บทที่ 5

Nonnormal Error

5.1 ความนำ

จากเค้าโครงของการศึกษาวิจัยโดยอาศัยเทคนิคการวิเคราะห์ความถดถอยที่กล่าวถึงแล้วในบทที่ 1 คือการวิจัยโดยอาศัยเทคนิคการวิเคราะห์ความถดถอยนั้น นักวิจัยจะต้องดำเนินงานเป็น 2 ระยะติดต่อกันคือ

ระยะที่ 1 วิเคราะห์สมการถดถอย Y = f(X's) ภายใต้ข้อตกลงของสมการถดถอย รวมถึงการตรวจสอบคุณภาพของค่าประมาณโดยนัยของวิธีการทางสถิติ และอื่น ๆ

ระยะที่ 2 ตรวจสอบว่า Residual และตัวแปรอิสระอื่น ๆ ว่ามีคุณสมบัติสอดคล้องกับข้อตกลงของสมการถดถอยหรือไม่ ถ้าเกิดความขัดแย้งกับข้อตกลงก็ให้แก้ไขให้ถูกต้อง และสอดคล้องกับข้อตกลง

การศึกษาในบทที่ 1 ถึง 4 เป็นการศึกษาเพื่อดำเนินการในระยะที่ 1 เท่านั้น การศึกษาในลำดับต่อจากนี้จะเป็นการศึกษาเพื่อดำเนินการในระยะที่ 2 รวมถึงการศึกษาถึงเทคนิค และพัฒนาการใหม่ที่นักศึกษาและนักวิจัยพึงทราบเพื่อประโยชน์ในงานวิจัยและช่วยให้งานวิจัยมีคุณภาพสูง

5.2 Nonnormal Error และผลกระทบต่องานวิเคราะห์ความถดถอย

Nonnormal Error หมายถึงสถานการณ์ที่ residual มีการแจกแจงแบบอื่น ๆ ที่มิใช่การแจกแจงแบบปกติที่กำหนดไว้ในข้อตกลง $(u_i \sim N(0,\sigma_u^2))$[1]

เมื่อ u_i มิได้มีการแจกแจงแบบปกติ ผลกระทบที่ติดตามมาก็คือ Y_i และ $\hat{\beta}$ จะไม่มีการแจกแจงแบบปกติ เช่น t-test, F-test, Z-test และอื่น ๆ รวมทั้งไม่อาจสร้างเขตเชื่อมั่น (Confidence Region) หรือช่วงเชื่อมั่นซึ่งอ้างอิงการแจกแจงปกติได้ แต่ทั้งนี้เวคเตอร์ $\hat{\beta}$ จะยังคงเป็น BLUE อยู่เช่นเดิม ไม่เปลี่ยนแปลงเมื่อ $\hat{\beta}$ คือ OLS-Estimator ทั้งนี้เพราะ OLS มิได้อาศัยข้อตกลงว่า $u_i \sim N(0,\sigma^2)$ หากอาศัยเพียงข้อตกลงว่า u_i เป็นตัวแปรสุ่มที่ $E(u_i) = 0$,

---

[1] ต่อไปจะเขียน σ^2 โดยละ Subscript "u" ไว้ในฐานที่เข้าใจ

$E(u_i u_j) = \begin{cases} 0 & ; i \neq j \\ \sigma^2 & , i = j \end{cases}$ เท่านั้น ผลสะท้อนดังกล่าวสามารถสรุปได้เป็นข้อ ๆ ดังนี้

1. $\hat{\beta} = (X'X)^{-1} X'Y$ เป็น BLUE และ Consistent

2. $\hat{\sigma}^2$ เป็นตัวประมาณค่าที่ปราศจากอคติของ σ^2 และ Consistent

3. $\hat{\beta}$ และ $\hat{\sigma}^2$ จะไม่มีคุณสมบัติของ Efficiency หรือ Asymptotic Efficiency

4. $\hat{\beta}$ จะไม่มีการแจกแจงแบบปกติ และ $(n-k)\hat{\sigma}^2/\sigma^2$ จะไม่มีการแจกแจงแบบไคสแควร์ (χ^2) ซึ่งในสถานการณ์เช่นนี้ t-test และ F-test จะไม่มีประสิทธิภาพในกรณี Finite Sample ทั้งนี้รวมตลอดถึง Test ทั้งหลายที่อ้างอิงการแจกแจงปกติเช่น Test ในกรณีที่มี prior Information[1]

5. EGLS - Estimator คือ $\hat{\beta} = (X'\hat{\Omega}^{-1} X)^{-1} X'\hat{\Omega}^{-1} Y$ ยังคงมี Asymptotic Property

5.3 การทดสอบความเป็นจริงของข้อตกลง (Tests for Normality)

การทดสอบความเป็นจริงของข้อตกลงว่า u_i มีการแจกแจงแบบปกตินั้น สามารถกระทำได้หลายวิธี วิธีที่นิยมใช้กันมากในปัจจุบันมี 3 วิธีคือ

วิธีที่ 1 การวัดค่า Skewness และ Kurtosis

วิธีที่ 2 Kolmogorv - Smirnov Test (KS)

วิธีที่ 3 Sharpiro - Wilk Test

ก. การวัดค่า Skewness และ Kurtosis (Peakedness - Flatness)

