ รุนแรงเพียงใด ตัวแปร (สดมภ์ของแมตริกซ์ X) ใดเกิดจากการประกอบ

---

[1] Kontsoyiannis, A., Op. Cit.. p.234

กันของตัวแปรอื่น และตัวแปรใดบ้างคือปัจจัยที่มีส่วนร่วมในการสร้างปัญหา Multicolinearity (หมายความว่าเวคเตอร์ใดบ้างทำหน้าที่เสมือน Basis ของ Column Space ของ X) เครื่องมือทดสอบปัญหา Multicolinearity ตามวิธีการของฟาร์ราร์และกลอเบอร์จึงประกอบด้วย test ต่าง ๆ ต่อไปนี้ ซึ่งจะต้องทดสอบอย่างต่อเนื่อง

1. χ^2-test สำหรับตรวจสอบว่ามีปัญหา Multicolinearity หรือไม่ รุนแรงเพียงใด

ปฏิเสธสมมุติฐานหลัก H_0 : ตัวแปร X's เป็นอิสระต่อกัน VS H_1 : ตัวแปร X's สัมพันธ์กัน ณ ระดับนัยสำคัญ α เมื่อ $^*\chi^2 > \chi^2_{1-\alpha,\nu}$

โดยที่

$$^*\chi^2 = -[n-1-\frac{1}{6}(2k+5)] \log_e |X'X|$$

เมื่อ n = ขนาดตัวอย่าง, k = จำนวนตัวแปรอิสระ

$|X'X|$ = ค่าดีเทอร์มิแนนท์ของแมตริกซ์ $X'X$ ที่ตัวแปรทุกตัวได้รับการแปลงรูปเป็น Standardized Variable หรือนัยหนึ่ง $|X'X| = |R_X|$

และ $\nu = \frac{1}{2}k(k-1)$

ถ้า $^*\chi^2 > \chi^2_{1-\alpha,\nu}$ แสดงว่าตัวแปรอิสระ X's ไม่เป็นอิสระต่อกัน (not Orthogonal) และยิ่ง $^*\chi^2$ มีค่าสูงกว่า $\chi^2_{1-\alpha,\nu}$ มากเพียงใด ปัญหา Multicolinearity จะยิ่งรุนแรงมากเพียงนั้น

เมื่อ $^*\chi^2 > \chi^2_{1-\alpha,\nu}$ ให้ทดสอบขั้นที่ 2 และ 3 ต่อไป

2. F - Test สำหรับตรวจสอบว่าตัวแปรอิสระใดเกิดจากการประกอบกันเชิงเส้นของตัวแปรอิสระอื่น

การตรวจสอบในชั้นนี้เพื่อตรวจดูว่าตัวแปรอิสระตัวใดเกิดขึ้นจากการประกอบกันเชิงเส้นของตัวแปรอื่นหรือไม่โดยตรวจสอบ R^2_j (หรือ $R^2_{j.23..,j-1,j+1,..k}$) ว่ามีค่าเท่ากับเท่าไร (อ่านตอน 8.4.4) แตกต่างจาก 0 อย่างมีนัยสำคัญหรือไม่[1] สมมุติฐานที่มุ่งทดสอบสำหรับกรณีนี้คือ

---

[1] ในตอน 8.4.4 เราพิจารณาแต่เพียงว่า R^2_j มีค่าสูงหรือไม่ ถ้า R^2_j มีค่าสูงแสดงว่าได้เกิดปัญหา Multicolinearity ขึ้นโดย X_j เกิดขึ้นจากการประกอบกันเชิงเส้นระหว่างตัวแปรอิสระอื่น ๆ ที่เหลืออยู่ แต่วิธีนี้มีลักษณะโน้มเอียงไปในเชิงอัตตนิยม (Subjective Judgement) เพราะไม่มีกฎเกณฑ์ใดบังคับไว้อย่างชัดเจนว่า R^2_j สูงเพียงใดจึงจะนับว่าเกิดปัญหา Multicolinearity ซึ่งเป็นสาเหตุให้ ฟาร์ราร์ และ กลอเบอร์สร้างเกณฑ์เปรียบเทียบขึ้นใช้สำหรับกรณีดังกล่าว

$$H_0 : R_j^2 = 0 \quad VS \quad H_1 : R_j^2 \neq 0$$

หรือเขียนให้เต็มรูปได้ดังนี้

$$H_0 : R^2_{j.23...j-1,j+1...k} = 0 \text{ VS } H_1 : R^2_{j.23...j-1,j+1...k} \neq 0$$

โดยเราจะปฏิเสธ H_0 ณ ระดับนัยสำคัญ เมื่อ $F^* > F_{k-1,n-k,1-\alpha}$ โดยที่

$$F^* = \frac{R_j^2/k-1}{(1-R_j^2)/n-k}$$

3. t - test เพื่อตรวจสอบดูว่าตัวแปรอิสระใดมีส่วนร่วมในการสร้างปัญหา Multi-colinearity

การตรวจสอบในขั้นนี้เป็นการตรวจสอบดูว่าตัวแปรอิสระคู่ใดบ้างที่มีความสัมพันธ์กัน โดยการดำเนินการตรวจสอบ จะดำเนินการตรวจสอบสหสัมพันธ์ที่แท้จริงคือ Partial Correlation ระหว่างตัวแบ่รอิสระ X_i กับ X_j ; $i \neq j = 2,3,\ldots,k$ ทีละคู่ทั้งสิ้น $\binom{k-1}{2} = \frac{(k-1)(k-2)}{2}$ คู่ Partial Correlation ระหว่างตัวแปรอิสระคู่ใดแตกต่างจาก 0 อย่างมีนัยสำคัญ แสดงว่าตัวแปรอิสระคู่นั้น ๆ ต้องรับผิดชอบในปัญหา Multicolinearity สมมุติฐานและเกณฑ์การตัดสินใจปรากฏดังนี้

$$H_0 : R_{ij.23...k} = 0 \text{ VS } H_1 : R_{ij.23...k} \neq 0$$

ปฏิเสธสมมุติฐานหลัก ณ ระดับนัยสำคัญ α เมื่อ $|t^*| > t_{n-k,1-\frac{\alpha}{2}}$ โดยที่

$$t^* = \frac{(R_{ij.23...k})\sqrt{n-k}}{\sqrt{1-R^2_{ij.23...k}}}$$

ถ้าปฏิเสธ H_0 แสดงว่า $R_{ij.23...k} \neq 0$ หรือ X_i และ X_j มีความสัมพันธ์กันอย่างเด่นชัด หรือ X_i และ X_j มีส่วนรับผิดชอบต่อปัญหา Multicolinearity[1]

Farrar - Glauber Test ได้รับความนิยมค่อนข้างกว้างขวางเพราะสามารถระบุ หรือชี้ให้เห็นปัญหา ความรุนแรงของปัญหา ตัวแปรที่เป็นปัญหาและตัวแปรที่ต้องรับผิดชอบต่อปัญหา อย่างไรก็ตาม Farrar - Glauber Test มีจุดอ่อนที่สำคัญที่ทำให้เกิดข้อโต้แย้งได้ดังนี้

(1) ฟาราร์ และกลอเบอร์ถือว่าค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i}, ..., X_{ki})$; $i = 1,2,...,n$ เป็นอิสระต่อกัน และสุ่มมาจากกลุ่มประชากรปกติ (Multivariate Normal Density) จึงมีผลให้ตัวสถิติ $-[n-1-\frac{1}{6}(2k+5)]\ell n|X'X|$ มีการแจกแจงแบบ χ^2 ได้ แต่ในทางปฏิบัติแล้วตัวแปรต่าง ๆ มิได้เป็นอิสระต่อกัน แต่กลับมีความสัมพันธ์กันอยู่บ้างในบางระดับเสมอ ซึ่งในกรณีเช่นนี้ตัวสถิติ $-[n-1-\frac{1}{6}(2k+5)\ln|X'X|$ จึงมิได้มีการแจกแจงแบบ χ^2 แม้เราจะยอมให้ $Z_i \sim N(\mu, V)$ ได้ก็ตาม

(2) H : ตัวแปรอิสระ X's เป็นอิสระต่อกัน (Orthogonal) หมายถึง X's ทั้งมวลประชากรคือ X's ที่พึงเป็นไปได้ทั้งหมดที่ควบคุมความเคลื่อนไหวของ Y หรือถ้ามองที่แมตริกซ์ $X'X$ (ซึ่งก็คือ R_X นั่นเอง) คำว่า Orthogonal หมายถึง R_X เป็น Population Correlation Matrix การสรุปหรืออนุมานสู่ Population Correlation Matrix จากข้อมูลของกลุ่มตัวอย่างซึ่งมิได้เป็นไปโดยสุ่ม[2] จึงเป็นเรื่องที่ไม่ถูกต้อง และมีความขัดแย้งกันในระหว่างเหตุผลทางทฤษฎี และวิธีดำเนินการในทางปฏิบัติ

8.5 การแก้ปัญหา Multicolinearity

การแก้ปัญหา Multicolinearity สามารถกระทำได้หลายวิธีดังนี้

---

[1] อ่านตอน 4.4 2 เรื่องเทคนิคการคำนวณหา $R^2_{1.23...k}$, R^2_j และ R^2_{ij}

[2] ขอให้ระลึกว่าเรากำหนดตัวแปรอิสระโดยอาศัยกฎเกณฑ์ทางทฤษฎี การคิดคำนึงและความเป็นไปได้ต่าง ๆ มิใช่กำหนดโดยสุ่ม

8.5.1 การเพิ่มข้อมูล

การเพิ่มข้อมูลเป็นวิธีหนึ่งที่สามารถแก้ปัญหา Multicolinearity ได้เพราะข้อมูลที่เพิ่มให้แก่แต่ละเวคเตอร์จะมีผลให้ธรรมชาติเดิมของเวคเตอร์เปลี่ยนแปลงไป แต่อย่างไรก็ตาม การเพิ่มข้อมูลให้แก่แต่ละเวคเตอร์โดยตรงอาจไม่ช่วยให้สถานการณ์ดีขึ้น ดังนั้นทางออกที่เหมาะสมก็คือการกำหนดเงื่อนไขให้แก่ข้อมูลที่จะเพิ่มเข้ามาใหม่ดังแนวคิดต่อไปนี้

สมมุติว่าแบบจำลอง $Y = X\beta + U$ มี Linear Dependency ปรากฏขึ้นระหว่าง Column Vector ของแมตริกซ์ X ($X'X$ ไม่มี Full Rank และมี Zero Characteristic Root)[1] และสมมุติว่ามีตัวแปรอิสระเพียงตัวเดียวเท่านั้น ที่เป็น Linear Combination ของเวคเตอร์อื่น ๆ ซึ่งแสดงว่า จะมี Normalized Characteristic Vector เพียงชุดเดียวคือ p ที่มีผลให้ $X'Xp = 0$[2] และทำให้ Linear Combination ของพารามิเตอร์คือ $p'\beta$ เป็นพารามิเตอร์ที่ไม่อาจประมาณค่าได้

แต่เนื่องจาก $X'Xp = 0$ ก็ต่อเมื่อ $Xp = 0$ หรือ Xp เป็น Zero Vector แสดงว่าปัญหา Multicolinearity นี้เกิดขึ้นได้เนื่องมาจากเวคเตอร์ต่าง ๆ ที่ประกอบเป็นแมตริกซ์ X ตั้งได้ฉาก (Orthogonal) กับเวคเตอร์ p ด้วยเหตุนี้ถ้าเราสามารถทำให้เวคเตอร์ต่าง ๆ ไม่ตั้งฉากกับเวคเตอร์ p เราก็ย่อมแก้ปัญหา Multicolinearity ได้ ซึ่งจากทางออกนี้แสดงว่าเราควรเพิ่มข้อมูลคือเวคเตอร์ $Z'_{n+1} = (1, X_{2,n+1}, X_{3,n+1}, \ldots, X_{k,n+1})$ ในลักษณะที่ $Z'_{n+1} = cp$ เมื่อ c คือตัวคงที่ใด ๆ

