บทที่ 10

Error in Variables

10.1 เหตุผล ความหมายและความเหมาะสมของข้อตกลง

คำว่า Error in Variables หมายถึงสถานการณ์ที่ค่าสังเกตหรือค่าที่เกิดจากการวัดของตัวแปรอิสระ X's มีข้อบกพร่องหรือ Error ผนวกอยู่ ซึ่งเป็นเหตุการณ์ที่ขัดแย้งกับข้อตกลงสมการถดถอยที่กำหนดไว้ว่า ตัวแปรอิสระต้องไม่มี Error ทั้งจากการวัดหรือจากการสังเกต กล่าวคือ $E(u_iX_j) = 0,\ i = 1,2,\ldots,n;\ j = 2,3,\ldots,k$

พิจารณา $E(u_iX_j)$ ถ้าหากจะมองกันในแง่ของความหมายทางสถิติและวัตถุประสงค์ จะพบว่า $E(u_iX_j)$ มีความหมายดังนี้

1. $E(u_iX_j)$ หมายความว่า เมื่อกำหนดให้ X_j มีค่าคงที่ ณ $X_j = X_{jo}$ การแจกแจงของตัวแปรสุ่ม u_i ณ จุดที่ทำการทดลองซ้ำ ๆ ที่ $X_j = X_{jo}$ จะมีลักษณะหนึ่ง เมื่อเปลี่ยนค่า X_j ไปซ้ำ (Replicate) ณ. จุดอื่นการแจกแจงของ u_i ก็จะมีลักษณะที่เป็นตัวของตัวเอง อาจมีลักษณะเดิมหรือเปลี่ยนลักษณะไปแต่ทั้งนี้ย่อมเป็นไปเพราะธรรมชาติของ u_i เองมิใช่เพราะได้รับอิทธิพลจากการที่ X_j เปลี่ยนแปลงค่า (เพิ่มขึ้น-ลดลง) แต่ประการใด และเนื่องจาก u คือแหล่งรวมของ Error ทั้งปวง อีกทั้ง u ยังมีการแจกแจงที่เป็นอิสระกับตัวแปร X's ก็แสดงว่า X's ต้องไม่ข้องเกี่ยวกับ Error ใด ๆ หรือ X's ต้องไม่มี Error ปะปนอยู่เลยทั้งในแง่ของการวัดค่าหรือการบันทึกค่าจากการสังเกต

2. X's ต้องเป็นปัจจัยภายนอกเท่านั้นจะมีฐานะเป็นปัจจัยภายในเช่นเดียวกับตัวแปรตามคือ Y มิได้ หรือนัยหนึ่งจะต้องมีความสัมพันธ์ระหว่าง Y กับ X's ในทิศทางเดียว (One-way causation) คือ $Y = f(X\text{'s})$ เท่านั้น จะเป็น $X = f(Y)$ มิได้

จากวัตถุประสงค์ข้อที่ 1 ผนวกกับข้อที่ 2 จะเห็นได้ว่า การบังคับให้ $E(u_iX_j) = 0$ หมายความว่า เมื่อทำการทดลองซ้ำ ณ จุด $X_j = X_{jo}$ ซึ่งมีผลให้ปรากฏตัวแปรสุ่ม u ในจำนวนเท่ากับจำนวนซ้ำ เช่น ทำการทดลองซ้ำ ณ. $X_j = X_{jo}$ ทั้งสิ้น r ครั้งจะปรากฏ u ณ จุดนี้ r ตัว ซึ่ง$(u_1,u_2,\ldots,u_r)$เหล่านี้จะเป็นตัวแปรสุ่มที่มี Sampling Distribution เป็น f(u) ส่งผลให้เกิด f(Y) เมื่อเลื่อนจุดซ้ำไป ก็จะเกิด f(u) ชุดใหม่และ f(Y) ชุดใหม่ ดังนี้เรื่อย ๆ ไป สมการที่พุ่งผ่านไปในจุดอันเป็นค่าเฉลี่ยของ Y|X หรือ Locus ของ $\mu_{Y/X}$ ก็คือสมการ $Y = f(X\text{'s})$ แสดงว่า การบังคับให้ $E(u_iX_j) = 0$ มีผลให้เกิดสมการ $Y = f(X\text{'s})$ เพียงลักษณะเดียวเท่านั้น จะเป็น $X = f(Y)$ มิได้