โดยปกติแล้วถ้าตัวแปรสุ่มใดมีการแจกแจงแบบปกติ (มีโค้งเป็นรูประฆังคว่ำ) ตัวแปรสุ่มนั้นจะมีค่าความเบ้ (Skewness) เท่ากับ 0 และค่าความโด่ง (Kurtosis) เท่ากับ 3 ซึ่งเรานิยมใช้ค่าทั้งสองนี้เป็นมาตรฐานเพื่อเปรียบเทียบดูว่าตัวแปรสุ่มที่สนใจมีการแจกแจงแบบปกติหรือไม่

สัมประสิทธิ์ทั้งสองดังกล่าว นิยามได้ดังนี้

(1) Coef. of Skewness $\hat{\gamma}_1 = \dfrac{1/n \ \Sigma(e_i - \bar{e})^3}{s_e^3}$

ถ้า $\hat{\gamma}_1 > 0$ แสดงว่า โค้งเบ้ขวา ถ้า $\hat{\gamma}_1 < 0$ แสดงว่าโค้งเบ้ซ้าย

ถ้า $\hat{\gamma}_1 = 0$ แสดงว่าโค้งสมมาตร

---

[1] อ่าน Judge, G.G. et. al., Op. Cit., CH. 3, P. 54-107

(2) Coef. of Kurtosis $\hat{\gamma}_2 = \dfrac{1/n\ \Sigma(e_i-\bar{e})^4}{s_e^4}$

ถ้า $\hat{\gamma}_2 > 3$ แสดงว่าโค้งแบนกว่าโค้งปกติ

ถ้า $\hat{\gamma}_2 < 3$ แสดงว่าโค้งโด่งกว่าโค้งปกติ

ถ้า $\hat{\gamma}_2 = 3$ แสดงว่าโค้งสมมาตรคล้ายโค้งปกติ

ข. Kolmogorov-Smirnov Test หรือ Kolmogov - Smirnov goodness of fit - test (KS)

KS คือตัวทดสอบที่อาศัยผลต่างระหว่าง Observed Cumulative Histogram กับ Cumulative Distribution Function F(·) ของตัวแปรสุ่มที่สนใจ ในที่นี้ F(·) คือ Normal Cumulative Distribution Function

ให้ $e_{(1)} < e_{(2)} < e_{(3)} < \ldots < e_{(n)}$ คือ Ordered Observation ของ residual $e_1, e_2, \ldots, e_n$

ให้ $\frac{i}{n}$ = Observed Cumulative ของ residual ลำดับที่ *i*

ให้ $F_u(e_{(i)}) = Pr\{U < e_{(i)}\} = Pr\ \{Z < \dfrac{e_{(i)}-\bar{e}}{s_e}\}$

= Hypothesized Cumulative ของ residual ลำดับที่ i

ดังนั้นโดยนัยของ KS เราจะปฏิเสธสมมุติฐานหลัก

H_0 : U มีการแจกแจงแบบปกติ vs H_1 : U มีการแจกแจงแบบอื่น ๆ เมื่อปริมาณความแตกต่างตัวที่ใหญ่ที่สุดของเซท $\{|\frac{i}{n} - F_u(e_{(i)})|\}$ มีค่าสูงกว่าค่าวิกฤต หรือนัยหนึ่ง ปฏิเสธสมมุติฐานหลัก ณ ระดับนัยสำคัญ α เมื่อ $D \geq k$ โดยที่

$D = \max\{|\frac{i}{n} - F_u(e_{(i)})|\}$ และ k คือค่าวิกฤตจากตาราง KS

ความจริงแล้วเราสามารถใช้ χ^2-test (goodness of Fit Test) สำหรับทดสอบสมมุติฐานข้างต้นนี้ได้ แต่ทั้งนี้จำเป็นต้องจัดกลุ่ม residual ออกเป็นกลุ่มย่อยหลายกลุ่ม แล้วคำนวณหาค่า χ^2 โดยที่

$$\chi_c^2 = \sum_i^k \frac{(0_i - np_{io})^2}{np_{i0}}$$

O_i = จำนวนความถี่ที่ค่าสังเกตปรากฏอยู่ในกลุ่มที่ i, i = 1, 2, ..., k

n = ขนาดตัวอย่าง

p_{io} = ค่าความน่าจะเป็นที่ตัวแปรสุ่ม u (ในที่นี้ $u \sim N(0, \sigma^2)$) จะปรากฏอยู่ในกลุ่มที่ i

ดังนั้น $np_{io} = E_i$ = Expected Frequency จากประชากรปกติ

อย่างไรก็ตาม ถ้าใช้ X^2-test เรามักประสบปัญหาเรื่องการจัดกลุ่มค่าสังเกตว่าควรจัดกลุ่มค่าสังเกตอย่างไร ทั้งนี้เพราะจำเป็นต้องคำนึงถึงช่วงของค่าตัวแปรสุ่มคือ $(-\infty, \infty)$ ด้วยซึ่งการแบ่งช่วงของตัวแปรสุ่ม $(-\infty, \infty)$ ออกเป็นช่วงสั้น ๆ จะมีผลให้สามารถคำนวณหาค่าความน่าจะเป็น p_{io} ได้ ด้วยเหตุนี้จึงพบว่า KS - test มีความยืดหยุ่นได้มากกว่าซึ่งสามารถสรุปให้เห็นได้ดังนี้