และในกรณีทั่วไปคือแมตริกซ์ $X'X$ ไม่ถึงกับเป็น Singular Matrix แต่เป็นเพียง Near Singular หรือ Characteristic Root บางค่ามีค่าต่ำ สมมุติ Characteristic Root ดังกล่าวคือ λ และ Normalized Characteristic Vector ที่สอดคล้องกับ λ คือ p จะพบว่าค่าสังเกตชุดใหม่ที่จะสุ่มเพิ่มเติมก็คือ ค่าสังเกตที่ควรขนานกับเวคเตอร์ p คือ $Z'_{n+1} = cp$ เช่นกัน

กล่าวโดยสรุปแล้ว การเพิ่มข้อมูลอาจช่วยแก้ปัญหา Multicolinearity ได้ และถ้าจะให้แน่ใจว่าจะสามารถแก้ปัญหาและเพิ่มความถูกต้องแม่นยำให้แก่ค่าประมาณของ β ค่าสังเกตที่เพิ่มเติมนั้นจะต้องได้รับการคัดเลือกมาในลักษณะที่เป็นพหุคูณของ Normalized Characteristic Vector ที่สอดคล้องกับ Characterist Root ที่มีค่าต่ำสุด ภารกิจของนักวิจัยที่จะแก้ปัญหาโดยวิธีนี้ก็คือ

---

1 ถ้าแมตริกซ์ A ใด ๆ มี Zero Root แสดงว่าแมตริกซ์นั้นเป็น Singular Matrix ดังทฤษฎีต่อไปนี้ "Singular Matrix A ใด ๆ จะมี Zero Characteristic Root เสมอ"

2 $AY = \lambda Y$ ถ้า $\lambda = 0$ แสดงว่า $AY = 0$

(1) หา Characteristic Root ของ $X'X$ คือ $\lambda_1, \lambda_2, \ldots, \lambda_k$

(2) หา Characteristic Vector ที่สอดคล้องกับ λ_i ที่มีค่าต่ำที่สุด แล้ว Normalized ด้วยการหารสมาชิกของ Characteristic Vector ดังกล่าวด้วยความยาวของเวคเตอร์ ได้เวคเตอร์ p[1]

(3) กำหนดค่า c ตามความเหมาะสม

(4) คูณ c เข้ากับสมาชิกทุกตัวของ p เวคเตอร์ cp คือค่าสังเกตชุดที่ (n+1) ตามต้องการ

สำหรับกรณีที่มี Characteristic Root หลายค่ามีค่าต่ำเราสามารถพิสูจน์ได้ในทำนองเดียวกัน และสำหรับปัญหาของการประมาณค่า Linear Combination ของพารามิเตอร์ β คือ $w'\beta$ ซิลวี (Silvey, S.D., *1969*)[2] สามารถพิสูจน์ให้เห็นได้ว่า ถ้าเลือกค่าสังเกต z_{n+1} ให้เป็นพหุคูณของเวคเตอร์ V เมื่อ $V = (I_k + \frac{1}{d^2}X'X)^{-1}w$ โดยที่ d^2 คือความยาวของเวคเตอร์ z_{n+1} แล้วค่าประมาณของ $w'\beta$ จะมีความถูกต้องแม่นยำขึ้น

8.5.2 Ridge Regression

8.5.2.1 หลักการและคุณสมบัติ

Ridge Regression ได้รับการพัฒนาขึ้นมาเพื่อแก้ปัญหา Multicolinearity โดยโฮเอร์ลและเคนนาร์ด (Hoerl, A.E. and Kennard, R.W., 1970)[3] และได้มีผู้อื่นร่วมกันพัฒนาอีกหลายท่าน[4]

หลักการโดยทั่วไปของ Ridge Regression ก็คือเมื่อพบว่า $|X'X| \to 0$ ซึ่งจะมีผลให้ $(X'X)^{-1}$ มีแนวโน้มที่จะไม่ปรากฏค่า และค่าประมาณของ β คือ $\hat{\beta} = (X'X)^{-1}X'Y$ มีค่าสูงผิดความจริง ให้พยายามปรับรูปแมตริกซ์ $X'X$ โดยการผนวกค่าคงที่ใด ๆ เข้ากับสมาชิกใน Main Diagonal ของ $X'X$ ซึ่งจะช่วยให้ Characteristic Root มีค่าสูงขึ้นและหลุดพ้นจากปัญหา Multicolinearity[5]

---

[1] อ่าน **มนตรี พิริยะกุล, คณิตศาสตร์สำหรับนักสถิติ (พีชคณิตเชิงเส้น)** หน้า 240-248, 311-321

[2] Silvey, S.D., "Multicolinearity and Imprecise Estimation," (Jour. of The royal Statistical Society, Series B., 35) p.67-75

[3] **Hoerl,** A.E. and Kennard, R.W., "Ridge Regression : Biased Estimation of Nonorthogonal Problems", (Technometrics, 12) p.55-67, 69-82

[4] Judge, G.G., et. al. Op. Cit., p.471-486

[5] อย่าลืมว่า $|X'X| = \lambda_1\lambda_2\ldots\lambda_k$ ถ้า λ บางค่ามีค่าต่ำ $|X'X|$ จะมีค่าใกล้ 0 ถ้าผนวกค่าคงที่ c ใด ๆ ให้แก่สมาชิกใน Main Diagonal ของ $X'X$ ทำให้แมตริกซ์ $X'X$ เปลี่ยนรูปเป็น $(X'X+cI_k)$ สมาชิกในแนว Main diagonal ของ $(X'X + cI_k)$ จะใหญ่กว่าของเดิมใน $X'X$ ซึ่งมีผลให้ Characteristic root ของ $(X'X + cI_k)$ มีค่าสูงขึ้นกว่าเดิม และ $|X'X + cI_k|$ จะมีค่าสูงขึ้น

Ridge Regression นอกจากพัฒนามาเพื่อแก้ปัญหา Multicolinearity แล้ว ยังสามารถแก้ไขจุดอ่อนบางประการของ OLS อีกด้วย กล่าวคือ โดยปกติแล้วเวคเตอร์ β คือ $\beta = \begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_k\end{bmatrix}$ จะมีความยาวคงที่[1] ค่าประมาณที่ดีของเวคเตอร์ β นั้นนอกจากจะประมาณค่าสมาชิกของเวคเตอร์ β ได้แม่นยำ (คือ $\hat{V}(\hat{\beta}_j) \to 0$) แล้วเวคเตอร์ของค่าประมาณคือ $\hat{\beta} = \begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{bmatrix}$ ควรมีความยาวโดยถัวเฉลี่ย (Expected Length) เท่ากับความยาวของ β

แต่จากการศึกษาความยาวถัวเฉลี่ย (Expected Length) ของ OLS-Extimator ของ β คือ $\hat{\beta}$ กลับพบว่า $\hat{\beta}$ มีความยาวถัวเฉลี่ยสูงกว่า β กล่าวคือ

$$
\begin{aligned}
E(\hat{\beta}'\hat{\beta}) &= E[(X'X)^{-1}X'Y]'[(X'X)^{-1}X'Y] \\
&= E[(X'X)^{-1}X'(X\beta+U)]'[(X'X)^{-1}X'(X\beta+U)] \\
&= E[\beta+(X'X)^{-1}X'U]'[\beta+(X'X)^{-1}X'U] \\
&= E[\beta'\beta+\beta'(X'X)^{-1}X'U+U'X(X'X)^{-1}\beta+U'X(X'X)^{-1}(X'X)^{-1}X'U] \\
&= \beta'\beta+0+0+E[U'\{X(X'X)^{-1}(X'X)^{-1}X'\}U] \\
&= \beta'\beta+\sigma^2 tr[X(X'X)^{-1}(X'X)^{-1}X'] \\
&= \beta'\beta+\sigma^2 tr(X'X)^{-1}
\end{aligned}
$$

แต่ $P'(X'X)P = D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ และ
$[P'(X'X)P]^{-1} = P'(X'X)^{-1}P = D^{-1} = \text{diag}(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_k})$ แสดงว่า $(X'X)^{-1}$ Similar กับ D^{-1} (P เป็น Orthogonal Modal Matrix)
ดังนั้น $tr\ P'(X'X)^{-1}P = tr(X'X)^{-1}PP' = tr(X'X)^{-1}I_k = tr(X'X)^{-1}$
แต่ $P'(X'X)^{-1}P = D$ แสดงว่า $tr(X'X)^{-1} = tr\ D^{-1} = \frac{1}{\lambda_1}+\frac{1}{\lambda_2}+\ldots+\frac{1}{\lambda_k}$

---

[1] ในที่นี้หมายถึง Square Length กล่าวคือเรานิยามความยาวของเวคเตอร์ V ได้ดังนี้ ความยาว (Length หรือ Norm) ของเวคเตอร์ $V = \sqrt{V'V} = \sqrt{\Sigma v_i^2}$ เมื่อ $V = (v_1, v_2, \ldots, v_n)$ ด้วยเหตุนี้จึงกล่าวถึงเฉพาะ $V'V$ ซึ่งหมายถึงกำลังสองของความยาว (Square Length)

นั่นคือ $E(\hat{\beta}'\hat{\beta}) = \beta'\beta + (\frac{\sigma^2}{\lambda_1} + \frac{\sigma^2}{\lambda_2} + \ldots + \frac{\sigma^2}{\lambda_k})$

$> \beta'\beta + \frac{\sigma^2}{\lambda_{min.}}$ เมื่อ $\lambda_{min.}$ คือค่าต่ำสุดของ λ_i; $i = 1,2,\ldots,k$

แสดงว่าเวคเตอร์ $\hat{\beta}$ มีความยาวโดยถัวเฉลี่ยสูงกว่าความยาวของเวคเตอร์ β อยู่แล้วเป็นปกติ

หมายเหตุ ในเชิงวิเคราะห์เราพิจารณาเฉพาะ λ_{min} ก็เพียงพอที่จะทำให้มองเห็นข้อเท็จจริงได้เพราะ λ_{min} ยิ่งมีค่าต่ำเพียงใด $\hat{\beta}$ ก็จะยิ่งมีความยาวถัวเฉลี่ยสูงกว่าความยาวของ β มากเพียงนั้น

ด้วยเหตุผล 2 ประการข้างต้นคือ (1) เพื่อแก้ปัญหา Multicolinearity และ (2) เพื่อให้ $\hat{\beta}$ มีความยาวโดยถัวเฉลี่ย (ความยาวคาดหมาย) ใกล้เคียงกับความยาวจริงของ β เราจึงควรแก้ปัญหาต่าง ๆ เหล่านี้ด้วยการทำให้ λ_{min} มีค่าสูงขึ้น ซึ่งกระทำได้ด้วยการเพิ่มขนาดสมาชิกในแนว Main Diagonal ของ $X'X$ ดังที่กล่าวมาแล้ว

ให้ $\hat{\beta}^*(c)$ คือ Ridge Regression Estimator ของ β จากสมการนอร์มอล

$$(X'X + cI_k)\beta = X'Y$$

โดยที่ c เป็นตัวคงที่ใด ๆ ที่ไม่ติดลบ คือ $0 \leq c \leq 1$ ดังนั้น[1]

$$\hat{\beta}^*(c) = (X'X + cI_k)^{-1}X'Y$$

จากการศึกษาคุณสมบัติทางทฤษฎีของ $\hat{\beta}^*(c)$ พบว่า $\hat{\beta}^*(c)$ มีคุณสมบัติดังนี้

1. $\hat{\beta}^*(c)$ เป็นเวคเตอร์ที่มีความยาวน้อยกว่าเวคเตอร์อื่น ๆ
2. SS (Residual) จะมีค่าสูงขึ้นตามค่าของ c
3. เวคเตอร์ $\hat{\beta}^*(c)$ มีความยาวน้อยกว่าเวคเตอร์ $\hat{\beta}$ เมื่อ $c \neq 0$ และเมื่อ c มีค่าสูงขึ้นเวคเตอร์ $\hat{\beta}^*(c)$ จะลดความยาวลง และเมื่อ $c \to \infty$ แล้ว $\hat{\beta}^*(c)'\hat{\beta}^*(c) \to 0$
4. เราสามารถหาค่าของ c; $c>0$ ที่มีผลให้ $MSE(\hat{\beta}^*(c)) < MSE(\hat{\beta})$ ได้เสมอ
5. Ridge Estimator $\hat{\beta}^*(c) = [I_k + c(X'X)^{-1}]^{-1}\hat{\beta} = Z\hat{\beta}$ เป็น Linear Transformation ของ OLS - Estimator กล่าวคือ