3. ตัวแปรอิสระที่ไม่ปรากฏอยู่ในสมการซึ่งอาจเพราะผู้วิจัยละเลย คาดคิดไม่ถึง หรือคัดทิ้งเนื่องจากเป็นตัวแปรที่ไม่อาจวัดค่าได้ จะต้องไม่เกี่ยวข้องกับตัวแปรอิสระ X's ในสมการ $Y = f(X's)$ แต่เนื่องจากอิทธิพลของตัวแปรดังกล่าวแม้จะมิได้ปรากฏร่วมเป็นตัวแปรอิสระ แต่ก็ยังคงส่งอิทธิพลต่อ Y อยู่ภายนอกซึ่งเราสามารถผนวกปัจจัยดังกล่าวไว้ในตัวแปรสุ่ม u การควบคุมให้ $E(u_i X_j) = 0$ จึงมีผลเสมือนควบคุมให้ตัวแปรอิสระทั้งที่ปรากฏในสมการและนอกสมการเป็นอิสระต่อกันไปในโอกาสเดียวกัน

อย่างไรก็ตามในทางปฏิบัติเรามักพบปัญหาที่เป็นสาเหตุให้ข้อตกลงว่า $E(u_i X_j) = 0$ ไม่เป็นจริงได้เสมอเพราะ X's จะมี Error ในรูปใดรูปหนึ่งคือ Error of Measurement หรือ Error of Observation ผนวกอยู่เสมอตามสถานการณ์ต่าง ๆ ต่อไปนี้คือ

1. ข้อมูลที่เก็บไว้ในแหล่งต่าง ๆ ซึ่งเป็นแหล่งทุติยภูมิ ส่วนใหญ่รวบรวมขึ้นจากงานสำรวจด้วยกลุ่มตัวอย่าง ในกระบวนการสุ่มตัวอย่างนั้นแม้ผู้รวบรวมข้อมูลจะควบคุมงานสำรวจได้ดีเพียงใด ความผิดพลาดก็เกิดขึ้นได้เสมอในหลายขั้นตอน เริ่มตั้งแต่ขั้นวางแผนเตรียมการ ขั้นปฏิบัติงานสนามเรื่อยไปจนถึงขั้นประมวลผลและเสนอผล[1] ข้อผิดพลาดอาจเกิดขึ้นจากผู้ให้ข้อมูลจงใจให้ข้อมูลผิดพลาด ผู้บันทึกข้อมูลบันทึกผิดพลาด แบบสำรวจขาดความถี่ถ้วนและไม่เจาะปัญหาให้ลึกลงไปในทุกด้าน ผู้ประมวลผลทำงานบกพร่องรวมตลอดถึงการปัดเศษตัวเลข และปรับค่าข้อมูลให้สอดคล้องกัน จะเห็นได้ว่าความผิดพลาดนั้นปรากฏอยู่แล้วเป็นปกติ และโดยปรัชญาของการทำงานสำรวจ เรายอมรับกันโดยทั่วไปแล้วว่า ความผิดพลาดจะต้องมีอยู่ไม่ระดับใดก็ระดับหนึ่ง ด้วยเหตุนี้ทั้งตัวแปร Y และ X's จึงผนวกเอาความผิดพลาดเหล่านี้ไว้ในตัวอยู่แล้วเป็นปกติ

2. ในบางสถานการณ์เราไม่อาจนำตัวแปรที่เป็นปัจจัยที่แท้จริงในการควบคุมความเคลื่อนไหวของ Y มาใช้ได้เพราะไม่อาจวัดค่าได้ทำให้ต้องเลือกใช้ตัวแปรทดแทน (Proxy Variable หรือ Proxy Index) มาใช้แทนเช่น การศึกษา (X) เป็นตัวแปรที่สำคัญที่สามารถควบคุมความเคลื่อนไหวของรายได้ (Y) แต่เนื่องจากการศึกษาเป็นตัวแปรที่วัดค่าได้ยากเพราะการศึกษาเป็นกระบวนการที่ต่อเนื่องและถี่ถ้วนลึกซึ้งเกินกว่าที่จะประเมินค่าเป็นตัวเลขได้ทำให้เราต้องใช้จำนวนปีที่ได้รับการศึกษาในสถาบันการศึกษามาใช้แทน ซึ่งก็พอจะแทนได้แต่ไม่อาจแทนได้ทั้งหมด ด้วยเหตุนี้เมื่อนำตัวแปร "จำนวนปีที่ได้รับการศึกษาในสถาบันการศึกษา" มาใช้แทน

---

[1] อ่าน มนตรี พิริยะกุล **เทคนิคการสำรวจด้วยกลุ่มตัวอย่าง** (บริษัท ประชาชน จำกัด กรุงเทพมหานคร 2525) บทที่ 1 หน้า 1-26