1. KS สามารถใช้ได้กับข้อมูลเดิมโดยตรง สามารถเปรียบเทียบค่าความถี่สะสมของตัวแปรสุ่มทุกค่าโดยไม่จำเป็นต้องแบ่งกลุ่มแล้วเปรียบเทียบเป็นรายกลุ่ม เช่น ใน X^2-test
2. KS สามารถใช้ได้กับขนาดตัวอย่างได้ทุกขนาด
3. KS ให้คำตอบได้รวดเร็วกว่า วิเคราะห์ง่ายกว่า

ค. Shapiro - Wilk Test

ตัวทดสอบนี้เสนอโดย ชาปีโร และ วิลค์ (Shapiro, S.S. and Wilk, M.B., 1965)[1] โดยเสนอตัวสถิติดังนี้

$$W = \frac{\left[\sum_{i=1}^{m} a_{(n-i+1)}\{e_{(n-i+1)} - e_{(i)}\}\right]^2}{\sum_{i}^{n}(e_i - \bar{e})^2} = \frac{b}{s^2}$$

โดยที่

1. $m = \frac{n}{2}$ หรือ $\frac{n-1}{2}$ ทั้งนี้ขึ้นอยู่กับ n ว่าเป็นเลขคู่ หรือ เลขคี่
2. $a_{(n-i+1)}$ ปรากฏในตารางท้ายเล่ม

---

[1] อ่าน Maddala, G.S., Op. Cit., p.306 และ Judge, G.G., et. al., Op. Cit., p.301 และ Shapiro, S.S. and Wilk, M.B., "An Analysis of Variance Test for Normality (Complete Sample)" (Biometrika, 1965, 52) p.602-6

3. $e_{(i)}$ คือ Ordered Observation ของ residual กล่าวคือ จาก residual $e_1, e_2, .., e_n$ เราจัดเรียงตามลำดับจากค่าน้อยไปมากเป็น $e_{(1)} < e_{(2)} < \ldots < e_{(n)}$

4. $\bar{e} = \frac{1}{n} \sum_{i}^{n} e_i$

5. ปฏิเสธ H_0 : Normality vs H_1: Non-Normality เมื่อ $W > W_\alpha$ (ดูตารางท้ายเล่ม)

สำหรับ W-test มีข้อสังเกตและสิ่งพึงทราบดังนี้

1. $\sum_{i}^{m} a_{(n-i+1)} \{e_{(n-i+1)} - e_{(i)}\}^2$ ก็คือ Sum Square ของผลต่างระหว่าง residual ค่าสูงกับ residual ค่าต่ำที่มีตำแหน่งสอดคล้องกันเมื่อนับจากปลายทั้งสองของ Sequence

$$e_{(1)} < e_{(2)} < e_{(3)} < \ldots < e_{(n-2)} < e_{(n-1)} < e_{(n)}$$

นั่นคือ $[\sum_{i}^{m} a_{(n-i+1)} \{ e_{(n-i+1)} - e_{(i)} \}]^2 = [a_{(n)} (e_{(n)} - e_{(1)}) + a_{(n-1)} (e_{(n-1)} - e_{(2)})$

$$+ a_{(3)} (e_{(n-2)} - e_{(3)}) + \ldots + a_{(n-\frac{n}{2}+1)} (e_{(\frac{n}{2}+1)} - e_{(\frac{n}{2})})]^2$$

2. ความจริงแล้ว ชาปิโร และ วิลค์ พัฒนาสูตรนี้ขึ้นมาโดยใช้ BLUS-residual[1] มิใช่

---

[1] BLUS = Best Linear unbiasd Scalar variance - covariance matrix ซึ่งเป็นวิธีประมาณค่าพารามิเตอร์ β ในกรณีที่มีปัญหา Serial Correlation เสนอโดยธีล (Theil, H., 1965, 1968) อ่านรายละเอียดใน Johnston, J., Op. Cit., p.254-8 ธีลเสนอวิธี BLUS ขึ้นมาเพื่อแก้ปัญหาของ Residual e_i ซึ่งโดยปกติจะไม่เป็นอิสระต่อกัน กล่าวคือ

$$\begin{aligned} e &= Y - \hat{Y} = Y - X\hat{\beta} = (X\beta + U) - X(X'X)^{-1}X'Y \\ &= X\beta + U - X(X'X)^{-1}X'(X\beta + U) \\ &= U - X(X'X)^{-1}X'U \\ &= (I_n - X(X'X)^{-1}X')U \\ &= MU \end{aligned}$$

ซึ่งมีผลให้ $E(ee') = E(MUU'M') = ME(UU')M' = \sigma_u^2 M I_n M' = \sigma_u^2 M \neq \sigma_u^2 I_n$ แสดงว่าเวคเตอร์ e มีสหสัมพันธ์ภายในต่อกันหรือ e_i ไม่เป็นอิสระกับ e_j ซึ่งในกรณีเช่นนี้ $(e_1, e_2, \ldots, e_n)$ จึงมิอาจถือว่าเป็น Random Sample ได้ ธีลได้พัฒนาวิธี BLUS ขึ้นเพื่อหาทางให้ $(e_1, e_2, \ldots, e_n)$ เป็น Random Sample อย่างไรก็ตามวิธี BLUS กระทำได้เพียงแต่คัดเอา e_i ที่ไม่อิสระต่อกันออกทำให้เหลือตัวอย่างของ e เพียง n-k ตัวที่เป็นอิสระต่อกัน