---

[1] ในกรณี Standardized Variable จะพบว่า $X'X$ ก็คือ Correlation Matrix R_x ซึ่งสมาชิกใน Main Diagonal ของ R_x จะมีค่าระหว่าง 0 ถึง 1 เราจึงปรับค่าแมตริกซ์ $X'X$ ที่มีปัญหาด้วยการผนวกค่าคงที่ c; $0 \leq c \leq 1$

$$\hat{\beta}^*(c) = (X'X + cI_k)^{-1}X'Y$$

$$= (X'X + cI_k)^{-1}[(X'X)(X'X)^{-1}]X'Y \quad {}^{1}$$

$$= (X'X + cI_k)^{-1}(X'X)[(X'X)^{-1}X'Y]$$

$$= [(X'X)^{-1}(X'X + cI_k)]^{-1}[(X'X)^{-1}X'Y] \quad {}^{2}$$

$$= I_k + c(X'X)^{-1}]^{-1}\hat{\beta}$$

หรือกล่าวนัยอีกนัยหนึ่งก็คือ $\hat{\beta}^*(c)$ สัมพันธ์กับ $\hat{\beta}$ ในลักษณะ $\hat{\beta}^*(c) = [I_k + c(X'X)^{-1}]^{-1}\hat{\beta}$ และขอให้สังเกตว่าถ้า $c = 0$ จะพบว่า $\hat{\beta}^*(c) = \hat{\beta}$ หรือ OLS เป็นกรณีเฉพาะของ Ridge Regression เมื่อ $c = 0$

6. $\hat{\beta}^*(c) = (X'X+cI_k)^{-1}X'Y = [I_k+c(X'X)^{-1}]^{-1}\hat{\beta}$ เป็น Biased Estimator ของ β และ Covariance Matrix ของเวคเตอร์ $\hat{\beta}^*(c)$ คือ

$$\mathrm{Cov}(\hat{\beta}^*(c)) = \sigma^2[X'X + cI_k]^{-1}(X'X)[X'X + cI_k]^{-1} \quad {}^{3}$$

พิสูจน์ ก. $E(\hat{\beta}^*(c)) = [I_k + c(X'X)^{-1}]^{-1}E(\hat{\beta})$

แต่ $E(\hat{\beta}) = \beta$ และเมื่อกำหนดให้ $Z = [I_k + c(X'X)^{-1}]^{-1}$ จึงพบว่า

$$E(\hat{\beta}^*(c)) = Z\beta \neq \beta$$

แสดงว่า $\hat{\beta}^*(c)$ เป็น Biased Estimator ของ β

ข. $$\mathrm{Cov.}(\hat{\beta}^*(c)) = E[\hat{\beta}^*(c) - E(\hat{\beta}^*(c))][\hat{\beta}^*(c) - E(\hat{\beta}^*(c))]'$$

$$= E[\hat{\beta}^*(c) - Z\beta][\hat{\beta}^*(c) - Z\beta]'$$

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

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

---

[1] $(X'X)^{-1}(X'X) = I_k$

[2] $(AB)^{-1} = B^{-1}A^{-1}$ ในที่นี้ $A = (X'X)^{-1}$ และ $B = (X'X + cI_k)$

[3] Marquardt, D.E. and Snee, R.D., "Ridge Regression", (The Univ. of Kentucky Conference on Regression with a Large Number of Predictor Variable, Lexington, Kentucky, Oct. 11-12, 1973), p.76

แต่ $\hat{\beta}-\beta = [\beta+(X'X)^{-1}X'U]-\beta = (X'X)^{-1}X'U$

$$\therefore \text{Cov}\ (\hat{\beta}^*(c)) = E[Z(X'X)^{-1}X'UU'X(X'X)^{-1}Z']$$

$$= Z(X'X)^{-1}X'E(UU')X(X'X)^{-1}Z'$$

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

แต่ $Z = [I_k+c(X'X)^{-1}]^{-1} = [I_k+c(X'X)^{-1}]^{-1}(X'X)^{-1}(X'X)$

$$= \{(X'X)[I_k+c(X'X)^{-1}]\}^{-1}\ (X'X)$$

$$= [X'X+cI]^{-1}\ (X'X)$$

$$\therefore \text{Cov}\ (\hat{\beta}^*(c)) = \sigma^2 Z(X'X)^{-1}Z'$$

$$= \sigma^2[X'X+cI]^{-1}(X'X)(X'X)^{-1}(X'X)[X'X+cI]^{-1}$$

$$= \sigma^2[X'X+cI]^{-1}(X'X)[X'X+cI]^{-1}$$

7. $$\text{MSE}(\hat{\beta}^*(c)) = \text{tr.Cov}\ (\hat{\beta}^*(c))+\beta'[Z-I_k]'[Z-I_k]\beta$$

$$= \text{Variance} + (\text{Bias})^2$$

$$= \sigma^2 \sum_{i}^{k} \frac{\lambda_i}{(\lambda_i+c_i)} + c^2\beta'[X'X+cI]^{-2}\beta$$ [1]

พิสูจน์ $\text{MSE}(\hat{\beta}^*(c)) = E[(\hat{\beta}^*(c))-\beta]'[(\hat{\beta}^*(c))-\beta]$ [2]

แต่ $\hat{\beta}^*(c) = [I_k+ c(X'X)^{-1}]^{-1}\hat{\beta} = Z\hat{\beta}$

$$\therefore \text{MSE}(\hat{\beta}^*(c)) = E[Z\hat{\beta}-\beta]'[Z\hat{\beta}-\beta]$$

พิจารณา $Z\hat{\beta} - \beta$ จะพบว่า

---

[1] อ่าน Hoerl, A.E. and Kennard, R.W., **Op. Cit.**, p.60

[2] จากข้อ 6 เราทราบว่า $\hat{\beta}^*(c)$ เป็น Biased Estimator ดังนั้นดัชนีที่ใช้วัด Dispression จึงต้องใช้ MSE แทน Variance โดยปกติ MSE จะเป็น Scalar และต้องวัดในรูปของ Deviation from true parameter มิใช่ Deviation from expected Value ของ estimator ขอให้เปรียบเทียบกับข้อ 6

$$Z\hat{\beta} - \beta = Z\hat{\beta}-Z\beta+Z\beta-\beta \qquad \text{บวกเข้าลบออกด้วย } Z\beta$$

$$= Z(\hat{\beta}-\beta)+(Z-I_k)\beta$$

$$[Z\hat{\beta}-\beta]' = (\hat{\beta}-\beta)'Z' + \beta'(Z-I_k)'$$

ดังนั้น $MSE(\hat{\beta}^*(c)) = E[(\hat{\beta}-\beta)'Z'+\beta'(Z-I_k)'][Z(\hat{\beta}-\beta)+(Z-I_k)\beta]$

$$= E[(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta)+(\hat{\beta}-\beta)'Z'(Z-I_k)\beta+\beta'(Z-I_k)'Z(\hat{\beta}-\beta)$$

$$+\beta'(Z-I_k)'(Z-I_k)\beta]$$

แต่เนื่องจาก $E(\hat{\beta}-\beta) = 0$ หรือ $E(\hat{\beta}) = \beta$

ดังนั้น $MSE(\hat{\beta}^*(c)) = E[(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta)]+\beta'(Z-I_k)'(Z-I_k)\beta$

$$= a + b$$

พิจารณา $a = E[(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta)]$ จะพบว่า $\hat{\beta}-\beta = \beta+(X'X)^{-1}X'U-\beta$

$$a = E[(\hat{\beta}-\beta)'Z'Z(\hat{\beta}-\beta)] = E[U'X(X'X)^{-1}Z'Z(X\ X)^{-1}X'U]$$

$$= E[U'\{X(X'X)^{-1}Z'Z(X'X)^{-1}X'\}U]$$

$$= \sigma^2 tr\{X(X'X)^{-1}Z'Z(X'X)^{-1}X'\}$$

$$= \sigma^2 tr\{(X'X)^{-1}Z'Z\} \text{ [1]}$$

แต่ $Z = [I_k+c(X'X)^{-1}]^{-1}$ เป็น Symmetric Matrix ขนาด $k \times k$ คือ $Z' = Z$

$$(X'X)^{-1}Z'Z = (X'X)^{-1}ZZ = (X'X)^{-1}[I_k+c(X'X)^{-1}]^{-1}[I_k+c(X'X)^{-1}]^{-1}$$

$$= [\{I_k+c(X'X)^{-1}\}(X'X)]^{-1}[I_k+c(X'X)^{-1}]^{-1}$$

$$= (X'X+cI)^{-1}[(X'X)(X'X)^{-1}\{I_k+c(X'X)^{-1}\}^{-1}]$$

---

[1] $tr(ABCDE) = tr(BCDEA)$ ดังนั้น $tr\ \{X(X'X)^{-1}Z'Z(X'X)^{-1}X'\}$

$= tr\ \{(X'X)^{-1}Z'Z(X'X)^{-1}X'X\} \quad = tr\ \{(X'X)^{-1}Z'Z\}$

$$= (X'X+cI)^{-1}(X'X)[\{I_k+c(X'X)^{-1}\}(X'X)]^{-1}$$

$$= (X'X+cI)^{-1}(X'X)(X'X+cI)^{-1}$$

นั่นคือ $a = \sigma^2 tr[(X'X+cI)^{-1}(X'X)(X'X+cI)^{-1}]$

$$MSE\ (\hat{\beta}^*(c)) = \sigma^2 tr[(X'X+cI)^{-1}(X'X)(X'X+cI)^{-1}]+\beta'(Z-I_k)'(Z-I_k)\beta$$

$$= tr\ Cov\ (\hat{\beta}^*(c))+\beta'(Z-I_k)'(Z-I_k)\beta$$

$$= tr\ Cov\ (\hat{\beta}^*(c))+(Bias)^2$$

และโดยอาศัยเทคนิคในทางพีชคณิตเชิงเส้นและความรู้เรื่อง Characteristic Value เราสามารถพิสูจน์ได้ว่า

$$MSE\ (\hat{\beta}^*(c)) = \sigma^2 \sum_{i}^{n} \frac{\lambda_i}{(\lambda_i+c_i)^2} + c^2\beta'(X'X+I_k)^{-2}\beta$$

ข้อสังเกต จาก MSE $(\hat{\beta}^*(c))$ จะพบว่า Variance Term จะมีค่าลดลงขณะที่ c มีค่าสูงขึ้น แต่ Bias Term จะมีค่าสูงขึ้นเมื่อ c มีค่าสูงขึ้น

8. SS (Residual) $= [Y-X\hat{\beta}^*(c)]'[Y-\hat{\beta}^*(c)] = Y'Y-\hat{\beta}^{*\prime}(c)'X'Y-k\hat{\beta}^{*\prime}(c)'\hat{\beta}^*(c)$

พิสูจน์

$$SSE = [Y-X\hat{\beta}^*(c)]'[Y-X\hat{\beta}^*(c)]$$

$$= Y'Y-Y'X\hat{\beta}^*(c)-\hat{\beta}^*(c)'X'Y+\hat{\beta}^*(c)'X'X\hat{\beta}^*(c)$$

$$= Y'Y-\hat{\beta}^*(c)'X'Y-[Y'X\hat{\beta}^*(c)-\hat{\beta}^*(c)'X'X\hat{\beta}^*(c)]$$

$$= Y'Y-\hat{\beta}^*(c)'X'Y-[Y'X-\hat{\beta}^*(c)'X'X]\hat{\beta}^*(c)$$

$$= Y'Y-\hat{\beta}^*(c)'X'Y-[X'Y-X'X\hat{\beta}^*(c)]'\hat{\beta}^*(c)$$

แต่ $(X'X+kI)\hat{\beta}^*(c) = X'Y$ คือ Normal Equation ของ Ridge Regression

ดังนั้น $SSE = Y'Y-\hat{\beta}^*(c)'X'Y-[(X'X+cI)\hat{\beta}^*(c)-X'X\hat{\beta}^*(c)]'\hat{\beta}^*(c)$

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

ข้อสังเกต ถ้า $c = 0$ จะมีผลให้ $\hat{\beta}^*(c) = \hat{\beta}$ และ $SSE = Y'Y - \hat{\beta}'X'Y$

8.5.2.2 การประมาณค่า c

จากตอนที่ 8.5.2.1 เราได้ศึกษาพบว่าถ้าเกิดปัญหา Multicolinearity ขึ้นแล้ววิธีแก้ปัญหานี้ก็คือให้นำตัวคงที่ c ใด ๆ ไปผนวกกับสมาชิกใน Main Diagonal ของ $X'X$ เพื่อจะได้ทำให้ λ_{min} มีค่าสูงขึ้น ปัญหาก็คือค่าที่เหมาะสมของ c คืออะไร หรือ c = ?