ตัวแปร "การศึกษา" ความผิดพลาดย่อมเกิดขึ้น

3. การใช้ตัวแปรดัมมี่เป็นตัวแปรอิสระจะมีผลให้เกิดความผิดพลาดขึ้นเสมอ ทั้งนี้เพราะตัวแปรดัมมี่มีค่าได้เพียง 2 ค่าคือ 0, 1 ขณะที่ตัวแปรที่แท้จริงนั้นสามารถมีค่าได้ถี่กว่านี้ เมื่อจำเป็นต้องใช้สเกลที่ค่อนข้างหยาบมาเป็นมาตรการวัดค่าตัวแปรที่มีสเกลถี่กว่า ความผิดพลาดจึงปรากฏขึ้นอย่างแน่นอนและไม่อาจเลี่ยงได้ เช่น ให้ X_3= รสนิยม เมื่อพิจารณา X_3จะพบว่าค่าของ X_3ค่อนข้างละเอียดเพราะเป็นตัวแปรที่ต้องอาศัยหลักเกณฑ์เชิงอัตตนิยมมาใช้ประเมินค่า เราไม่อาจสรุปได้ตรงกันว่านาย ก. มีรสนิยมดีเลวเพียงใด ทั้งนี้ขึ้นอยู่กับผู้พิจารณาว่ามีภูมิหลังแตกต่างกันเพียงใด ผู้พิจารณามีภูมิหลังต่างกันจะมองนาย ก. ต่างกัน ดังนี้เป็นต้น แต่เมื่อจำเป็นต้องมีตัวแปร X_3ไว้ในสมการ เราก็จำเป็นต้องสร้างค่าของ X_3ให้ได้ซึ่งกระทำได้โดยกำหนดให้ X_3 เป็นตัวแปรดัมมี่อาจให้ $X_3 = 1$ ถ้ารสนิยมดี $X_3 = 0$ ถ้ารสนิยมไม่ดี ดังนี้เป็นต้น ขอให้สังเกตว่าการกำหนดให้ X_3 เป็นตัวแปรดัมมี่แสดงว่าเรายอมให้ X_3 มีค่าหยาบกว่าความเป็นจริง ความผิดพลาดในค่าของ X_3จึงปรากฏขึ้น

4. การใช้เลขดัชนีเป็นค่าของตัวแปรแทนที่จะใช้ค่าที่แท้จริงของตัวแปรมีผลให้เกิดความผิดพลาดได้ ทั้งนี้เพราะเลขดัชนีเป็นตัวเลขที่ผนวกเอา Error ไว้ในตัวเองมากอยู่แล้วโดยธรรมชาติ กล่าวคือ การคำนวณเลขดัชนีต้องอาศัยการกำหนดปีฐาน ซึ่งหากกำหนดปีฐานไม่ถูกต้องเลขดัชนีก็มีความบกพร่อง ประการต่อมาเลขดัชนีต้องอาศัยปริมาณ (Quantity) และราคา (Price)[1] และอื่น ๆ ซึ่งต้องอาศัยผลจากการสำรวจความผิดพลาดในข้อมูลเหล่านี้ย่อมปรากฏอยู่แล้วเป็นปกติ และประการสุดท้ายเลขดัชนีเป็นค่าประมาณที่เรานิยมปัดเศษทิ้ง ความผิดพลาดของข้อมูลเพราะการปัดเศษจึง ปรากฏอยู่ เมื่อนำเลขดัชนีมาใช้เป็นค่าของตัวแปรอิสระหรือตัวแปรตาม ตัวแปรเหล่านี้จึงผนวกเอาความผิดพลาดไว้ในตัวอย่างไม่อาจเลี่ยงได้

10.2 **ผลสะท้อนของ** Error in Variables

เมื่อเกิดปัญหา Error in Variable คือ $E(u_i X_j) \neq 0;\ i = 1,2,\ldots,n\ ;\ j = 2, 3,\ldots,k$ จะเกิดปัญหาทางสถิติ 2 ประการคือ

1. $\hat{\beta}$ เป็น Biased Estimator คือ $\hat{\beta}$ ประมาณ β ได้ต่ำกว่าความเป็นจริง

2. $\hat{\beta}$ เป็น Inconsistent Estimator กล่าวคือ แม้จะใช้กลุ่มตัวอย่าง n ให้มีขนาดใหญ่เพียงใดก็ตาม ปริมาณความเอียงเฉ (Bias) ก็จะยังคงปรากฏอยู่มิได้หมดไปเช่นในสถานการณ์ของ

---

[1] อ่านสุทธิชัย โง้วศิริ, **สถิติเบื้องต้น** (โรงพิมพ์มหาวิทยาลัยรามคำแหง กรุงเทพมหานคร, 2523) หน้า 258-284

$E(u_i X_j) = 0$

ในที่นี้จะพิสูจน์ให้เห็นเฉพาะกรณีของ Simple Regression เท่านั้น กรณีของ Multiple Regression ก็คงพิสูจน์ได้โดยนัยเดียวกัน ผู้สนใจสามารถติดตามศึกษาได้จากเอกสารอ้างอิง