OLS-residual ที่เสนอไว้ข้างบนนี้ แต่จากการศึกษาในลำดับต่อ ๆ มา ทั้งโดยนัยของวิธีมอนติคาร์โล และอื่น ๆ พบว่าถ้าคำนวณหาค่า W โดยอาศัย OLS-residual W-test จะมี Power สูงกว่าเมื่อคำนวณหาโดยใช้ BLUS-residual

3. W-test มี Power สูงกว่า KS และ Skewness-Kurtosis Criteria จากการศึกษาโดยนัยของการทดลองมอนติคาร์โลพบว่า W-test มี Power สูงที่สุดขณะที่ KS มี Power ต่ำที่สุด และ Skewness - Kutosis Criteria มี Power สูงแต่ก็ไม่สูงเท่า W-test

ตัวอย่างการคำนวณสำหรับ W-test ปรากฏดังนี้

สมมุติว่ามี Residual อยู่ 7 ค่า คือ $e_1=6$, $e_2=1$, $e_3=-4$, $e_4=8$, $e_5=-2$, $e_6=5$, $e_7=0$

(1) จัดลำดับของ Residual จะพบว่า $-4<-2<0<1<5<6<8$

(2) คำนวณหาค่า $\sum_i^7 (e_i-\bar{e})^2$ ได้ $\sum_i^7 (e_i-\bar{e})^2 = 147-28 = 118$

(3) จากตารางสัมประสิทธิ์ $\{a_{(n-i+1)}\}$ สำหรับ $n=7$ จะพบว่า

$$a_7 = 0.6233,\ a_6 = 0.3031,\ a_5 = 0.1401,\ a_4 = 0.000$$

ดังนั้น

$$b = 0.6233(8+4)+0.3031(6+2)+0.1401(5-0) = 10.6049$$

(4) $W = (10.6049)^2/118 = 0.9530$

(5) จากตารางค่าวิกฤติของ W พบว่า $W > W_{.50}$ แสดงว่ามี Non-Normality

5.4 ROBUST Estimator

จากที่ศึกษาในบทและตอนที่ผ่านมาเราทราบว่า ถ้าตัวแปรสุ่ม u มีการแจกแจงแบบ

ปกติแล้ว OLS-Estimator จะเป็น Efficient Estimator[1] แต่ถ้า u มิได้มีการแจกแจงแบบปกติ OLS-Estimator จะไม่เป็น Efficient Estimator ซึ่งในสถานการณ์เช่นนี้น่าจะมี Estimator อื่นของ β ที่ประมาณค่า β ได้ดีกว่า OLS-Estimator ตัวประมาณที่กล่าวถึงเรียกว่า ROBUST Estimator

จากการศึกษาพบว่าเราสามารถหา ROBUST Estimator ได้หลายตัว แต่จะเสนอในที่นี้เพียง 2 วิธีเพื่อเป็นแนวทางดังนี้

5.4.1 Minimum Absolute Deviation (MAD)

วิธี MAD เป็นวิธีประมาณค่า β โดยการ minimize $\Sigma|e_i|$ ค่าประมาณของ β ที่ได้เรียกว่า MAD - Estimator ($\hat{\beta}_m$) ซึ่งโดยปกติเรามักจะคำนวณหา $\hat{\beta}_m$ โดยวิธี Search หรือ Linear Programming มากกว่าการแก้สมการ Normal Equation

วิธีที่พัฒนาล่าสุดเพื่อหา MAD - Estimator คือวิธี Iterated WLS ซึ่งเสนอโดย Schossmacher และ Fair ดังอัลกอริทึมต่อไปนี้

(1) จากสมการ $Y = X\beta + U$ ให้ประมาณค่า β โดยอาศัยวิธี OLS กล่าวคือประมาณค่าเวคเตอร์ β ด้วย $\hat{\beta} = (X'X)^{-1}X'Y$ แล้วคำนวณหา Residual $e = Y - \hat{Y}$

(2) ใช้ Residual $e = Y - \hat{Y}$ ในขั้นที่ (1) เป็นตัวถ่วงน้ำหนัก แล้วประมาณค่าเวคเตอร์ β จากสมการถ่วงน้ำหนักดังนี้คือ

$$S = \sum_{i}^{n} \hat{w}_i \{Y_i - (\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})\}^2$$

เมื่อ $$\hat{w}_i = \frac{1}{|e_i|} \; ; \; i = 1, 2, \ldots, n$$

หรืออีกนัยหนึ่งให้อาศัย Residual e ในขั้นที่ (1) เป็นตัวถ่วงน้ำหนักในรูป $w_i = \frac{1}{|e_i|}$ แล้ว คำนวณหาค่าประมาณของ β โดยวิธี WLS ได้ $\hat{\beta}_m$ ดังนี้

$$\hat{\beta}_m = (X'W^{-1}X)^{-1}X'W^{-1}Y$$

---

[1] Efficient Estimator มีคุณสมบัติสำคัญ 2 ประการคือ

1. เป็น Minimum Variance Unbiased Estimator
2. เป็น Sufficient Estimator

เมื่อ

$$\hat{W} = \begin{bmatrix} 1/|e_1| & & & \\ & 1/|e_2| & & \\ & & \ddots & \\ & & & 1/|e_n| \end{bmatrix}_{nxn} \quad \text{หรือ} \quad \hat{W}^{-1} = \begin{bmatrix} |e_1| & & & \\ & |e_2| & & \\ & & \ddots & \\ & & & |e_n| \end{bmatrix}$$