การประมาณค่า c สามารถกระทำได้หลายวิธีทั้งโดยวิธี Ridge Trace วิธี Iterate ตลอดจนวิธีของเบยส์ ในที่นี้จะเสนอวิธีต่าง ๆ ไว้ 3 วิธีพอเป็นแนวทางดังนี้

วิธีที่ 1 Ridge Trace

วิธีนี้เสนอโดยโฮเอร์ล และเคนนาร์ด (Hoerl, A.E. and Kennard, R.W., 1970)[1] หลักการโดยสรุปก็คือ

(1) จาก $\hat{\beta}^*(c) = (X'X+cI)^{-1}X'Y$ ให้แปรค่าของ c จาก 0 ไปถึง 1 โดยที่การแปรค่าครั้งหนึ่งให้คำนวณหา $\hat{\beta}^*(c)$ ครั้งหนึ่ง และหาค่า SSE = $Y'Y-\hat{\beta}^*(c)'X'Y-k\hat{\beta}^*(c)'\hat{\beta}^*(c)$ ครั้งหนึ่ง ซึ่งจะมีผลให้ $\hat{\beta}_j^*$; $j = 1,2,\ldots,k$ และ SSE มีค่าผันแปรไปตามค่าของตัวคงที่ c

(2) นำค่าของ $\hat{\beta}_j^*$; $j = 1,2,\ldots,k$ และ SSE ไปพล็อตในกระดาษกราฟ โดยให้แกนตั้ง คือค่าของ $\hat{\beta}_j^*$ และ SSE ส่วนแกนนอนคือแกนของค่าคงที่ c จะได้โค้ง (Trace) ทั้งสิ้น k โค้ง ($\hat{\beta}_j^*$ หนึ่งตัวต่อ 1 โค้ง) และอีกโค้งหนึ่งคือโค้งของ SSE

(3) วิเคราะห์จากกลุ่มของ Trace เพื่อวินิจฉัยเลือกค่า c ที่เหมาะสมดังนี้

ก. เลือกค่า c จากแนวของ Trace ที่ไม่เปลี่ยนแปลงมากนัก (System Stabilization)[2]

ข. เลือก c จากโค้งของ SSE ที่มีค่าต่ำ

ค. เลือก c จากเฉพาะกลุ่มของ Trace ที่ไม่มีเครื่องหมาย (+ หรือ -) สับไปสับมา กล่าวคือ Trace ของ $\hat{\beta}_j^*$ จะต้องอยู่เหนือแกน หรือใต้แกนเสมอ ถ้าเคยอยู่เหนือแกนแล้วกลับไปอยู่ใต้แกนแสดงว่า $\hat{\beta}_j^*$ มีเครื่องหมายเปลี่ยนแปลงไปตามค่า c ที่สูงขึ้น ซึ่งไม่ใช่สิ่งที่เราต้องการ

---

[1] Hoerl, A.E. and Kennard, R.W., **Ibid.**, p.58 และ Hoerl, A.E. and Kennard, R.W., "Ridge Regression : Application to Nonorthogonal Problems", (Technometrics, 12, Feb., 1970), p.69-82 ซึ่งเสนอเฉพาะ Ridge Trace.

[2] โฮเอร์ลและเคนนาร์ดแนะนำว่า เราใช้ ข้อ ก. นี้เพียงประการเดียวก็เพียงพอ

ให้ตัด $\hat{\beta}_j^*$ ทิ้งไปแล้วหา Ridge Trace ใหม่[1]

หมายเหตุ นักวิจัยควรเลือกค่า c โดยใช้เกณฑ์ทั้ง 3 ข้อ (ก, ข, ค) ผสมกันไม่ควรใช้ข้อใดข้อหนึ่ง

ตัวอย่างเช่น ข้อมูลจาก กอร์แมน และโทแมน (Gorman and Toman) ซึ่งประกอบด้วยตัวแปรอิสระ 10 ตัว ปรากฏ Correlation Matrix และ λ_i; $i=1,2,...,10$ ดังนี้[2]

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y
X_1	1										
X_2	-0.04	1									
X_3	0.51	-0.00	1								
X_4	0.12	-0.16	0.00	1							
X_5	-0.71	0.06	-0.59	-0.07	1						
X_6	-0.87	0.09	-0.65	-0.09	0.84	1					
X_7	-0.09	0.24	-0.02	0.03	0.38	0.13	1				
X_8	-0.00	0.01	0.34	0.08	-0.36	-0.20	-0.48	1			
X_9	-0.09	0.09	-0.08	0.02	-0.14	0.40	0.07	-0.18	1		
X_{10}	-0.36	-0.30	-0.44	-0.09	0.54	0.45	0.40	-0.46	0.05	1	
Y	-0.81	-0.10	-0.63	-0.10	0.56	0.81	0.04	0.06	0.46	0.45	1

$\lambda_1 = 3.692$ $\qquad \lambda_6 = 0.659$

$\lambda_2 = 1.542$ $\qquad \lambda_7 = 0.357$

$\lambda_3 = 1.293$ $\qquad \lambda_8 = 0.220$

$\lambda_4 = 1.046$ $\qquad \lambda_9 = 0.152$

$\lambda_5 = 0.972$ $\qquad \lambda_{10} = 0.068$

---

[1] โฮเอร์ลและเคนนาร์ดแนะนำให้คง X_j ไว้ตามเดิม ถ้าต้องการขจัดทิ้งให้กระทำได้โดยแทนค่า X_{ji} ด้วย $\bar{X}_j$ ซึ่งจะให้ผลให้ $\hat{\beta}_j = 0$ โดยมิได้ทำให้ต้องเปลี่ยนแปลงขนาดของแมตริกซ์ $(X'X+cI)$ แต่ประการใด

[2] อ่านรายละเอียดใน Hoerl, A.E. and Kennard, R.W. **Ibid.,** p.70-75

จะเห็นว่า $\lambda_{min} = 0.068 \neq 0$

ผลการทดลองคำนวณหาค่า $\hat{\beta}^*(c)$ และ SSE เมื่อแปรค่า c จาก 0 ถึง 1 ปรากฏ Trace ของ $\hat{\beta}_j^*$ ดังภาพ

จากภาพเราใช้แกนตั้งด้านซ้ายมือคือค่าของ $\hat{\beta}_j^*$; $j = 1,2,\ldots,10$ แกนตั้งด้านขวามือคือแกนของ SSE แกนนอนคือแกนของค่าคงที่ c (สัญญลักษณ์ที่ใช้เดิมใน Literature คือ k ผู้เขียนใช้อักษร c แทนเพราะเกรงว่านักศึกษาจะนำไปสับสนกับอักษร k ที่เคยใช้แสดงรหัสของตัวแปร) โดยที่ c มีค่าตั้งแต่ 0 ถึง 1 และ ณ. c = 0 ค่าของ $\hat{\beta}_j^*$ ในแกนตั้ง (ซ้าย) ก็คือ $\hat{\beta}_j$(OLS - Estimator ของβ_j)

จากภาพจะเห็นว่า เมื่อ c = 0 ขนาด (absolute magnitude) ของ $\hat{\beta}_j^*$ (ซึ่งก็คือ $\hat{\beta}_j$ นั่นเอง) มีค่าสูงมาก และเมื่อ c มีค่าสูงขึ้นขนาดของ $\hat{\beta}_j^*$ จะลดลงอย่างรวดเร็ว และเมื่อ c มีค่าระหว่าง 0.2 ถึง 0.3 ขนาดของ $\hat{\beta}_j^*$ จะค่อย ๆ ทรงตัว หรือมีอัตราการเปลี่ยนแปลงช้าลงซึ่งเมื่อ c มีค่าสูงขึ้น Trace ของ $\hat{\beta}_j^*$ ก็มิได้เปลี่ยนไปมากนัก ขณะเดียวกันเมื่อพิจารณาโค้งของ SSE ซึ่งแสดงด้วยเส้นประ จะเห็นว่า SSE ค่อย ๆ มีค่าสูงขึ้นเมื่อ c มีค่าสูงขึ้น และ ณ จุดที่ $0.2<c<0.3$ จะพบว่า SSE มิได้มีค่าสูงเกินไป

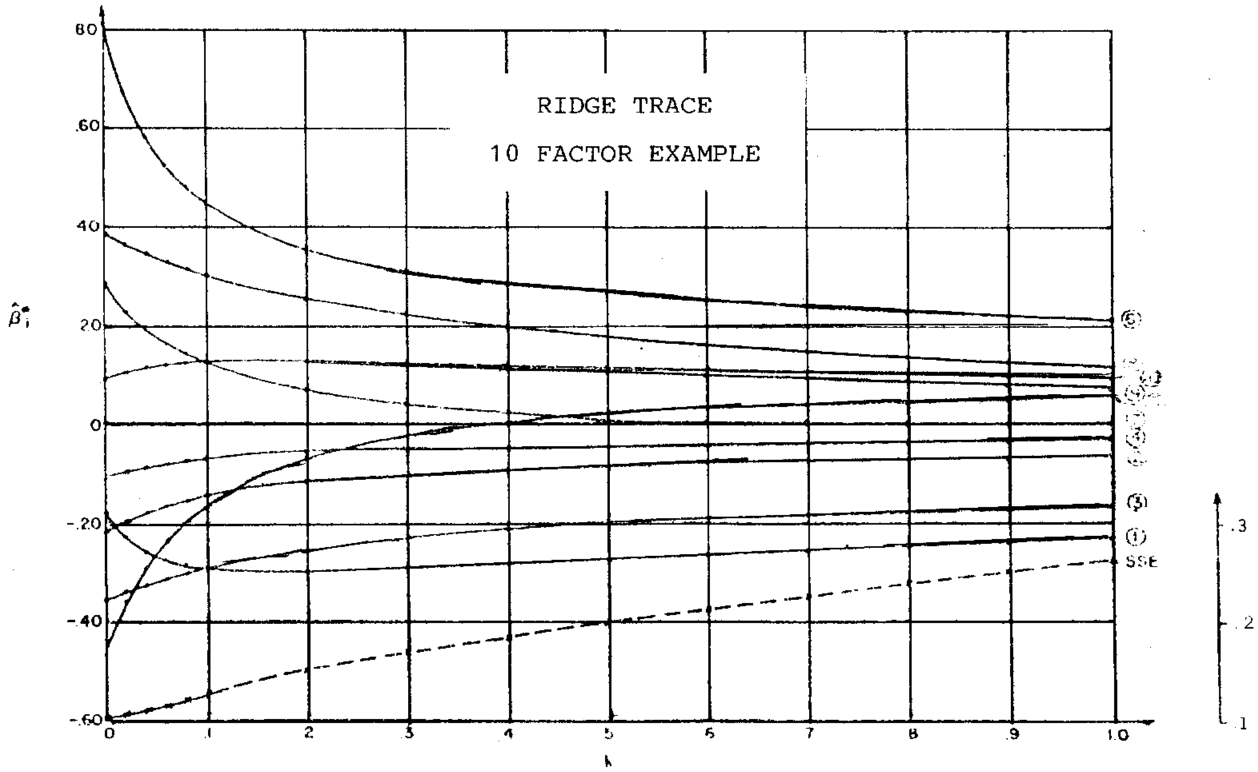
ด้วยเหตุผลดังกล่าว เมื่อกำหนดให้ c = 0.25 จะทำให้ $\hat{\beta}_j^*$ มีค่าคงที่ (System Stabilization) เราจึงใช้ c = 0.25