ให้ X_i และ Y_i เป็นค่าสังเกตของตัวแปร X และ Y ที่มีค่าจริง ณ วาระที่ i เป็น X_i^* และ Y_i^* ตามลำดับ เมื่อ X_i และ Y_i เป็นตัวแปรที่มี Measurement Error แสดงว่า

$$X_i = X_i^* + u_i$$

$$Y_i = Y_i^* + v_i$$

โดยที่ u_i และ v_i คือ Error of Measurement หรือ Error of Observation

สมมุติว่าตัวแปร Y และ X มีรูปแบบความสัมพันธ์ที่ถูกต้องแท้จริง (Exact Relation) ดังนี้คือ

$$Y_i^* = \alpha + \beta X_i^* \qquad \ldots\ldots(1)$$

เมื่อแทนที่ X_i^* ด้วย $X_i - u_i$ และแทนที่ Y_i^* ด้วย $Y_i - v_i$ จะพบว่า

$$Y_i - v_i = \alpha + \beta(X_i - u_i)$$

หรือ

$$Y_i = \alpha + \beta X_i + (v_i - \beta u_i)$$

ให้ $w_i = (v_i - \beta u_i)$

$$Y_i = a + \beta X_i + w_i \qquad \ldots\ldots(2)$$

พิจารณา Stochastic Term w_i ในสมการที่ (2) ถ้าเรายังคงข้อตกลงเดิมเอาไว้คือ $E(u_i v_i) = 0$ $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2$, $E(v_i) = 0$, $E(v_i^2) = \sigma_v^2$, $E(u_i X_i^*) = 0$, $E(v_i Y_i^*) = 0$ จะพบว่า $E(w_i X_i) \neq 0$ ตามข้อตกลงเดิม กล่าวคือ

$$E[X_i w_i] = E[X_i - E(X_i)]w_i = E[u_i(v_i - \beta u_i)]$$

$$= E[u_i v_i - \beta u_i^2]$$

$$= -\beta\sigma_u^2 \neq 0$$

ซึ่งแสดงว่าเมื่อเกิด Error in Variable ซึ่งเป็นเหตุการณ์ปกติของงานวิจัยทั่วไปจะพบว่า $Cov(X_i w_i) = -\beta\sigma_u^2 \neq 0$ ซึ่งขัดแย้งกับข้อตกลงเดิม และเมื่อพิจารณาตัวประมาณค่า $\hat{\beta}$ จะพบว่า $\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$ เป็นตัวประมาณค่าที่มีอคติ (Bias Estimator) กล่าวคือ

$$E(\hat{\beta}) = E[\frac{\Sigma x_i y_i}{\Sigma x_i^2}] = E\sum_i^n c_i Y_i \quad \text{เมื่อ} \quad c_i = \frac{x_i}{\Sigma x_i^2}$$

$$= E\sum_i^n c_i(\alpha + \beta X_i + w_i)$$

$$= \beta + E\sum_i^n c_i w_i \neq \beta$$

ปริมาณความเอียงเฉ (Bias) คือ $E(\hat{\beta}) - \beta = E\sum_i^n c_i w_i = E\frac{\Sigma(X_i - \bar{X})w_i}{\Sigma(X_i - \bar{X})^2}$

แต่โดยปกติปริมาณความเอียงเฉ (Bias) ควรหายไปถ้าเราขยายตัวอย่างให้มีขนาดใหญ่ขึ้น ซึ่งถ้าตัวประมาณค่าใดแม้จะเป็น Biased Estimator แต่เมื่อ n มีค่าสูงขึ้นคือ $n \to \infty$ แล้ว ปริมาณความเอียงเฉหายไป เราเรียกตัวประมาณค่านั้นว่า Consistent Estimator หรือ Assymptotic Unbiased Estimator กล่าวคือถ้า $\lim_{n \to \infty} Pr\{|\hat{\theta} - \theta| < \varepsilon\} = 1$ แสดงว่า $\hat{\theta}$ เป็น Consistent Estimator ของ θ

พิจารณา $\hat{\beta}$ ในกรณีมี Error in Variable จะพบว่า $\hat{\beta}$ เป็น Biased Estimator ของ β โดยมีปริมาณความเอียงเฉเท่ากับ $E\left[\frac{\Sigma(X_i - \bar{X})w_i}{\Sigma(X_i - \bar{X})^2}\right]$