แล้วคำนวณหา Residual (เรียกว่า MAD Residual)

(3) นำ Residual ในขั้นที่ (2) มาเป็นตัวถ่วงน้ำหนัก แล้วประมาณค่า β โดยวิธี WLS เดียวกันกับในขั้นที่ (2)

(4) ดำเนินการเช่นนี้เรื่อยไปจนกระทั่งค่าของ $\hat{\beta}_m$ คงที่หรือ Converge

หมายเหตุ ถ้าพบว่า Residual ตัวใดมีค่าเท่ากับ 0 หรือใกล้ 0 ซึ่งจะมีผลให้ตัวถ่วงน้ำหนัก $\hat{w}_i = 1/|e_i| \to \infty$ ให้ตัด Residual ค่านั้นทิ้งหรือกำหนดให้ Residual ค่านั้นมีค่าเท่ากับ 0.00001

สำหรับการหา Confidence Region ของเวคเตอร์ β_m หรือ Confidence Interval ของ β_j; $j=1,2,\ldots,k$ นั้นนับว่ามีปัญหาเพราะเท่าที่เคยปฏิบัติมาในตอนก่อนเราจะอาศัยข้อตกลงของ u ว่า u มีการแจกแจงแบบปกติ

อย่างไรก็ตาม ด้วยการอาศัยข้อตกลงดังนี้คือ

1. u_i เป็นอิสระต่อกัน และเป็นตัวแปรสุ่มแบบต่อเนื่องที่มีค่า Median เท่ากับ 0

2. $\lim_{n\to\infty} \frac{X'X}{n} = Q$ เป็น Positive Definite Matrix

เราสามารถหา Assymptotic Distribution ของ $\hat{\beta}_m$ ได้ดังนี้

$$\sqrt{n}(\hat{\beta}_m - \beta) \underset{ass}{\sim} N(0, w^2 Q^{-1})$$ [1]

หรือ $\sqrt{n}(\hat{\beta}_{mj} - \beta_j) \underset{ass}{\sim} N(0, w^2 q_{jj})$ เมื่อ q_{jj} คือสมาชิก ณ ตำแหน่งที่ (j,j) ของเมตริกซ์ Q^{-1} ; $j=1,2,\ldots,k$

---

[1] $\underset{ass}{\sim}$ หมายถึง assymptotic distributed

และจากการศึกษาในลำดับต่อมาพบว่า

$$w = \{2f(0)\}^{-1}$$

เมื่อ $\hat{f}(0) = \frac{r - s - 1}{n(e_{(r)} - e_{(s)})}$ โดยที่ $r \simeq \frac{3n}{4}$; $s \simeq \frac{n}{4}$ และ $e_{(r)}$ และ $e_{(s)}$ คือ Ordered Statistics ตัวที่ r และ s จาก Squence

$$e_{(1)} < e_{(2)} < \ldots < e_{(r)} < \ldots < e_{(s)} < \ldots < e_{(n)}$$

ด้วยการอาศัยค่าประมาณของ f(o) เราจึงสามารถกำหนด Assymptotic Distribution ของ $\sqrt{n}\ (\hat{\beta}_m - \beta)$ ได้ดังนี้

$$\sqrt{n}\ (\hat{\beta}_m - \beta) \underset{ass}{\sim} N\{0, \tfrac{1}{4}[\hat{f}(o)]^{-2}(X'X)^{-1}\}$$

นั่นคือ $\sqrt{n}(\beta_{mj} - \beta_j) \underset{ass}{\sim} N\{0, \frac{1}{4}[\hat{f}(o)]^{-2} c_{jj}\}$ เมื่อ c_{jj} คือสมาชิก ณ ตำแหน่งที่ (j,j) ของเมตริกซ์ $(X'X)^{-1}$; $j=1,2,\ldots,k$

หรือ

$$\frac{\sqrt{n}\ (\hat{\beta}_{mj} - \beta_j)}{\frac{1}{2}[\hat{f}(o)]^{-1}\sqrt{c_{jj}}} \underset{ass}{\sim} N(0,1) \ ; \ j=1,2,\ldots,k$$

ด้วยความรู้เหล่านี้ เราจึงสามารถหา Confidence Interval ของ β_j รวมทั้งตรวจสอบนัยสำคัญของ β_j ได้โดยง่าย

5.4.2 Box - Cox transformation

บ๊อกซ์ และ ค๊อกซ์ (Box, G.E.P. and Cox, D.R., 1964)[1] ได้เสนอ Class of Transformation ขึ้นมาโดยสมมุติว่า เราสามารถคำนวณหาค่าตัวคงที่ λ จากสมการแปลงรูปต่อไปนี้ได้คือ[2]

$$\frac{Y_i^{\lambda} - 1}{\lambda} = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \ ; \ i=1,2,\ldots,n^{3}$$