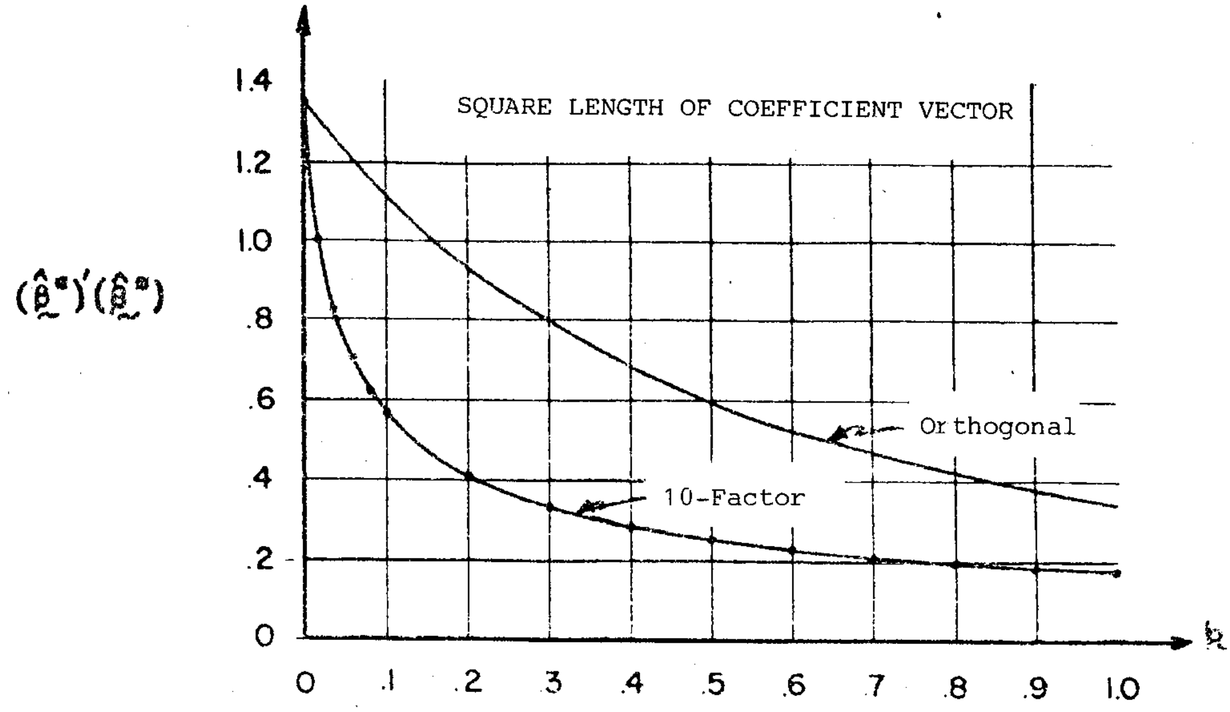
นอกจากนี้หากพิจารณา Trace ของ $\hat{\beta}_5^*$ และ $\hat{\beta}_7^*$ จะพบว่า $\hat{\beta}_5^*$ มีค่าเปลี่ยนจากลบเป็นบวก เมื่อ c มีค่าสูงขึ้น ขณะที่ $\hat{\beta}_7^*$ มีค่าลดลงสู่ 0 เมื่อ c มีค่าสูงขึ้น แสดงว่า X_5 และ X_7 มีอิทธิพลไม่แน่นอนสมควรขจัดทิ้ง และเมื่อทดลองตัด X_5 และ X_7 ทิ้งพบว่าเวคเตอร์ $\hat{\beta}^*(c)$ มีความยาวลดลงทั้งยังมีธรรมชาติคล้ายความยาวของ Orthogonal Coefficient Vector แต่เรื่องนี้ โฮเอร์ล และเคนนาร์ดแนะนำว่า เราไม่จำเป็นต้องคัดตัวแปรใดออก ตัวแปรใดที่ปรารถนาจะคัดออกสามารถกระทำได้โดยไม่ต้องลดขนาดของแมตริกซ์ $(X'X+cI)$ คือ แทนที่x_{ji} ด้วย $\bar{x}_j$ ซึ่งจะมีผลเสมือนการกำหนดให้ $\beta_j = 0$

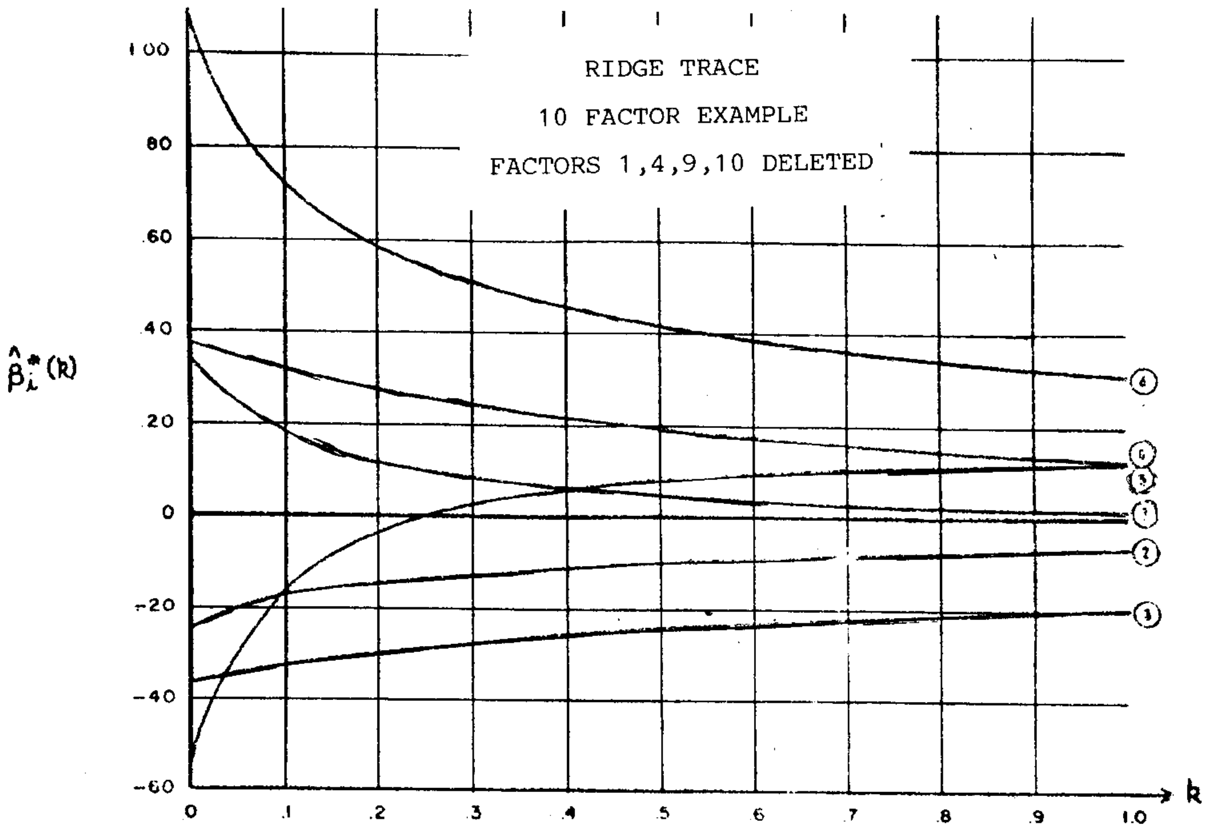
จาก Ridge Trace เมื่อกำหนดให้ c = 0.025 จะพบว่า

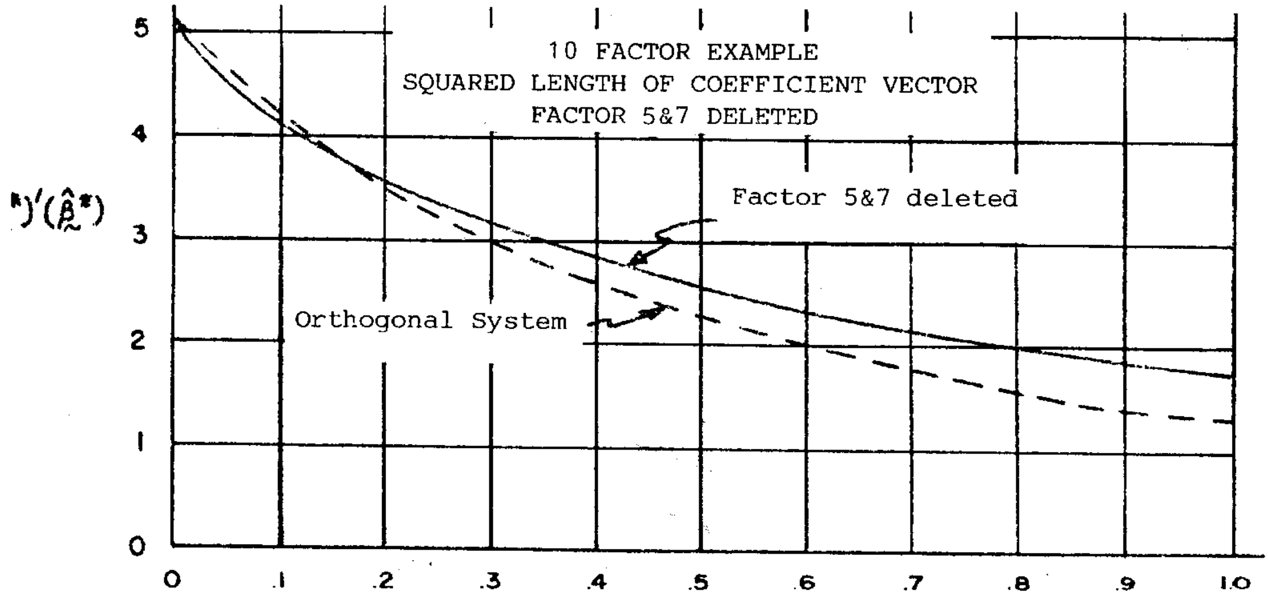
$\hat{\beta}_1^* = -0.295$		$\hat{\beta}_6^* = 0.325$
$\hat{\beta}_2^* = -0.110$		$\hat{\beta}_7^* = 0.050$
$\hat{\beta}_3^* = -0.245$		$\hat{\beta}_8^* = 0.240$
$\hat{\beta}_4^* = -0.050$		$\hat{\beta}_9^* = 0.125$
$\hat{\beta}_5^* = -0.040$		$\hat{\beta}_{10}^* = 0.125$

SSE = 0.16




SQUARE LENGTH OF COEFFICIENT VECTOR
Orthogonal
10-Factor
1.4
1.2
1.0
.8
.6
.4
.2
0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0


RIDGE TRACE
10 FACTOR EXAMPLE
FACTORS 1,4,9,10 DELETED
100
80
60
40
20
0
-20
-40
-60
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0



วิธีที่ 2 Hoerl - Kennard - Baldwin estimator[1]

$$c_{HKB} = \frac{k\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

โดยที่ k = จำนวนตัวแปรอิสระ, $\hat{\beta} = (X'X)^{-1}X'Y$ คือ OLS - Estimator ของ β และ

$$\hat{\sigma}^2 = \frac{1}{n-k}(Y'Y-\hat{\beta}'X'Y)$$

จากการศึกษาพบว่าวิธีนี้จะมีคุณภาพสูง (เป็น Minimax Estimator) ถ้าหาก

$k > 3$ และ $\lambda_{min} < \frac{(4+k)\lambda_{max}}{2k}$ และพบว่าถ้า $k < .2$ จะไม่ดีไปกว่า OLS เลย

วิธีที่ 3 McDonald - Galarneau Estimator[2]

แมคโดนาล์ด และ กาลาร์นู พิจารณาจากข้อเท็จจริงที่ว่า OLS - Estimator คือเวคเตอร์ $\hat{\beta}$ นั้น โดยปกติจะมีความยาวสูงกว่าความยาวของเวคเตอร์จริงคือ β อยู่ทั้งสิ้น $\sigma^2 tr(X'X)^{-1}$ กล่าวคือ