พิจารณาปริมาณความเอียงเฉ $E(\hat{\beta} - \beta) = E\left[\frac{\Sigma(X_i - \bar{X})w_i}{\Sigma(X_i - \bar{X})^2}\right]$ จะพบว่า

$$\lim_{n\to\infty} E\left[\frac{\Sigma(X_i-\bar{X})w_i}{\Sigma(X_i-\bar{X})^2}\right]^{1} = \lim_{n\to\infty} E\left[\frac{\Sigma(X_i-\bar{X})w_i/n}{\Sigma(X_i-\bar{X})^2/n}\right]$$

$$= \frac{\lim_{n\to\infty}[\Sigma E(X_i w_i)/n - \Sigma\bar{X}E(w_i)/n]^2}{\lim_{n\to\infty}\sum_{i}^{n}(X_i-\bar{X})^2/n}$$

$$= \frac{\lim_{n\to\infty}[\Sigma(-\beta\sigma_u^2)/n-0]}{\lim_{n\to\infty}\Sigma(X_i-\bar{X})^2/n}$$

$$= \frac{\lim_{n\to\infty}(-\beta\sigma_u^2)}{\sigma_x^2} \text{ เมื่อ } \sigma_x^2 = \lim_{n\to\infty}\Sigma(X_i-\bar{X})^2/n$$

$$= -\frac{\beta\sigma_u^2}{\sigma_x^2}$$

นั่นคือ $E(\hat{\beta}-\beta) = -\frac{\beta\sigma_u^2}{\sigma_x^2}$ หรือ $E(\hat{\beta}) = \beta - \frac{\beta\sigma_u^2}{\sigma_x^2} < \beta$ แสดงว่าเมื่อเกิดปัญหา Error in Value $\hat{\beta}$ จะเป็น Downward Biased Estimator ของ β และถึงแม้จะขยาย n ออกไปเพียงใดก็ตาม ($n \to \infty$) ปริมาณความเอียงเฉ (Bias) ก็มิได้หายไปซึ่งแสดงว่า $\hat{\beta}$ เป็น Inconsistent Estimator

หมายเหตุ สำหรับปัญหา Error in Variables นี้จะไม่นับเป็นปัญหา ถ้าผู้วิจัยหรือผู้วินิจฉัยสั่งการในกิจการต่าง ๆ ที่ต้องอาศัยข้อมูลที่มีอยู่ตามที่ส่วนราชการพิมพ์เผยแพร่ หรืออาศัยจากข้อมูลที่มีอยู่ตามแหล่งข้อมูลต่าง ๆ โดยมิได้สนใจหรือพาดพิงไปถึงค่าที่แท้จริงของตัวแปรนั้น ๆ หรือนัยหนึ่ง ปัญหา Error in Variable จะเกิดขึ้นก็ต่อเมื่อผู้วิจัยหรือผู้เกี่ยวข้องกับงานวิจัยหรืออาศัยผลจากงานวิจัยต้องเกี่ยวข้องอยู่กับค่าที่แท้จริงของตัวแปรและทราบว่าข้อมูลที่รวบรวมไว้ในแหล่งต่าง ๆ คลาดเคลื่อนไปจากความเป็นจริง ด้วยเหตุนี้ถ้างานวิจัยใดที่ต้องอาศัยข้อมูลจากแหล่งต่าง ๆ

---

[1] เรียกว่า Assymptotic Expectation

[2] $E(w_i) = E[v_i-\beta u_i] = E(v_i) - \beta E(u_i) = 0$

ที่พิมพ์เผยแพร่เอาไว้หรือบันทึกไว้ ผู้วิจัยก็ยังคงใช้ OLS ได้ตามปกติ หากจะพูดกันอย่างตรงไปตรงมาที่สุดก็คือ ถ้าเราไม่ทราบค่าที่แท้จริงของตัวแปร X's และ Y ปัญหา Error in Variable ก็ไม่เกิดขึ้น และในสถานการณ์ทางปฏิบัติก็นับว่าเป็นเรื่องยากที่จะทราบค่าที่แท้จริงของตัวแปรที่เกี่ยวข้องได้

อย่างไรก็ตาม การที่เราจะยึดเอา "ความไม่รู้" มาเป็นข้ออ้างเพื่อจะได้ไม่ต้องศึกษาปัญหาอันเนื่องมาจากการใช้ค่าที่ไม่ใช่ค่าที่แท้จริงของตัวแปรนั้นนับว่าเป็นเรื่องไม่สมควร ประเด็นต่าง ๆ ที่จะศึกษาในลำดับต่อไปนี้จะถือว่าผู้วิจัยกำลังเกี่ยวข้องกับค่าที่แท้จริงของตัวแปรและการแก้ปัญหาเมื่อข้อมูลที่มีอยู่มิใช่ค่าที่แท้จริงของตัวแปร

10.3 **การทดสอบปัญหา** Error in Variable (Regressors)

สำหรับปัญหา Error in Variable นับว่าเป็นปัญหาที่ยังไม่ค่อยมีการพัฒนาวิธีทดสอบกว้างขวางเท่าปัญหาอื่น ๆ วิธีหนึ่งที่พอใช้ทดสอบได้ก็คือ การตรวจสอบนัยสำคัญของสหสัมพันธ์ระหว่าง $|e_i|$ กับ X ซึ่งมีขั้นตอนดังนี้

