ช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยเชิงเส้นอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปรสำหรับขนาดตัวอย่างเล็ก

Confidence Interval of Simple Linear Regression Coefficient with Errors in Variables for Small Sample Sizes

วิสุนีย์ ผักกาด[1], มานะชัย รอดชื่น[2*], พุฒิพงษ์ พุกกะมาน[2] และ บัณฑิตา พลับอินทร์[2]

Wisunee Puggard[1], Manachai Rodchuen[2*], Putipong Bookkamana[2] and Bandhita Plubin[2]

[1] บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่

[2] ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

[1] Graduate School, Chiang Mai University

[2] Department of Statistics, Faculty of Science, Chiang Mai University

Received : 11 May 2018

Accepted : 14 June 2018

Published online : 22 June 2018

บทคัดย่อ

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบช่วงความเชื่อมั่นสำหรับค่าสัมประสิทธิ์การถดถอย (β_1) ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปรสำหรับขนาดตัวอย่างเล็ก เมื่อทราบค่าความแปรปรวนของค่าคลาดเคลื่อนในตัวแปร X (σ_δ^2) วิธีที่พัฒนาขึ้นใหม่นี้ได้มาจากการปรับปรุงช่วงความเชื่อมั่นกำลังสองน้อยที่สุด และเปรียบเทียบกับช่วงความเชื่อมั่นเชิงเส้นกำกับ ช่วงความเชื่อมั่นแซนด์วิช โดยใช้วิธีการจำลองข้อมูลมอนติคาร์โล เพื่อเปรียบเทียบค่าประมาณความน่าจะเป็นคุ้มรวม และค่าความกว้างเฉลี่ยของทั้ง 3 วิธี ผลการศึกษาพบว่าเมื่อค่าอัตราส่วนความเชื่อถือได้ (k_ξ) มีค่าน้อยกว่า 0.3 ($k_\xi < 0.3$) พบว่าค่าประมาณความน่าจะเป็นคุ้มรวมของช่วงความเชื่อมั่นทั้ง 3 วิธี ไม่ใกล้เคียงกับระดับของความเชื่อมั่นที่กำหนด ส่วนกรณีที่ $0.3 \le k_\xi \le 0.7$ ช่วงความเชื่อมั่นกำลังสองน้อยที่สุดปรับปรุงเป็นช่วงความเชื่อมั่นที่มีค่าประมาณความน่าจะเป็นคุ้มรวมใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด และมีความกว้างเฉลี่ยแคบที่สุด ในขณะที่ $k_\xi > 0.7$ ช่วงความเชื่อมั่นแซนด์วิชจะเป็นช่วงความเชื่อมั่นที่มีค่าประมาณความน่าจะเป็นคุ้มรวมใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนดและมีความกว้างเฉลี่ยแคบที่สุด

คำสำคัญ : ช่วงความเชื่อมั่น, สัมประสิทธิ์การถดถอย, ความคลาดเคลื่อนในตัวแปร

*Corresponding author. E-mail : r.manachai@gmail.com

Abstract

This study aims to improve and compare the confidence intervals of simple linear regression coefficient (β_1) with errors in variables for small sample sizes when the variance of errors in X (σ_δ^2) is known. Improved ordinary least square confidence interval (IOLS) is the new developing method which improved from ordinary least square confidence interval. Comparing IOLS with asymptotic confidence interval (ACI) and sandwich confidence interval (SCI), a Monte-Carlo simulation is conducted to evaluate the performance of IOLS for comparison. The estimated coverage probability ($\hat{CP}$) and average lengths (AL) will be used as performance criteria. The simulation study indicates that when reliability ratio (k_ξ) less than 0.3 ($k_\xi < 0.3$), $\hat{CP}$ of three methods are not close to specified confidence coefficient. For $0.3 \le k_\xi \le 0.7$, $\hat{CP}$ of IOLS method is quite close to specified confidence coefficient and AL of IOLS method is the shortest. While $k_\xi > 0.7$, $\hat{CP}$ of SCI method is quite close to specified confidence coefficient and AL of SCI method is the shortest.