---

1 Judge, G.G. et. al., **Op. Cit.**, p.475-6

2 **Ibid.**, p.277

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

หรือ

$$E[\hat{\beta}'\hat{\beta} - \sigma^2 tr(X'X)^{-1}] = \beta'\beta$$

ด้วยเหตุนี้ถ้าเราจะใช้ Ridge Regression เราควรเลือกตัวคงที่ c ที่มีผลให้เวคเตอร์ $\hat{\beta}^*(c)$ มีความยาวเท่ากับความยาวจริง (โดยประมาณ) นั่นก็คือควรเลือกค่า c โดยวิธี Search จากสมการ

$$\hat{\beta}^*(c)'\hat{\beta}^*(c) = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 tr(X'X)^{-1}$$

เมื่อ $\hat{\sigma}^2 = \frac{1}{n-k}(Y'Y - \hat{\beta}'X'Y)$

ยังมีวิธีประมาณค่า c อีกมากมายหลายวิธี ผู้สนใจสามารถติดตามศึกษาได้จากเอกสารอ้างอิง และวารสารทางวิชาการโดยเฉพาะอย่างยิ่งคือ Technometrics, JASA และอื่น ๆ

8.5.3 Principal Component Regression (PCR)

PCR เป็นกรณีเฉพาะที่สำคัญของ Factor Analysis หลักการโดยสรุปของ PCR ก็คือ พยายามสร้างตัวแปรอิสระตัวใหม่ คือ Z's ที่เกิดจาก Linear Combination ของตัวแปรอิสระ X's ขึ้นมาในลักษณะที่ตัวแปร Z's เหล่านี้เป็นอิสระเชิงเส้นต่อกัน (เราไม่เรียกว่า Independent แต่เรียกว่า Artificial Independent) ซึ่งตัวแปร z_j แต่ละตัวจะดึงดูดเอาคุณลักษณะร่วมต่าง ๆ จาก X's มาไว้ในตัวของมันโดย z_1 จะดึงดูดเอาคุณลักษณะต่าง ๆ ของ X's ไว้มากที่สุด กาก (Residual) ที่เหลือจะถูก z_2 ดูดเอาไว้ กากที่เหลือจากที่ z_2 ดูดไปแล้วจะถูก z_3 ดูดเอาไว้ ดังนี้เรื่อยไป ซึ่งก็คงพอมองเห็นได้ว่าจำนวนตัวแปร Z อาจมีจำนวนได้มากที่สุดเท่ากับจำนวนตัวแปร X's แต่โดยทั่วไปแล้วจำนวนตัวแปร Z มักน้อยกว่าจำนวนตัวแปร X ทั้งนี้ขึ้นอยู่กับธรรมชาติของ Z ว่าดูดซับเอาคุณลักษณะร่วมของ X's ไว้ในตัวได้รวดเร็วเพียงใด จากนั้นจึงใช้ Z's เป็นตัวแปรอิสระสำหรับวิเคราะห์สมการถดถอย Y = f(Z's) ปัญหา Multicolinearity ย่อมไม่ปรากฏในสมการ Y = f(Z's) เพราะ Z's ถูกสร้างให้เป็นตัวแปรที่อิสระต่อกันอยู่แล้ว ปัญหาที่ปรากฏพบก็คือเราไม่ทราบว่า z_j แต่ละตัวหมายถึงอะไร เพราะ z_j เป็น Linear Combination ของ X's ทำให้ประสบปัญหาในการอภิปรายผล

8.5.3.1 หลักเกณฑ์และการพัฒนา

การพัฒนา PCR นั้นกระทำได้ 2 แนวคือ แนวที่อาศัย Characteristic Value แนวหนึ่ง และแนวที่อาศัย Correlation Matrix อีกแนวหนึ่ง นักศึกษาเลือกใช้วิธีใดก็ได้ เพราะให้ผลเสมอกัน จะต่างกันที่ความยากง่ายของการพัฒนา ซึ่งหากนักศึกษามีพื้นฐานความรู้เรื่องพีชคณิตเชิงเส้นดีพอ แนวทางทั้งสองล้วนเป็นเรื่องง่ายไม่ยากแก่การติดตาม ในที่นี้จะกล่าวถึงเฉพาะแนวที่อาศัย Characteristic Value เท่านั้น ส่วนแนวที่อาศัย Correlation Matrix ซึ่งความจริงก็คือกรณีที่พัฒนาต่อจากเดิมไปอีกเพียงเล็กน้อยนั้น นักศึกษาสามารถศึกษาได้จาก Koutsoyiannis, A., **Op. Cit.**, p.414-20 และ Johnston, J., **Op. Cit.**, p.325-6 และเพื่อให้แนวทางทั้งสองกลายเป็นแนวทางเดียวกัน (แต่ต่างกันเล็กน้อยในวิธีปฏิบัติ) เราจะพัฒนา PCR จากกรณีที่ตัวแปร X's และ Y ได้รับการแปลงรูปให้เป็น Standardized Variable ซึ่งในกรณีเช่นนี้แมตริกซ์ $X'X$ ก็คือ Correlation Matrix R_X นั่นเอง

จากสมการ $\frac{y}{s_1} = \beta_2\frac{x_2}{s_2} + \beta_3\frac{x_3}{s_3} + \ldots + \beta_k\frac{x_k}{s_k} + \frac{u-\bar{u}}{s_u}$ หรือเสนอในรูปแมตริกซ์ได้เป็น $Y = X\beta+U$ โดยที่ X เป็นแมตริกซ์ขนาด $n\times(k-1)$ β เป็นเวคเตอร์ขนาด $(k-1)\times 1$[1]

สมมุติว่าเราตรวจพบว่ามีปัญหา Multicolinearity ปรากฏขึ้น การพัฒนา PCR จะเริ่มจากการแปลงรูปสมการ $Y = X\beta+U$ ดังนี้

ให้ P คือ Orthogonal Modal Matrix ที่แต่ละสดมภ์ของ P คือ Normalized Characteristic Vector $p_2, p_3, \ldots, p_k$ ขนาด $(k-1)\times 1$ ที่สอดคล้องกับ Characteristic Value $\lambda_2, \lambda_3, \ldots, \lambda_k$ ของแมตริกซ์ $X'X$ กล่าวคือ

$$P = (p_2, p_3, \ldots, p_k) \quad \text{โดยที่} \quad p_i'p_j = \begin{cases} 1 & \text{ถ้า } i = j \\ 0 & \text{ถ้า } i \neq j \end{cases} \quad \text{หรือ} \; P'P = PP' = I$$

---

[1] ผู้เขียนใช้สมการ $Y_i = \beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\ldots+\beta_kX_{ki}+u_i$ ตามแบบเดิมที่เคยเสนอไว้ในบทที่ 3 ดังนั้นเมื่อ Standardize จึงมีผลให้สมการถดถอยไม่มีเทอมคงที่ β_1 ปรากฏอยู่ แมตริกซ์ X ซึ่งมีขนาด $n\times(k-1)$ แต่ตาม Literature นิยมใช้สมการรูปต่อไปนี้คือ $Y_i = \beta_0+\beta_1X_{2i}+\ldots+{}_{k-1}X_{ki}+u_i$ ซึ่งภายหลังจาก Standardize แมตริกซ์ X จะมีขนาด $n\times k$ อย่างไรก็ตามผลลัพธ์จากสมการ 2 รูปนี้จะตรงกันเสมอเพียงแต่ต้องระมัดระวังเรื่องค่าของ k อยู่บ้าง กล่าวคือตามรูปของผู้เขียน k หมายถึงจำนวนพารามิเตอร์ β k-1 หมายถึงจำนวน Regression Coefficient คือ $\beta_2, \beta_3, \ldots, \beta_k$ และจำนวน Regressor คือ $X_2, X_3, \ldots, X_k$ ขณะที่รูปที่ใช้ใน Literature k หมายถึงจำนวน Regression Coefficient และ Regressor คือ $\beta_1, \beta_2, \ldots, \beta_k$ และ $X_1, X_2, \ldots, X_k$

นำแมตริกซ์ P ไปแปลงรูปสมการ $Y = X\beta+U$ โดยอาศัยความจริงว่า $I = PP'$ จะพบว่า

$$\begin{aligned} Y &= X(PP')\beta+U \\ &= (XP)(P'\beta)+U \\ &= (XP)\theta+U && : \text{ ให้ } P'\beta = \theta \\ &= Z\theta + U && : \text{ ให้ } XP = Z \text{ เรียกว่า Principal Component Matrix} \end{aligned}$$

การประมาณค่าพารามิเตอร์ θ จากสมการ $Y = Z\theta + U$ สามารถกระทำได้โดยอาศัย OLS ตามปกติคือ $\hat{\theta} = (Z'Z)^{-1}Z'Y$ แต่เนื่องจาก $\theta = P'\beta$ ค่าประมาณของ θ จึงเป็น Linear Function ของ β ดังนั้นเพื่อให้ได้ค่าประมาณของ β เราจึงต้องปรับรูปคืนสู่ β ตามเดิมคือ

จาก $\hat{\theta} = P'\beta$ โดยที่ P เป็น Orthogonal Matrix $(P'P = PP' = I)$

$\therefore \hat{\beta}_{PC} = P\hat{\theta}$

อย่างไรก็ตามในทางปฏิบัตินั้นเราไม่จำเป็นต้องนำตัวแปร Z ทุกตัวคือ $Z_2, Z_3, \ldots, Z_k$ เข้าร่วมการวิเคราะห์ทั้งหมดเพราะ Z_j บางตัวอาจ**มิได้มีความสำคัญ** คือมิได้ดึงดูดเอาอิทธิพลร่วมของ X's ไว้ในตัวเลย เราจึงจำเป็นต้องตัดตัวแปร Z บางตัวทิ้งไปก่อนการวิเคราะห์ สมมุติว่ามี $Z_2, Z_3, \ldots, Z_r$ เท่านั้นที่มีความสำคัญเราก็ตัด $Z_{r+1}, Z_{r+2}, \ldots, Z_k$ ทิ้งไป ด้วยเหตุนี้การวิเคราะห์สมการถดถอย $Y = Z\theta + U$ จึงต้องอาศัยวิธี Partition ดังนี้คือ

$$Z = [Z_1 \mid Z_2]\begin{bmatrix}\theta_1 \\ \hline \theta_2\end{bmatrix} + U = Z_1\theta_1 + Z_2\theta_2 + U$$

โดยที่

$$Z = [Z_2, Z_3, \ldots, Z_r \mid Z_{r+1}, Z_{r+2}, \ldots, Z_k] = [Z_1 \mid Z_2]$$

และ

$$\theta = \begin{bmatrix}\theta_2 \\ \theta_3 \\ \vdots \\ \theta_r \\ \hline \theta_{r+1} \\ \vdots \\ \theta_k\end{bmatrix} = \begin{bmatrix}\theta_1 \\ \hline \theta_2\end{bmatrix}$$

แล้วให้ถือเสมือนว่า $\theta_2 = 0$ คือ $\theta_{r+1} = 0,\ \theta_{r+2} = 0, \ldots, \theta_k = 0$ ดังนั้นสมการถดถอยที่ต้องวิเคราะห์จึงลดรูปลงเหลือเพียง $Y = [Z_1 \vdots Z_2] \begin{bmatrix} \theta_1 \\ \hdashline 0 \end{bmatrix} + U$ นั่นคือ

$$Y = Z_1\theta_1 + U \text{ และ } \hat{\theta}_1 = (Z_1'Z_1)^{-1}Z_1'Y$$

ดังนั้นเมื่อย้อน $\hat{\theta}_1$ สู่ $\hat{\beta}$ จึงพบว่า $\hat{\beta}_{PC} = P_1\hat{\theta}_1$ โดยที่ $P_1 = (p_2, p_3, \ldots, p_r)$

หมายเหตุ ถ้าใช้ตัวแปร Z ทุกตัววิเคราะห์สมการถดถอยจะพบว่า $\hat{\beta}_{PC} = \hat{\beta}_{OLS}$ ซึ่งจะไม่ช่วยแก้ปัญหา Multicolinearity ได้แต่ประการใด

8.5.3.2 ลำดับความสำคัญตัวแปร Z และการทดสอบความสำคัญ

พิจารณาสมการ $Y = (XP)\theta + U$ จะพบว่าเมื่อกำหนดให้ Z = XP ก็แสดงว่า Principal Component Matrix Z เกิดขึ้นจาก Orthogonal Transformation ของแมตริกซ์ X หรือ $Z = XP = X(p_2, p_3, \ldots, p_k) = (Xp_2, Xp_3, \ldots, Xp_k) = (Z_2, Z_3, \ldots, Z_k)$ เราเรียก Z_j ว่า Principal Component Variabvle ตัวที่ j