1. วิเคราะห์สมการถดถอย $Y = f(X_j)$ โดยวิธี OLS และคำนวณหา Residual $e_i = Y_i - \hat{Y}_i$; $i = 1,2,\ldots,n$; $j = 2,3,\ldots,k$

2. คำนวณหาสหสัมพันธ์ระหว่าง $|e|$ กับ x_j ดังนี้

$$r = \frac{\Sigma|e_i|x_{ji}}{\sqrt{\Sigma e_i^2 \ \Sigma x_{ji}^2}} \quad ; j = 2,3,\ldots,k$$ [1]

3. ทดสอบสมมุติฐาน $H_0 : \rho = 0$ VS $H_1 : \rho \neq 0$

โดยเราจะปฏิเสธสมมุติฐานหลัก ณ ระดับนัยสำคัญ α เมื่อ $|t_c| > t_{n-2,1-\frac{\alpha}{2}}$ เมื่อ

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

---

[1] เหตุที่ต้องใช้ $|e_i|$ เพราะ $\Sigma e_i x_i = \Sigma(y_i-\hat{y}_i)x_i = \Sigma x_i y_i - \Sigma\hat{y}_i x_i = \Sigma x_i y_i - \hat{\beta}\Sigma x_i^2$
$= \Sigma x_i y_i - \frac{\Sigma x_i y_i}{\Sigma x_i^2} \cdot \Sigma x_i^2 = 0$

หรือนัยหนึ่งถ้าพบว่า $|t_c| > t_{n-2, 1-\frac{\alpha}{2}}$ เราย่อมสรุปได้ว่า X_j มีปัญหา Error in Variable คือ $E(u_i X_j) \neq 0$

10.4 การแก้ปัญหา Error in Variable

10.4.1 Inverse Least Square (ILS)

วิธี ILS เป็นวิธีที่ใช้ได้กับเฉพาะกรณีที่ X มี Error แต่ Y ไม่มี Error ขอให้สังเกตว่าวิธีนี้กลับกันกับเหตุการณ์ปกติที่สามารถใช้กับ OLS ที่เราถือว่า X ไม่มี Error แต่ Y มี Error) กล่าวคือถ้า χ คือค่าจริงของตัวแปรอิสระ X เมื่อค่าสังเกต X มี Error แสดงว่า

$$X_i = \chi_i + v_i \text{ หรือ } \chi_i = X_i - v_i$$

ดังนั้น ความสัมพันธ์จริงระหว่างตัวแปร Y และ X คือ

$$Y_i = \beta_0 + \beta_1 \chi_i \quad ; \quad i = 1,2,\ldots,n$$

$$Y_i = \beta_0 + \beta_1 (X_i - v_i) \qquad \ldots\ldots(1)$$

แต่เนื่องจาก Y_i ไม่มี Error ขณะที่ X_i มี Error ดังนั้น Y จึงทำหน้าที่เสมือนตัวแปรอิสระดังนี้

$$X_i = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} Y_i + v_i \quad ; \; i = 1,2,\ldots,n \qquad \ldots\ldots(2)$$

ให้ $-\frac{\beta_0}{\beta_1} = b_0$ และ $\frac{1}{\beta_1} = b_1$ สมการ (1) จึงเปลี่ยนรูปเป็น

$$X_i = b_0 + b_1 Y_i + v_i \quad ; \quad i = 1,2,\ldots,n \qquad \ldots\ldots(3)$$

จากสมการ (3) ขอให้สังเกตว่า $E(v_i Y_i) = 0$ เพราะในที่นี้ Y_i ไม่มี Error Y_i จึงเป็น Nonstochastic Variable การกลับรูปจาก Y = f(X) มาเป็น X = f(Y) จึงแก้ปัญหา Error in Variable ได้

สำหรับการประมาณค่า β_0 และ β_1 ให้กระทำโดยอาศัย OLS ดังนี้

$$\hat{b}_1 = \frac{\Sigma x_i y_i}{\Sigma y_i^2} \quad , \quad \hat{b}_0 = \bar{X} - \hat{b}_1 \bar{Y} \quad , \quad \hat{\sigma}^2 = \frac{1}{n-2} (\Sigma y_i^2 - \hat{b}_1 \Sigma x_i y_i)$$

ซึ่งค่าประมาณของ b_0 และ b_1 นี้เป็นค่าประมาณของสมการ (3) ซึ่งแปลงรูปมาจากสมการเดิม คือ สมการ (1) วิธีประมาณค่าพารามิเตอร์ β_0 และ β_1 ให้ย้อนสู่รูปเดิมโดยอาศัยการแก้สมการ

$$-\frac{\beta_0}{\beta_1} = \hat{b}_0 \quad \text{และ} \quad \frac{1}{\beta_1} = \hat{b}_1$$

และพบว่า

$$\hat{\beta}_1 = 1/\hat{b}_1 \quad , \quad \hat{\beta}_0 = -\hat{b}_0\hat{\beta}_1$$