Keywords : confidence interval, regression coefficient, errors in variables

บทนำ

การถดถอยเชิงเส้นอย่างง่าย (Simple linear regression) คือ วิธีการทางสถิติที่ใช้ในการศึกษาความสัมพันธ์เชิงเส้นระหว่างตัวแปร 2 ตัวแปร ประกอบด้วย ตัวแปรตาม (Dependent variable) และตัวแปรอิสระ (Independent variable) โดยสามารถเขียนตัวแบบความสัมพันธ์ได้ดังนี้

$$Y_i = \beta_{0R} + \beta_{1R}X_i + \varepsilon_i \tag{1}$$

เมื่อ Y คือ ตัวแปรตาม
X คือ ตัวแปรอิสระ และเป็นค่าคงที่ที่ทราบค่า
β_{0R} คือ ระยะตัดแกน Y
β_{1R} คือ ความชันของเส้นถดถอย หรือสัมประสิทธิ์การถดถอยเชิงเส้นอย่างง่าย
ε_i คือ ความคลาดเคลื่อนเชิงสุ่ม (Kutner, *et al.*, 2005)

ตัวแบบดังกล่าวจะพิจารณาความคลาดเคลื่อนเฉพาะความคลาดเคลื่อนที่เกิดจากตัวแปรตาม Y เพียงอย่างเดียว (Casella and Berger, 2002) แต่ในความเป็นจริงปัญหาของการวิเคราะห์การถดถอยนั้นความคลาดเคลื่อนอาจจะเกิดได้ทั้งในตัวแปรตามและตัวแปรอิสระ (Dachodomphant, 2003) เนื่องจากข้อมูลที่นำมาวิเคราะห์อาจจะเป็นข้อมูลที่ไม่สามารถวัดค่าได้โดยตรง เช่น ในปี ค.ศ. 1800 – 1899 นักฟิสิกส์ชาวสกอตแลนด์ได้ศึกษาการประมาณระดับความสูงเหนือน้ำทะเล จากอุณหภูมิ ณ จุดที่ต้องการวัด ซึ่งโดยทั่วไปการหาความสูงเหนือระดับน้ำทะเลสามารถคำนวณได้จากความสูงเหนือระดับน้ำทะเล = (760 – ความดันบรรยากาศ ณ จุดนั้น) x 11 แต่ในปี ค.ศ. 1800 - 1899 นั้นบารอมิเตอร์ซึ่งเป็นเครื่องมือ

สำหรับวัดค่าความดันบรรยากาศค่อนข้างเปราะและแตกหักง่าย ดังนั้นจึงต้องประมาณค่าความดันบรรยากาศ (Y) จากอุณหภูมิ ณ จุดที่ต้องการวัด (X) แล้วจึงนำค่าที่ได้ไปคำนวณเพื่อหาค่าระดับความสูงเหนือน้ำทะเล เป็นต้น (Casella and Berger, 2002) หรือในบางกรณีข้อมูลที่ได้อาจมาจากการทดลอง (Experimental data) เช่น การศึกษาความสัมพันธ์ระหว่างผลผลิตข้าวโพด (Y) และปริมาณไนโตรเจนในดิน (X) ในการศึกษาดังกล่าวมีวิธีการเก็บรวบรวมข้อมูลโดยการวัดค่าปริมาณผลผลิตข้าวโพดที่ได้ และวัดปริมาณไนโตรเจนจากการเก็บตัวอย่างดินเพื่อนำไปตรวจสอบในห้องปฏิบัติการ เป็นต้น (Fuller, 1987) จะเห็นว่าจากตัวอย่างขั้นต้นข้อมูลที่ได้อาจเกิดความคลาดเคลื่อนในการวัด (Measurement error) หรือความคลาดเคลื่อนจากการสังเกต (Observation error) (Dachodomphant, 2003) หรือความคลาดเคลื่อนจากการเลือกตัวอย่าง (Sampling error) (Fuller, 1987) ดังนั้นการสร้างตัวแบบการถดถอยควรคำนึงถึงคุณลักษณะของตัวแปรอิสระร่วมด้วย