พิจารณา Z_j จะพบว่า $Z_j = Xp_j$ โดยที่ p_j คือ Normalized Characteristic Vector ที่สอดคล้องกับ Characteristic Root λ_j ของแมตริกซ์ $X'X$ แสดงว่า Z_j คือ Linear Combination ของเวคเตอร์ X's กล่าวคือ

$$Z_j = Xp_j = [X_2, X_3, \ldots, X_k] \begin{bmatrix} p_{j2} \\ p_{j3} \\ \vdots \\ p_{jk} \end{bmatrix}$$

$$= p_{j2}X_2 + p_{j3}X_3 + \ldots + p_{jk}X_k$$

แสดงว่า Z_j เกิดจากการประกอบกันของเวคเตอร์ $X_2, X_3, \ldots, X_k$ และ Z_j จะมีความสำคัญมากน้อยเพียงใดขึ้นอยู่กับธรรมชาติของ Z_j เองว่าสามารถดึงดูดเอาคุณลักษณะต่าง ๆ ของ

$X_2, X_3, \ldots, X_k$ ไว้ในตัวได้มากน้อยเพียงใด Z_j ตัวใดดึงดูดเอาอิทธิพลร่วมของ X's ไว้ได้มากที่สุด ถือว่ามีความสำคัญที่สุด Z_j ตัวใดดึงดูดอิทธิพลร่วมของ X's ไว้ได้น้อยถือว่ามีความสำคัญน้อยลดหลั่นกันไปจนถึง Z_j ตัวท้าย ๆ ที่แทบจะมิได้ดึงดูดเอาอิทธิพลร่วมของ X's ไว้ได้เลย (เพราะ Z_j ตัวก่อนหน้านั้นดึงดูดอิทธิพลร่วมของ X's ไว้หมดหรือเกือบหมด) ก็จะไม่มีความสำคัญเลย และสามารถตัดทิ้งไปจากกระบวนการวิเคราะห์สมการถดถอย $Y = Z\theta + U$ ได้

พิจารณาตัวแปร $Z_j = Xp_j$ จะพบว่า

$$
\begin{aligned}
Z_3'Z_3 &= (Xp_j)'(Xp_j) \\
&= p_j'(X'X)p_j^{\,1} \\
&= \lambda_j
\end{aligned}
$$

และ

$$
\begin{aligned}
Z'Z &= (XP)'(XP) \\
&= P'(X'X)P \\
&= D \\
&= \text{diag}\,(\lambda_2, \lambda_3, \ldots, \lambda_k) \\
&= \begin{bmatrix} \lambda_2 & & & & \\ & \lambda_3 & & & \\ & & \ddots & & \\ & & & \lambda_j & \\ & & & & \ddots \\ & & & & & \lambda_k \end{bmatrix}
\end{aligned}
$$

---

[1] จาก Characteristic Equation $AY = \lambda Y$ เราจะพบว่า $Y'AY = Y'\lambda Y = \lambda Y'Y = \lambda$ ถ้า Y เป็น Orthonormal Vector ดังนั้นจาก $p_j'(X'X)p_j$ เราทราบว่า $(X'X)p_j = \lambda_j p_j$ เพราะฉะนั้น $p_j'(X'X)p_j = \lambda_j p_j'p_j = \lambda_j$

เมื่อเรียงลำดับค่าของ Characteristic root λ ตามขนาดจากมากไปหาน้อยในลักษณะของ Order Statistic สมมุติว่าได้ดังนี้

$$\lambda_2 > \lambda_3 > \lambda_4 > \ldots > \lambda_k$$ [1]

เราจึงให้ $Z_2 = Xp_2$ เป็น PC ตัวที่ 1 $Z_3 = Xp_3$ เป็น PC ตัวที่ 2 ... $Z_k = Xp_k$ เป็น PC ตัวที่ k-1[2]

เครื่องมือที่ใช้วัดความสำคัญของ PC ก็คือ Relative Magnitude ของ λ ซึ่งเป็นอัตราส่วนเปรียบเทียบระหว่าง λ กับ S ซึ่งเป็นผลรวมความผันแปรของตัวแปร X's ซึ่งพัฒนาได้ดังนี้

$$S = \sum_{i}^{n}(X_{2i}-\bar{X}_2)^2 + \sum_{i}^{n}(X_{3i}-\bar{X}_3)^2 + \ldots + \sum_{i}^{n}(X_{ki}-\bar{X}_k)^2$$

$$= \sum_{i}^{n}x_{2i}^2 + \sum_{i}^{n}x_{3i}^2 + \ldots + \sum_{i}^{n}x_{ki}^2$$

= ผลรวมของสมาชิกใน Main Diagonal ของแมตริกซ์ X'X เมื่อตัวแปร x's และ Y ได้รับการแปลงรูปให้เป็น Deviated Form หรือ Standardized Form

$$= tr\ (X'X)$$ [3]

แต่ $tr\ [P'(X'X)P] = tr\ (X'X)$ และ $tr\ [P'(X'X)P] = tr\ [(XP)'(XP)]$

$= tr\ [Z'Z] = tr\ D$ เมื่อ $D = diag\ (\lambda_2, \lambda_3, \ldots, \lambda_k)$

ดังนั้น

$$S = tr(X'X) = tr(Z'Z) = tr\ D = \lambda_2+\lambda_3+\ldots+\lambda_k$$

แต่เนื่องจากตัวแปร X's และ Y เป็น Standardize Form ดังนั้น $X'X = R_X$ ดังนั้น

---

[1] ผู้เขียนใช้ดัชนีของ λ ตามดัชนีของตัวแปร x โปรดอย่าได้สงสัยว่าเพราะเหตุใดจึงไม่มี λ_1

[2] อ่าน Johnston, J., **Op. Cit.**, p. 323-4ซึ่งแสดงการหาลำดับที่ของ PC โดยอาศัย Lagrangian Multiplier

[3] อ่านตอน 4.4.1 และ 8.4.6

$$s = tr(X'X) = tr\ R_x = k - 1$$

$$\lambda_2 + \lambda_3 + \ldots + \lambda_k = k - 1$$ [1]

หรือ $$\frac{\lambda_2}{k-1} + \frac{\lambda_3}{k-1} + \ldots + \frac{\lambda_k}{k-1} = 1 \text{ หรือ } 100\%$$

ด้วยเหตุนี้ ดัชนีที่ใช้วัดว่า z_j มีความสำคัญเพียงใด ควรคัดทิ้งไปก่อนการวิเคราะห์สมการถดถอย Y = f(Z's) หรือไม่ก็คือ $\frac{\lambda_j}{k-1} \times 100\%$ ถ้า z_j ตัวใดให้ค่า $\frac{\lambda_j}{k-1}$ สูงก็แสดงว่า z_j ตัวนั้นมีความสำคัญสูงเพราะดึงดูดเอาความผันแปรร่วมของ X's ไว้ได้มาก z_j ตัวใดให้ค่า $\frac{\lambda_j}{k-1}$ ต่ำก็แสดงว่า z_j ตัวนั้นมีความสำคัญน้อยเพราะดึงดูดเอาความผันแปรร่วมของ X's ไว้ได้น้อย ด้วยเหตุนี้และโดยอาศัยกฎเกณฑ์ดังกล่าว เราจึงสามารถขจัดตัวแปร z_j ที่ไม่มีความสำคัญทิ้งไปก่อนที่จะนำไปวิเคราะห์สมการถดถอย Y = f(Z's)

8.5.3.3 หลักเกณฑ์การขจัดตัวแปร z_j ก่อนการวิเคราะห์สมการถดถอย Y = f(Z's)

ดังที่กล่าวมาแล้วในตอนต้นว่าตัวแปร Z อาจมีจำนวนได้สูงสุดไม่เกินจำนวนตัวแปรอิสระ X แต่โดยทั่วไปจะน้อยกว่าเพราะ z_j บางตัวอาจไม่มีความหมาย หรือความสำคัญเลยเพราะดึงดูดความผันแปรร่วม S ไว้ได้น้อยเกินไป และในทางปฏิบัติเราควรคัดตัวแปร z_j ที่ไม่จำเป็นทิ้งไปเสียบ้างเพื่อสงวน df หลักเกณฑ์คัดเลือก z_j ทิ้งไปหรือคง z_j เอาไว้มีหลายวิธีดังนี้

1. Kaiser's Criterion

วิธีนี้ให้พิจารณา z_j ทิ้งถ้าพบว่า $\lambda_j < 1$ โดยที่ $z_j = Xp_j$ เมื่อ p_j คือ Normalized Characteristic Vector ที่สอดคล้องกับ λ_j (หรือคัด z_j ทิ้ง ถ้าพบว่า $z_j'z_j < 1$))

วิธีปฏิบัติสำหรับกฎเกณฑ์นี้สามารถกระทำได้โดยง่ายเพียงแต่ตรวจสอบแมตริกซ์ D โดยที่ $D = diag(\lambda_2, \lambda_3, \ldots, \lambda_k)$ หากพบว่า λ ตัวใดมีค่าต่ำกว่า 1 ก็ให้ตัดทิ้งไปซึ่งมีผล

---

[1] ขอให้สังเกตว่า k-1 ก็คือจำนวนราก λ หรือจำนวนตัวแปรอิสระ X นั่นเอง โดยนัยนี้ถ้าหากว่างานวิจัยประกอบด้วยตัวแปร X ทั้งสิ้น 5 ตัว คือ X_2, X_3, X_4, X_5, X_6 จะพบว่า $\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6 = 5$ ความผันแปรร่วมของ $X_2, X_3, \ldots, X_6$ ที่ z_j แต่ละตัวดึงดูดเอาไว้ก็คือ $\frac{\lambda_j}{5}$ เช่น $\frac{\lambda_2}{5}, \frac{\lambda_3}{5}, \ldots, \frac{\lambda_6}{5}$ สมมุติว่า $\lambda_2 = 2.5$, $\lambda_3 = 1.25$, $\lambda_4 = 1.00$, $\lambda_5 = 0.20$, $\lambda_6 = 0.05$ จะพบว่า $\frac{\lambda_2}{5} = \frac{2.5}{5} = .50$ $\frac{\lambda_3}{5} = \frac{1.25}{5} = .25$, $\frac{\lambda_4}{5} = \frac{1}{5} = .20$, $\frac{\lambda_5}{5} = \frac{0.20}{5} = .04$, $\frac{\lambda_6}{5} = \frac{0.05}{5} = .01$ หรือ z_2, z_3, z_4, z_5, z_6 ดึงดูดอิทธิพลร่วมของ X's ไว้ในตัวได้ 50%, 25%, 20%, 4% และ 1% ตามลำดับ

เสมือนกำหนดให้ λ ตัวนั้นมีค่าเป็น 0 และเมื่อเราจัดเรียงค่าของ λ ตามลำดับจากมากไปหาน้อย คือ $\lambda_2 > \lambda_3 > \ldots > \lambda_k$ ค่าของ λ ที่อยู่ท้าย ๆ ก็ถูกขจัดทิ้ง หรือถูกกำหนดให้เป็น 0 สมมุติ $\lambda_{r+1} < 1$ ผลที่ตามมาก็คือ $\lambda_{r+2}, \lambda_{r+3}, \ldots, \lambda_k$ ล้วนมีค่าน้อยกว่า 1 และถูกคัดทิ้ง (คือกำหนดให้เท่ากับ 0) ดังนี้

$$D = \begin{bmatrix} \lambda_2 & & & & & & \\ & \lambda_3 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

และแมตริกซ์ $P = [p_2, p_3, \ldots, p_k]$ โดยที่ p_j คือ Normalized Vector ที่สอดคล้องกับ λ_j ก็จะลดรูปเป็น $P = [p_2, p_3, \ldots, p_r, 0, \ldots, 0] = [P_1 \; P_2]$ เมื่อ 0 คือ Zero Vector