เมื่อแทนค่า $\hat{\beta}_0$ และ $\hat{\beta}_1$ ลงในสมการ (1) เราจะได้สมการประมาณค่า $Y = \hat{\beta}_0 + \hat{\beta}_1 X$ ซึ่งไม่มีปัญหา Error in Variable ตามต้องการ

10.4.2 Two - Group Method

วิธีนี้เสนอโดยวาลด์ (Wald, A, 1940)[1] เพื่อใช้แก้ปัญหาเมื่อทั้งตัวแปร Y และ X ต่างก็มี Error

ในกรณีที่ทั้งตัวแปร X และ Y มี Error ด้วยกันทั้งคู่แสดงว่าค่าสังเกตของ Y และ X คลาดเคลื่อนจากค่าจริงดังนี้คือ

$X_i = \chi_i + v_i$ เมื่อ χ_i คือค่าจริงของ X_i

$Y_i = \psi_i + w_i$ เมื่อ ψ_i คือค่าจริงของ Y_i

โดยที่ความสัมพันธ์จริงของตัวแปร Y และ X คือ

$$\psi_i = \beta_0 + \beta_1\chi_i + u_i \ ; \quad i = 1,2,..,n \qquad ....(1)$$

เมื่อแทนที่ $\chi_i = X_i - v_i$ และ $\psi_i = Y_i - w_i$ ในสมการ (1) ความสัมพันธ์ระหว่างตัวแปร Y และ X จะปรากฏดังนี้

$$Y_i - w_i = \beta_0 + \beta_1 (X_i - v_i) + u_i$$

$$Y_i = \beta_0 + \beta_1 X_i + (u_i - \beta_1 v_i + w_i) \ ; \ i = 1,2,...,n \quad .....(2)$$

[1] Koutsoyiannis, A., Op. Cit, p.257

วาลด์เสนอแนะให้ประมาณค่า β_0 และ β_1 เป็นขั้น ๆ ดังนี้

1. ให้จัดเรียงลำดับค่าสังเกต $Z_i = (Y_i, X_i)$; $i = 1,2,...,n$ จากน้อยไปหามากตามขนาด (magnitude) ของตัวแปรอิสระ X

2. แบ่งตัวอย่างค่าสังเกตที่จัดเรียงแล้วในข้อ 1 เป็น 2 กลุ่มเท่า ๆ กัน ถ้า n เป็นเลขคี่ให้ตัด Median ของ Z (Central Observation) ทิ้ง แล้วคำนวณหาค่าเฉลี่ยของตัวแปร X และ Y ในตัวอย่างย่อยทั้งสองกลุ่มดังกล่าวคือ $(\bar{X}_1, \bar{Y}_1)$ สำหรับกลุ่มแรก และ $(\bar{X}_2, \bar{Y}_2)$ สำหรับกลุ่มที่ 2 พร้อมทั้งหาค่าเฉลี่ยของตัวแปร X และ Y จากกลุ่มตัวอย่างขนาด n คือ $(\bar{X}, \bar{Y})$

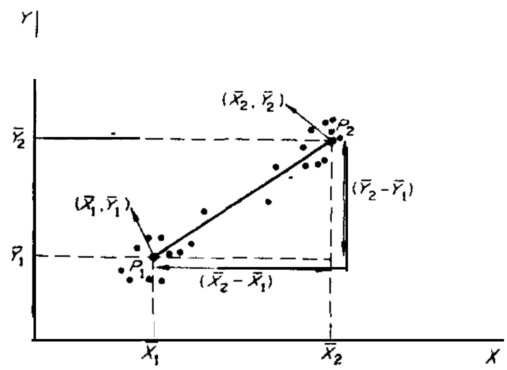
3. คำนวณค่าความชัน($\hat{\beta}_1$)โดยวิธี Two - Point Form ค่าความชัน($\hat{\beta}_1$)ก็คือความชันของเส้นตรงที่เชื่อมระหว่างจุด 2 จุดในระนาบคือ $(\bar{X}_1, \bar{Y}_1)$ กับ $(\bar{X}_2, \bar{Y}_2)$ พบว่า

$$\hat{\beta}_1 = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{X}_2 - \bar{X}_1}$$

4. คำนวณหาค่าประมาณของ β_0 โดยวิธี OLS พบว่า

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$

ลักษณะการกระจายตัวของค่าสังเกต Z ในกลุ่มตัวอย่างย่อยทั้งสองกลุ่มและที่ตั้งของจุด $(\bar{X}_1, \bar{Y}_1)$ กับ $(\bar{X}_2, \bar{Y}_2)$ ปรากฏดังภาพ