การวิเคราะห์การถดถอยเมื่อมีความคลาดเคลื่อนในตัวแปร (Regression with errors in variables or measurement error model : EIV) เป็นการวิเคราะห์การถดถอยที่ตัวแปรตาม Y และตัวแปรอิสระ X เป็นตัวแปรสุ่ม ดังนั้นการพิจารณาความคลาดเคลื่อนจะพิจารณาจากทั้ง 2 ตัวแปร ซึ่งมีตัวแบบของความสัมพันธ์ ดังนี้

$$Y_i = \eta_i + \varepsilon_i \tag{2}$$

เมื่อ $\eta_i = \beta_0 + \beta_1\xi_i$ และ $X_i = \xi_i + \delta_i$ โดยที่ ตัวแปร η_i และ ξ_i เรียกว่า ตัวแปรแฝง (Latent variable) และ ε_i และ δ_i คือ ความคลาดเคลื่อนเชิงสุ่มในตัวแปร Y และ X ตามลำดับ (Cheng and Van Ness, 1999) หากกำหนดให้ตัวแปร ξ_i คือ ตัวแปรที่มีค่าคงตัวและทราบค่า จะเรียกความสัมพันธ์ของสมการ (2) ว่า Linear functional relationship และหากตัวแปร ξ_i คือ ตัวแปรสุ่ม จะเรียกว่า Linear structural relationship (Casella and Berger, 2002) โดยในการศึกษาครั้งนี้จะศึกษาภายใต้ความสัมพันธ์แบบ Linear structural relationship เนื่องจากภายใต้ความสัมพันธ์ดังกล่าวจะกำหนดให้ ξ_i คือ ตัวแปรสุ่ม ดังนั้นจึงมีการพิจารณาความคลาดเคลื่อนจากตัวแปร ξ_i ร่วมด้วย ซึ่งมีข้อตกลงเบื้องต้น ดังนี้

$$(\xi_i, \delta_i, \varepsilon_i)^T \sim N_3((\mu_x, 0, 0)^T, \mathrm{diag}(\sigma^2, \sigma_\delta^2, \sigma_\varepsilon^2)) \tag{3}$$

เมื่อ $N_3((\mu_x, 0, 0)^T, \mathrm{diag}(\sigma^2, \sigma_\delta^2, \sigma_\varepsilon^2))$ คือ การแจกแจงปรกติ 3 ตัวแปรที่มีค่าเฉลี่ยเท่ากับ $(\mu_x, 0, 0)^T$ และมีความแปรปรวนเท่ากับ $\mathrm{diag}(\sigma^2, \sigma_\delta^2, \sigma_\varepsilon^2)$ โดยที่ $\mathrm{diag}(\sigma^2, \sigma_\delta^2, \sigma_\varepsilon^2)$ คือ เมตริกซ์ทแยงมุมขนาด 3×3 ที่มีสมาชิกทุกตัวที่ไม่ได้อยู่บนเส้นทแยงมุมหลักมีค่าเป็นศูนย์ และจะได้ว่า

$$(X_i, Y_i)^T \sim N_2((\mu_x, \mu_y)^T, \Sigma) \tag{4}$$

เมื่อ $\mu_y = \beta_0 + \beta_1\mu_x$ และ $\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} = \begin{bmatrix} \sigma^2 + \sigma_\delta^2 & \beta_1\sigma^2 \\ \beta_1\sigma^2 & \beta_1^2\sigma^2 + \sigma_\varepsilon^2 \end{bmatrix}$ จากข้อตกลงเบื้องต้นดังกล่าวจะเห็นว่าตัวแปรสุ่ม X และ Y มีการแจกแจงปรกติสองตัวแปร มีค่าพารามิเตอร์ที่ไม่ทราบค่าอยู่ทั้งหมด 5 ค่า ($\mu_x, \mu_y, \sigma_{xx}, \sigma_{xy}, \sigma_{yy}$) แต่