---

[1] Judge, G.G. et. al., **Op. Cit.**, p.318

[2] ในทางปฏิบัติให้คำนวณหาค่า λ โดยวิธี Iteration

[3] ขอให้สังเกตว่า ถ้า $\lambda=1$ จะมีผลให้ $\frac{Y_i^{\lambda}-1}{\lambda} = Y_i$ และถ้า $\lambda \to 0$ จะพบว่า $\frac{Y_i^{\lambda}-1}{\lambda} \to \log Y_i$

ซึ่งการแปลงรูปตัวแปรตามลักษณะนี้จะมีผลให้ตัวแปรตาม มีคุณสมบัติสำคัญ 3 ประการคือ

1. $\frac{Y_i^{\lambda}-1}{\lambda}$ หรือเขียนย่อ ๆ ว่า $Y_i^{(\lambda)}$ จะมีการแจกแจงแบบปกติ

2. มีค่าความแปรปรวนคงที่ (Homoscedasticity)

3. $Y_i^{(\lambda)}$ เป็น Linear Function ของ พารามิเตอร์ β

หรือกล่าวโดยสรุปได้ว่า ถ้าแปลงรูปสมการ $Y_i = \beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}+u_i$ เป็นสมการใหม่คือ $\frac{Y_i^{\lambda}-1}{\lambda} = \beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}+u_i$ แล้วสมการใหม่นี้จะเป็นสมการที่ $u_i \sim N(0,\sigma^2)$ โดยประมาณ

ปัญหาก็คือ λ ควรมีค่าเท่าไรจึงจะเหมาะสม และมีผลให้ $u_i \sim N(0,\sigma^2)$

การคำนวณหาค่าของ λ เรากระทำได้โดยวิธี Maximum Likelihood Method (ML) ผสมกับ Iteration Method กล่าวคือค่อย ๆ เปลี่ยนค่า λ ไปทีละน้อยโดย $(0<\lambda\leq 1)$ ในแต่ละครั้งที่เปลี่ยน λ ไปก็ให้คำนวณหา $\hat{\beta}$ และ $\hat{\sigma}^2$ (ความจริงต้องเขียนว่า $\hat{\beta}(\lambda)$ และ $\hat{\sigma}^2(\lambda)$ เพราะ $\hat{\beta}$ และ $\hat{\sigma}^2$ มีค่าผันแปรไปตาม λ) ทุกครั้ง λ ค่าใดที่มีผลให้ $\hat{\sigma}^2$ มีค่าต่ำสุดก็ถือว่า λ ค่านั้นคือ MLE ของ λ ให้ใช้ λ ดังกล่าวในสมการแปลงรูปคือ

$$\frac{Y_i^{\lambda}-1}{\lambda} = \beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}+u_i \quad ; \; i=1,2,\ldots,n$$

แล้ววิเคราะห์สมการถดถอยตามปกติต่อไป

เทคนิคในการหา MLE ของ λ ซึ่งจะสรุปเป็นอัลกอริทึมไว้ในตอนท้ายมีที่มาซึ่งสามารถพิสูจน์ให้เห็นเป็นขั้น ๆ ดังนี้

1. ให้สมการ

$$Y_i = \beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}+u_i \quad ; \; i=1,2,\ldots,n$$

ซึ่งจากการทดลองนำไปวิเคราะห์โดยอาศัยค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i}, \ldots, X_{ki})$; $i=1,2,\ldots,n$ และทดสอบแล้วพบว่า u_i มิได้มีการแจกแจงแบบปกติ

เมื่อทดลองนำ Geometric Mean ของ $Y_1, Y_2, \ldots, Y_n$ คือ $G = \sqrt[n]{Y_1 Y_2 \ldots Y_n}$

ไปหาร Y_i ; $i=1,2,\ldots,n$ ได้ $Y_{i*} = \frac{Y_i}{G}$; $i=1,2,\ldots,n$

จะพบว่า
$$\sum_i^n \ln Y_{i*} = \sum_i^n \ln\left(\frac{Y_i}{G}\right) = \sum_i^n \ln Y_i - \sum_i^n \ln G$$

$$= \sum_i^n \ln Y_i - \sum_i^n \ln \sqrt[n]{Y_1 Y_2 \ldots Y_n}$$

$$= \sum_i^n \ln Y_i - \sum_i^n \frac{1}{n}\left(\sum_i^n \ln Y_i\right)$$

$$= \sum_i^n \ln Y_i - \frac{n}{n}\sum_i^n \ln Y_i = 0$$

แสดงว่าค่าสังเกต $Z_i = \left(\frac{Y_i}{G}, X_{2i}, X_{3i}, \ldots, X_{ki}\right)$; $i=1,2,\ldots,n$ นั้น $\sum_i^n \ln\left(\frac{Y_i}{G}\right) = 0$ ซึ่งความจริงข้อนี้จะนำไปใช้ประโยชน์ในขั้นที่ 2 ต่อไป

2. จากสมการ $\frac{Y_{i*}^{\lambda}-1}{\lambda} = \beta_1+\beta_2 X_{2i}+\beta_3 X_{3i}+\ldots+\beta_k X_{ki}+u_i$; $i=1,2,\ldots,n$

ซึ่งบ๊อกซ์ และ ค๊อกซ์ ถือว่าในกรณีเช่นนี้ $u_i \sim N(0,\sigma^2)$ ผลที่ตามมาก็คือ

$$Y_{i*}^{(\lambda)} \sim N(\beta_1+\beta_2 X_{2i}+\ldots+\beta_k X_{ki}, \sigma^2)$$