วิธีประมาณค่าของ β_0 และ β_1 ตามนัยนี้เป็นวิธีที่ง่าย และยังพบว่า $\hat{\beta}_0$ และ $\hat{\beta}_1$ เป็น Consistent Estimator แต่ไม่เป็น Best Estimator (ไม่มี Minimum Variance property)

10.4.3 Three - Group Method

วิธีนี้เสนอโดยบาร์ตเลตต์ (Bartlett, M.S., 1949)[1] ความจริงวิธีนี้ก็คือวิธีที่ขยายจำนวนตัวอย่างย่อยเพิ่มขึ้นจาก 2 กลุ่มตามวิธีของวาลด์มาเป็น 3 กลุ่ม เพื่อให้ $\hat{\beta}_0$ และ $\hat{\beta}_1$ มีคุณสมบัติทางสถิติสูงขึ้นกว่าเดิม กล่าวคือวิธีของ บาร์ตเลตต์ทำให้ $\hat{\beta}_0$ และ $\hat{\beta}_1$ เป็น Best Estimator นอกเหนือไปจากคุณสมบัติของ Unbiasness และ Consistent ตามวิธีของวาลด์

ขั้นตอนการประมาณค่าพารามิเตอร์ β_0 และ β_1 ของสมการ $Y_i = \beta_0 + \beta_1 X_i + (u_i - \beta_1 v_i + w_i)$; $i = 1,2,\ldots,n$ ปรากฏดังนี้

1. จัดเรียงลำดับค่าสังเกต $Z_i = (Y_i, X_i)$; $i = 1,2,\ldots,n$ จากน้อยไปหามากตามขนาดของค่าของตัวแปร X

2. จำแนกค่าสังเกต Z ในข้อ 1 ออกเป็น 3 ส่วน (ตัวอย่างย่อย) เท่า ๆ กันหรือเท่ากันโดยประมาณคือ n_1, n_2 และ n_3 ชุดตามลำดับโดยที่ $n_1 \cong n_2 \cong n_3$ และ $n_1 + n_2 + n_3 = n$

3. ตัดกลุ่มตัวอย่างชุดกลางขนาด n_2 (Central Observation of Size n_2) ทิ้งแล้วหาค่าเฉลี่ยของตัวแปร X และ Y ในกลุ่มตัวอย่างย่อย 2 กลุ่มที่เหลืออยู่คือ $(\bar{X}_1, \bar{Y}_1)$ และ $(\bar{X}_3, \bar{Y}_3)$ พร้อมทั้งหาค่าเฉลี่ยของตัวแปร X และ Y จากกลุ่มตัวอย่างขนาด n คือ $(\bar{X}, \bar{Y})$

4. คำนวณหาความชันของเส้นตรงที่เชื่อมจุด $(\bar{X}_1, \bar{Y}_1)$ และ $(\bar{X}_3, \bar{Y}_3)$ ตามวิธี Two-Point Form ได้ดังนี้

$$\hat{\beta}_1 = \frac{(\bar{Y}_3 - \bar{Y}_1)}{(\bar{X}_3 - \bar{X}_1)}$$

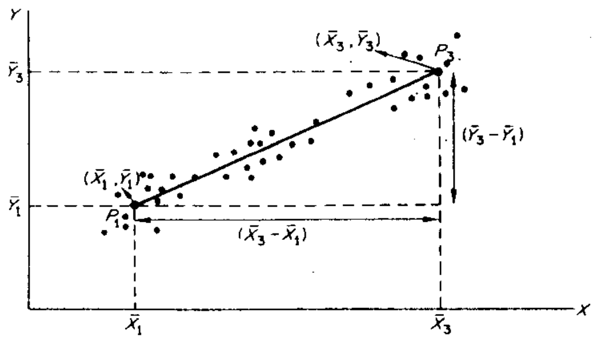
5. คำนวณหา $\hat{\beta}_0$ ตามวิธี OLS คือ

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

---

[1] Koutsoyiannis, A., Ibid., p.258

ผลการประมาณค่า β_0 และ β_1 ทำให้ได้สมการประมาณค่า $Y = \hat{\beta}_0 + \hat{\beta}_1 X$ ตามต้องการ ลักษณะการจัดจำแนกค่าสังเกตและทิศทางเส้นสมการปรากฏดังภาพ



นอกจากวิธีการทั้ง 3 ดังกล่าวแล้ว ยังมีวิธีประมาณค่าแบบอื่น ๆ อีกหลายวิธีเช่น วิธี Instrumental Variable วิธี Durbin's Rank[1] วิธี Two-Stage Least Square วิธี Ml วิธีของเบย์ วิธี Factor Analysis และวิธีอื่น ๆ ผู้สนใจสามารถติดตามศึกษาได้จาก Judge, G.G. et. al., Ibid., p.513-554

---

[1] วิธี Two - Group วิธี Three - Group วิธี Durbin's Rank เป็นกรณีเฉพาะของวิธี Instrumental Variable