ภายใต้ข้อตกลงเบื้องต้นของตัวแบบนั้นมีค่าพารามิเตอร์ที่ไม่ทราบค่าอยู่ทั้งหมด 6 ค่า ($\mu_x, \beta_0, \beta_1, \sigma^2, \sigma_\delta^2, \sigma_\varepsilon^2$) ซึ่งเซตของพารามิเตอร์ในแต่ละตัวแบบภายใต้ค่าพารามิเตอร์ 6 ค่านั้นอาจมีได้หลายค่าแตกต่างกัน แต่เมื่อพิจาณาการแจกแจงของตัวแปรสุ่ม X และ Y ภายใต้ค่าพารามิเตอร์ 5 ค่าแล้วนั้นพบว่าตัวแปรสุ่ม X และ Y ของแต่ละตัวแบบมีการแจกแจงเดียวกัน จึงทำให้ตัวแบบไม่มีความเป็นเอกลักษณ์ (Fuller, 1987) ดังนั้นต้องมีการกำหนดข้อตกลงในการประมาณค่าพารามิเตอร์ภายใต้ข้อตกลงใดข้อตกลงหนึ่ง ดังต่อไปนี้ (1) ทราบค่าอัตราส่วนของความแปรปรวนของความคลาดเคลื่อน ($\lambda = \sigma_\varepsilon^2/\sigma_\delta^2$) (2) ทราบค่าอัตราส่วนความเชื่อถือได้ ($k_\xi = \sigma^2/(\sigma^2 + \sigma_\delta^2)$) ซึ่งแสดงให้เห็นถึงความคลาดเคลื่อนที่เกิดขึ้นในตัวแปรอิสระ X (Fuller, 1987) เช่น ถ้า k_ξ มีค่าสูงแสดงว่า σ_δ^2 มีค่าน้อย ดังนั้นความคลาดเคลื่อนในตัวแปรอิสระ X จะมีค่าน้อย ทำให้ความผิดพลาดในการประมาณค่าพารามิเตอร์ β_1 ลดน้อยลง (3) ทราบค่า σ_δ^2 (4) ทราบค่า σ_ε^2 (5) ทราบค่า σ_δ^2 และ σ_ε^2 และ (6) ทราบค่า β_0 และ $E(X) \neq 0$ (Cheng and Van Ness, 1999) สำหรับการศึกษาในครั้งนี้จะพิจารณาภายใต้กรณีทราบค่า σ_δ^2

ตัวแบบการถดถอยเมื่อมีความคลาดเคลื่อนในตัวแปรนั้นโดยทั่วไปจะไม่ทราบค่าพารามิเตอร์ที่แท้จริงของตัวแบบ ดังนั้นจึงต้องทำการประมาณค่าโดยใช้ข้อมูลที่ได้จากตัวอย่าง ในทางสถิติการประมาณค่าพารามิเตอร์มีวิธีการประมาณ 2 วิธี คือ การประมาณค่าแบบจุด (Point estimation) และการประมาณค่าแบบช่วง (Interval estimation) โดยการประมาณค่าแบบจุดนั้นจะประมาณค่าพารามิเตอร์ด้วยค่าสถิติค่าเดียว ซึ่งค่าสถิติที่ใช้ในการประมาณนั้นอาจมีค่าเท่ากับหรือไม่เท่ากับค่าพารามิเตอร์ก็ได้ ในขณะที่การประมาณค่าแบบช่วงเป็นการประมาณค่าพารามิเตอร์ของประชากรว่าจะอยู่ในช่วงใดช่วงหนึ่ง ช่วงของการประมาณจะบอกถึงค่าต่ำสุดและค่าสูงสุดของพารามิเตอร์โดยที่ช่วงดังกล่าวจะคลุมค่าพารามิเตอร์ด้วยความน่าจะเป็นตามที่กำหนด

ในทางปฏิบัติการกำหนดระดับความเชื่อมั่น (Confidence level) ซึ่งเขียนสัญลักษณ์แทนด้วย $(1-\alpha)100\%$ โดย $0 < \alpha < 1$ และค่า $1-\alpha$ เรียกว่า สัมประสิทธิ์ของความเชื่อมั่น (Confidence coefficient) การประมาณค่าแบบช่วงจึงเหมาะสมมากกว่าการประมาณแบบจุด โดยเฉพาะเมื่อลักษณะของประชากรมีความแปรปรวนสูง และขนาดตัวอย่างเล็กซึ่งอาจส่งผลให้ตัวประมาณแบบจุดมีความคลาดเคลื่อนสูง ดังนั้นถ้าใช้การประมาณแบบช่วงอาจทำให้การประมาณดีขึ้นและมีความคลาดเคลื่อนน้อยลง (Kiranandana, 2002) สำหรับตัวแบบการถดถอยอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปรนั้นเป็นตัวแบบที่พิจารณาความคลาดเคลื่อนจากทั้งตัวแปร X และตัวแปร Y โดยความคลาดเคลื่อนดังกล่าวจะส่งผลต่อการประมาณค่าพารามิเตอร์ β_1 หากความคลาดเคลื่อนมีค่าสูงก็จะทำให้การประมาณค่าแบบจุดสำหรับ β_1 นั้นมีความผิดพลาดสูง และการประมาณค่าแบบช่วงจะได้ช่วงความเชื่อมั่นที่กว้าง (Fuller, 1987; Gleser and Hwang, 1987) ซึ่งวิธีการที่จะทำให้ตัวประมาณมีความถูกต้องมากยิ่งขึ้น คือ การเพิ่มขนาดตัวอย่าง แต่ในความเป็นจริงนั้นการเพิ่มขนาดตัวอย่างทำได้ค่อนข้างยาก เนื่องจากถ้าผู้วิจัยต้องการขนาดตัวอย่างใหญ่ ค่าใช้จ่ายสำหรับการได้มาซึ่งข้อมูลนั้นย่อมสูงขึ้น (Tsai, 2013) ดังนั้นการเลือกใช้วิธีการประมาณค่าที่เหมาะสมเมื่อขนาดตัวอย่างเล็กจึงเป็นสิ่งสำคัญที่ต้องพิจารณา