ดังนั้น Likelihood Function ของ $Y_{i*}^{(\lambda)}$ คือ $L = \prod_{i.}^{n} f(Y_{i*}^{(\lambda)})$

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_i^n \left(Y_{i*}^{(\lambda)}-\beta_1-\beta_2 X_{2i}-\ldots-\beta_k X_{ki}\right)^2\right\} . J$$

เมื่อ $J = \begin{bmatrix} Y_{1*}^{(\lambda)}, & Y_{2*}^{(\lambda)}, & \ldots, & Y_{n*}^{(\lambda)} \\ Y_{1*}, & Y_{2*}, & \ldots, & Y_{n*} \end{bmatrix}$ คือ Jakobian of Transformation ที่แปลงรูปจาก $Y_{i*}^{(\lambda)}$ สู่ Y_{i*} ; $i=1,2,\ldots,n$

ดังนั้น
$$J = \begin{vmatrix} \frac{\partial Y_{1*}^{(\lambda)}}{\partial Y_{1*}} & \frac{\partial Y_{2*}^{(\lambda)}}{\partial Y_{1*}} & \cdots & \frac{\partial Y_{n*}^{(\lambda)}}{\partial Y_{1*}} \\ \frac{\partial Y_{1*}^{(\lambda)}}{\partial Y_{2*}} & \frac{\partial Y_{2*}^{(\lambda)}}{\partial Y_{2*}} & \cdots & \frac{\partial Y_{n*}^{(\lambda)}}{\partial Y_{2*}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial Y_{1*}^{(\lambda)}}{\partial Y_{n*}} & \frac{\partial Y_{2*}^{(\lambda)}}{\partial Y_{n*}} & \cdots & \frac{\partial Y_{n*}^{(\lambda)}}{\partial Y_{n*}} \end{vmatrix} = \begin{vmatrix} Y_{1*}^{\lambda-1} & & & \\ & Y_{2*}^{\lambda-1} & & \\ & & \ddots & \\ & & & Y_{n*}^{\lambda-1} \end{vmatrix}$$

$$= Y_{1*}^{\lambda-1} \; Y_{2*}^{\lambda-1} \; \ldots \; Y_{n*}^{\lambda-1}$$

$$= \prod_{i}^{n} Y_{i*}^{\lambda-1}$$

นั่นคือ

$$L = \left(\frac{1}{2\pi\sigma}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i}^{n}\left(Y_{i*}^{(\lambda)}-\beta_1-\beta_2X_{2i}-\ldots-\beta_kX_{ki}\right)^2\right\} \prod_{i}^{n} Y_{i*}^{\lambda-1}$$

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma - \frac{1}{2\sigma^2}\sum_{i}^{n}\left(Y_{i*}^{(\lambda)}-\beta_1-\beta_2X_{2i}-\ldots-\beta_kX_{ki}\right)^2 + (\lambda-1)\sum_{i}^{n}\ln Y_{i*}$$

และเราทราบมาจากขั้นที่ 1 ว่า $\sum_{i}^{n} \ln Y_{i*} = 0$ ดังนั้น[1]

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma - \frac{1}{2\sigma^2}\sum_{i}^{n}(Y_{i*}^{(\lambda)} - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2$$

หรือเสนอในรูปของเมตริกซ์ได้ดังนี้

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma - \frac{(Y_*^{(\lambda)} - X\beta)'(Y_*^{(\lambda)} - X\beta)}{2\sigma^2}$$

MLE ของ $\beta(\lambda)$ และ $\sigma^2(\lambda)$ สามารถหาได้จากสมการ $\frac{\partial}{\partial\beta}\ln L = 0$ และ $\frac{\partial}{\partial\sigma}\ln L = 0$ และพบว่า[2]

$$\hat{\beta}(\lambda) = (X'X)^{-1}X'Y_*^{(\lambda)}$$

$$\hat{\sigma}^2(\lambda) = \frac{1}{n}(Y_*^{(\lambda)} - X\hat{\beta}(\lambda))'(Y_*^{(\lambda)} - X\hat{\beta}(\lambda))$$

จึงเห็นได้ว่าเราสามารถคำนวณหาค่า λ ที่เหมาะสมได้จากอัลกอริทึมต่อไปนี้

(1) จากค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i}, X_i, \ldots, X_{ki})$; $i=1,2,\ldots,n$ ให้นำ Geometric Mean ของ $Y_1, Y_2, \ldots, Y_n$ คือ $G = \sqrt[n]{Y_1 Y_2 \ldots Y_n}$ หารค่าของ Y ทุกค่าได้ $Y_{i*} = \frac{Y_i}{G}$; $i=1,2,\ldots,n$

ดังนั้นค่าสังเกตที่ได้คือ $Z_i = (Y_{i*}, X_{2i}, X_{3i}, \ldots, X_{ki})$; $i=1,2,\ldots,n$

(2) กำหนดให้ $0<\lambda<1$ โดยค่อย ๆ แปรค่า λ ไปทีละน้อย และในทุกค่าของ λ ดังกล่าวให้นำไปใช้ในการวิเคราะห์สมการถดถอย

---

[1] ขอให้สังเกตว่าการแปลงค่าจาก Y_i เป็น $Y_{i*} = \frac{Y_i}{G}$ ซึ่งมีผลให้ $\sum_{i}^{n}\ln Y_{i*}$ นั้น เรามีจุดมุ่งหมายเพื่อปรับรูปของ L ได้ง่าย