ดังนั้น $$Z = XP = X[p_2, p_3, \ldots, p_r, \vdots\, 0, \ldots, 0] = X[P_1 \vdots P_2]$$
$$= [Xp_2, Xp_3, \ldots, Xp_r, \vdots\, 0, \ldots, 0] = [XP_1 \vdots 0]$$
$$= [Z_2, Z_3, \ldots, Z_r, \vdots\, 0, 0, \ldots, 0] = [Z_1 \vdots Z_2]$$

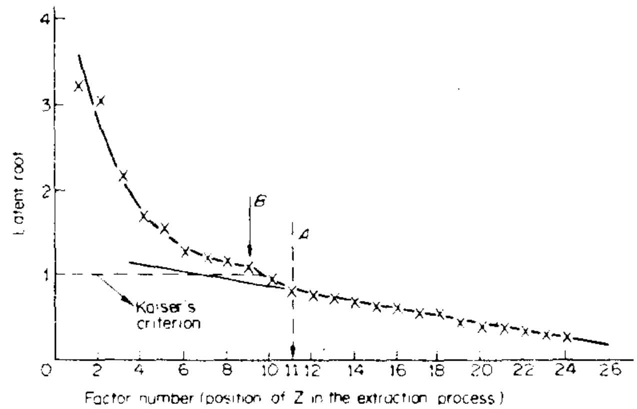
ดังนั้นสมการถดถอย $Y = X(PP')\beta + U$ จึงเปลี่ยนรูปเป็น $Y = X(P_1P_1')\beta + U$ หรือ $Y = Z_1\theta_1 + U$ และค่าประมาณของ θ_1 ก็คือ $\hat{\theta}_1 = (Z_1'Z_1)^{-1}Z_1'Y$ เมื่อย้อนคืนสู่ β จะพบว่า $\hat{\theta}_1 = P_1'\beta$ ซึ่งเมื่อคูณตลอดด้วย P_1 จะได้

$$P_1\hat{\theta}_1 = P_1P_1'\beta$$

หรือ $$\hat{\beta} = P_1\hat{\theta}_1$$

2. Cattell's Scree test'

การทดสอบวิธีนี้ให้พิจารณาลักษณะการเปลี่ยนโค้งของคู่ลำดับ (j, λ_j) กล่าวคือให้ใช้แกนนอนแทนระหัส (เลขที่) ของ λ และแกนตั้งแทนค่าของ λ เมื่อนำคู่ลำดับ (j, λ_j) ไปพล๊อตลงในแกน Coordianate ดังกล่าว ทางเดินของจุด (j, λ_j) จะมีลักษณะของโค้งที่ลาดเอียงลงทางขวาดังภาพ



ขอให้สังเกตว่าเส้นโค้งจะลาดลงตามลำดับจนใกล้ทีจะตัดแกนนอน (แต่ต้องไม่ตัดกับแกนนอน เพราะถ้าตัดแกนนอนแสดงว่า $\lambda_j = 0$ หรือ $X'X$ เป็น Singular Matrix มิใช่ Near Singularity ซึ่งเป็นประเด็นที่เรากำลังศึกษาอยู่)

การตัดสินใจให้พิจารณาจากโค้งโดยคง λ_j ทุกตัวที่อยู่ก่อนจุดเปลี่ยนโค้ง (Curvature) เอาไว้ และตัด λ_j ทุกตัวที่อยู่ถัดจุดเปลี่ยนโค้งทิ้ง หรือกล่าวให้เข้าใจง่ายขึ้นก็คือให้ตัด λ_j ทุกตัวที่ทำให้โค้งมีลักษณะยืดตัวเป็นแนวเส้นตรง

จากภาพจะเห็นว่า $\lambda_{12}, \lambda_{13}, \ldots, \lambda_{24}$ สอดคล้องกับโค้งที่ยืดตัวเป็นเส้นตรง ให้ตัด $\lambda_{12}, \lambda_{13}, \ldots, \lambda_{24}$ ทิ้งและคง $\lambda_2, \lambda_3, \ldots, \lambda_n$ เอาไว้ เมื่อเปรียบเทียบกับ Kaiser's Criterion จะพบว่า Kaiser's Criterion คง λ_j ไว้น้อยกว่าคือ ถ้าใช้ Kaiser's Criterion เราจะคง $\lambda_2, \lambda_3, \ldots, \lambda_9$ เอาไว้ นอกนั้นตัดทิ้ง

---

[1] Koutsoyiannis, A., Op. Cit., p.423.4

3. Bartlett's Criterion

หลักเกณฑ์ของวิธีนี้ก็คือ ภายหลังจากได้ตรวจสอบ $\lambda_2 > \lambda_3 > \dots > \lambda_k$ แล้ว พบว่า $\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_k$ มีค่าต่ำขณะที่ $\lambda_2, \lambda_3, \dots, \lambda_r$ มีค่าสูงอยู่ในระดับที่พอยอมรับได้[1] ปัญหาที่จะต้องกระทำก็คือเราควรคง λ_j ตัวใดกลุ่ม ($\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_k$) ไว้โดยตั้งเป็นข้อสมมุติฐานว่า λ_j เหล่านี้มีค่าไม่ต่างกัน คือ

$$H_0 : \lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_k \quad \text{VS} \quad H_1 : \lambda \text{ not all equal}$$

ซึ่งถ้ายอมรับ H_0 ก็แสดงว่า $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_k$ เมื่อ $\lambda_k = \lambda_{min}$ หรือยอมรับว่า λ_j ทุกตัวในกลุ่ม ($\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_k$) มีค่าต่ำเท่ากับ λ_{min} ซึ่งโดยปกติจะต่ำมากจนไม่ถือว่ามีความสำคัญ แล้วตัดสินใจตัด $Z_{r+1}, Z_{r+2}, \dots, Z_k$ ทิ้งไป แต่ถ้าปฏิเสธ H_0 ก็แสดงว่าเราเชื่อว่า λ_{r+1} ต้องไม่มีค่าต่ำเท่ากับ λ_{min}[2] ให้ตัดสินใจคง Z_{r+1} ไว้ แล้วทดสอบสมมุติฐาน $H_0 : \lambda_{r+2} = \lambda_{r+3} = \dots = \lambda_k$ ต่อไป หากปฏิเสธ H_0 ก็ให้คง Z_{r+2} ไว้ แล้วทดสอบ $H_0 : \lambda_{r+3} = \lambda_{r+4} = \dots = \lambda_k$ ต่อไป ทำดังนี้เรื่อยไปจนกว่าจะยอมรับ H_0 สมมุติว่าเมื่อทดสอบถึง $H_0 : \lambda_{r+5} = \lambda_{r+6} = \dots = \lambda_k$ แล้วปรากฏว่าเรายอมรับสมมุติฐานนี้ก็แสดงว่าเราจะคง $Z_2, Z_3, \dots, Z_r, Z_{r+1}, \dots, Z_{r+4}$ เอาไว้เพื่อวิเคราะห์สมการถดถอย $Y = Z_1\theta_1 + U$ หรือ $Y = \gamma_2 Z_2 + \gamma_3 Z_3 + \dots + \gamma_{r+4} Z_{r+4} + u$ ต่อไป

ตัวทดสอบสำหรับวิธีนี้ก็คือ

$$\chi^{2^*} = n \log_e\left\{(\lambda_{r+1}\lambda_{r+2}\cdots\lambda_k)^{-1}\left(\frac{\lambda_{r+1}+\lambda_{r+2}+\dots+\lambda_k}{k-1-r}\right)^{k-1-r}\right\}$$

โดยปฏิเสธ H_0 ณ ระดับนัยสำคัญ α เมื่อ $\chi^{2^*} > \chi^2_{\frac{\alpha}{2}, \nu}$ โดยที่ $v = \frac{1}{2}(k-r-2)(k-r+1)$

8.5.3.4 การประมาณค่าพารามิเตอร์ β

การประมาณค่า β ใน PCR สามารถกระทำได้ 2 วิธีดังนี้

---

[1] ควรใช้ Kaiser's Criterion หรือ Cattel's Scree Test ตรวจสอบดู

[2] โดยอาศัย Sequence $\lambda_2 > \lambda_3 > \dots > \lambda_k$ แสดงว่า λ_{r+1} ต้องสูงกว่า $\lambda_{r+2}, {}_{r+3}, \dots, \lambda_k$ อยู่แล้ว เมื่อปฏิเสธ H_0 (ยอมรับ H_1) ก็แปลว่าเรายอมรับว่า λ_{r+1} มีค่าสูงพอที่จะคงเอาไว้ได้

วิธีที่ 1 อาศัย Inverse Transformation

จากสมการ $Y = Z_1\theta_1 + U$ เมื่อ $Z_1 = XP_1$ และ $\theta_1 = P_1'\beta$ โดยที่ $P_1 = [p_2, p_3, \ldots, p_r]$ และโดยอาศัย OLS เราจึงสามารถประมาณค่า θ_1 ได้โดยง่ายคือ

$$\hat{\theta}_1 = (Z_1'Z_1)^{-1}Z_1'Y$$

เมื่ออาศัย Inverse Transformation จึงพบว่า $\hat{\beta} = P_1\hat{\theta}_1$

วิธีนี้มีผลเช่นเดียวกับ Restricted Least Square ที่ประมาณค่า β จากสมการ $Y = Z\beta + U$ ภายใต้ข้อจำกัดว่า $P_2'\theta_2 = 0$

วิธีที่ 2 ใช้ OLS

สมมุติว่าภายหลังจากพิจารณาคัดตัวแปร Z ที่ไม่สำคัญบางส่วนทิ้งไปแล้วเหลือตัวแปร Z ที่สำคัญ r-1 ตัวคือ $Z_2, Z_3, \ldots, Z_r$ ให้ดำเนิน การวิเคราะห์สมการถดถอย Y = f(Z's) โดยอาศัย OLS ตามปกติคือ

$$\text{จาก} \quad Y = \gamma_2 Z_2 + \gamma_3 Z_3 + \ldots + \gamma_r Z_r + u$$

แล้วประมาณค่า $\gamma_2, \gamma_3, \ldots, \gamma_r$ โดยอาศัย OLS ได้ $\hat{\gamma}_2, \hat{\gamma}_3, \ldots, \hat{\gamma}_r$ สมการประมาณค่าที่ได้ก็คือ $Y = \hat{\gamma}_2 Z_2 + \hat{\gamma}_3 Z_3 + \ldots + \hat{\gamma}_r Z_z$

แต่เนื่องจาก $Z_j = Xp_j = p_{j2}X_2 + p_{j3}X_3 + \ldots + p_{jk}X_k \; ; \; j = 2, 3, \ldots, k$

ให้แทนที่ Z_j ด้วย Xp_j จะได้สมการ Y = f(X's) ตามเดิมคือ

$$\begin{aligned} Y &= \hat{\gamma}_2(p_{22}X_2 + p_{23}X_3 + \ldots + p_{2k}X_k) + \hat{\gamma}_3(p_{32}X_2 + p_{33}X_3 + \ldots p_{3k}X_k) \\ &\quad + \ldots + \hat{\gamma}_k(p_{k2}X_2 + p_{k3}X_3 + \ldots + p_{kk}X_k) \\ &= (\hat{\gamma}_2 p_{22} + \hat{\gamma}_3 p_{32} + \ldots + \hat{\gamma}_k p_{k2})X_2 + (\hat{\gamma}_2 p_{23} + \hat{\gamma}_3 p_{33} + \ldots + \hat{\gamma}_k p_{k3})X_3 \\ &\quad + \ldots + (\hat{\gamma}_2 p_{2k} + \hat{\gamma}_3 p_{3k} + \ldots + \hat{\gamma}_k p_{kk})X_k \\ &= \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \ldots + \hat{\beta}_k X_k \end{aligned}$$

จะเห็นได้ว่า $\hat{\beta}_j$ เป็น Linear Combination ของ $\hat{\gamma}_2, \hat{\gamma}_3, \ldots, \hat{\gamma}_k$ และขอให้สังเกตว่า ถ้าเราคงตัวแปร Z ทุกตัวเอาไว้ ค่าประมาณของ β_{PC} จะมีผลตรงกันกับ $\hat{\beta}_{OLS}$ ทุกประการ

ข้อสังเกต

ขอให้สังเกตว่าวิธีที่ 1 และวิธีที่ 2 ก็คือวิธีเดียวกัน