[2] เรียกว่า Conditional ML Estimator ของ β และ σ^2 เพราะค่าของ $\hat{\beta}$ และ $\hat{\sigma}^2$ จะผันแปรค่าไปตามค่าของ λ

$$Y_{i*}^{(\lambda)} = \frac{Y_{i*}^{\lambda}-1}{\lambda} = \beta_1+\beta_2 X_{2i}+\ldots+\beta_k X_{ki} \quad ; \; i=1,2,\ldots,n$$

โดยใช้สูตร

$$\hat{\beta}(\lambda) = (X'X)^{-1}X'Y_*^{(\lambda)} \qquad \hat{\sigma}^2(\lambda) = \frac{(Y_*^{(\lambda)}-X\hat{\beta}(\lambda))'(Y_*^{(\lambda)}-X\hat{\beta}(\lambda))}{n}$$

(3) เลือกค่า λ ที่มีผลให้ $\hat{\sigma}^2(\lambda)$ ต่ำที่สุดไปใช้ประโยชน์ต่อไป ค่า λ ที่เลือกนี้ก็คือ MLE ของ λ นั่นเอง

สำหรับการวิเคราะห์สมการถดถอยเพื่อแก้ปัญหาตามวิธีของบ๊อกซ์ และค๊อกซ์ เพื่อแก้ปัญหา Nonnormal Error สามารถสรุปวิธีดำเนินการได้ดังนี้

(1) คำนวณค่าที่เหมาะสมของ λ จากอัลกอริทึมข้อ (1) - (3)

(2) นำค่า $\hat{\lambda}$ ที่ได้ไปแปลงรูปตัวแปรตามจาก Y_i เป็น $Y_i^{(\hat{\lambda})} = \frac{Y_i^{(\hat{\lambda})}-1}{\hat{\lambda}}$ ได้ค่าสังเกต Z ดังนี้คือ $Z_i = (Y_i^{(\hat{\lambda})}, X_{2i}, X_{3i}, \ldots, X_{ki})$; $i=1,2,\ldots,n$

(3) วิเคราะห์สมการถดถอยตามวิธี OLS โดยอาศัยค่าสังเกต Z ในข้อ (2)

5.5 สรุป และ ข้อเสนอแนะ

ภาระกิจของนักวิจัยที่ดำเนินการวิจัยโดยอาศัยเทคนิคการวิเคราะห์ความถดถอยนั้นนอกเหนือไปจากการประมาณค่าเวคเตอร์ β จากสมการ $Y=X\beta+U$ รวมถึงการตรวจสอบนัยสำคัญของ β แล้วยังจำเป็นต้องตรวจสอบดูว่าข้อตกลงต่าง ๆ ของสมการถดถอยจะยังคงเป็นจริงหรือไม่ เพราะถ้าตรวจพบว่าไม่เป็นจริง และเรายังคงใช้ $\hat{\beta} = (X'X)^{-1}X'Y$ ตามเดิม ค่า $\hat{\beta}$ นี้จะเป็นตัวประมาณค่าที่ไม่มีคุณภาพ สำหรับปัญหา Nonnormal Error นี้ ถ้านักวิจัยตรวจพบว่า u มิได้มีการแจกแจงแบบปกตินักวิจัยถึงปฏิบัติดังนี้

1. ถ้าเราสามารถทราบเค้ามูลใด ๆ (Prior Information) ที่ยืนยันว่า u มีการแจกแจงแบบใดแบบหนึ่งที่แน่นอน ให้แก้ปัญหานี้ด้วยการประมาณค่า ด้วยวิธี ML

2. ถ้าไม่อาจทราบว่า u มีการแจกแจงแบบใด ให้ตัดสินใจใช้ OLS แต่ทั้งนี้จะกระทำได้เฉพาะในกรณีที่มีจำนวน Outlier ไม่สูงนัก[1]

---

[1] Outlier คือ residual ที่มีขนาด (magnitude) สูงกว่า Residual อื่น ๆ 3 SD ถึง 4 SD หมายความว่า เมื่อนำค่า Residual e_i มาหาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานพร้อมทั้งสร้างช่วงเชื่อมั่น 99% แล้ว Residual ตัวใดตกอยู่นอก 99% Confidence Band เราเรียกว่า Outlier

2. ถ้ามี Outlier ในอัตราที่สูงมากให้เลือกใช้ ROBUST Estimator วิธีใดวิธีหนึ่งที่กล่าวมาแล้ว

4. ถ้าทดสอบด้วย Skewness แล้วพบว่าตัวแปรสุ่ม u มาจากการแจกแจงที่เป็น Skew Distribution ให้แก้ปัญหาด้วย Box - Cox Transformation

5. ถ้ามีความจำเป็นต้องกำหนดฟังก์ชั่นของตัวแปร เช่น ในกรณีของ Box - Cox Transformation ($Y_i^{(\lambda)} = \frac{Y_i^{\lambda} - 1}{\lambda}$) นักวิจัยสมควรกำหนดฟังก์ชั่นของ u ด้วย กรณีของ Box - Cox Transformation นั้นกำหนดเฉพาะฟังก์ชั่นของ Y โดยไม่สนใจ u จึงนับว่ามีปัญหาอยู่