สำหรับช่วงความเชื่อมั่นของ β_1 ในตัวแบบการถดถอยเมื่อมีความคลาดเคลื่อนในตัวแปร Gleser และ Hwang (1987) ได้กล่าวว่า ช่วงความเชื่อมั่นที่แท้จริงของ β_1 จะไม่สามารถหาขอบเขตได้ ในทางกลับกันช่วงความเชื่อมั่นของ β_1 ที่ได้จากการประมาณจะสามารถหาขอบเขตของความเชื่อมั่นได้ แต่ช่วงความเชื่อมั่นที่ได้จะมีค่าประมาณความน่าจะเป็นคุ้มรวม (Estimated coverage probability : $\widehat{CP}$) ไม่ใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด และมีค่าความกว้างเฉลี่ย (Average lengths : AL) ที่ค่อนข้างกว้างเมื่อค่า k_ξ มีค่าต่ำ จากการศึกษาที่ผ่านมา Tsai (2010) ได้นำวิธีการสร้างช่วง

ความเชื่อมั่นนัยทั่วไป (Generalized confidence interval : GCI) ซึ่งเป็นวิธีที่ถูกเสนอโดย Weerahandi (1993) มาประยุกต์ใช้ในการสร้างช่วงความเชื่อมั่นสำหรับ β_1 ในตัวแบบการถดถอยอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปร และศึกษาเปรียบเทียบกับช่วงความเชื่อมั่นเชิงเส้นกำกับ (Asymptotic confidence interval : ACI) (Fuller, 1987) และช่วงความเชื่อมั่นแบบเที่ยงตรงเชิงเส้นกำกับ (Asymptotically honest confidence interval : AHCI) (Huwang, 1996) ต่อมา Tsai (2013) ได้เสนอช่วงความเชื่อมั่นแซนด์วิช (Sandwich confidence interval : SCI) วิธีนี้ได้ทำการพัฒนามาจากวิธี ACI แต่ใช้วิธีการประมาณค่าพารามิเตอร์ด้วยวิธี M-estimator และประมาณค่าความแปรปรวนของตัวประมาณด้วยวิธีแซนด์วิช (Sandwich estimated) แล้วศึกษาเปรียบเทียบกับวิธี ACI วิธี AHCI และวิธี GCI ผลการศึกษาจากทั้ง 2 งานวิจัยข้างต้นพบว่า เมื่อ ค่า k_ξ มีค่าน้อย วิธี GCI จะมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นแคบกว่าวิธี ACI และวิธี SCI และ Tsai (2013) ยังกล่าวว่าช่วงความเชื่อมั่นวิธี GCI นั้นถึงแม้จะมีค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่แคบกว่าวิธี ACI และวิธี SCI แต่เมื่อพิจารณาค่า $\hat{CP}$ พบว่ามีค่า $\hat{CP}$ ที่ค่อนข้างสูงกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดมากเกินไป ในขณะที่เมื่อขนาดตัวอย่างใหญ่ และ/หรือ ค่า k_ξ มีค่ามาก ค่า $\hat{CP}$ ของทั้ง 4 วิธีจะมีค่าใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด และเมื่อพิจารณาความกว้างเฉลี่ยของช่วงความเชื่อมั่น พบว่า วิธี ACI หรือวิธี SCI จะมีความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่แคบกว่าวิธี GCI และวิธี AHCI แต่วิธี ACI และวิธี SCI จะถูกจำกัดด้วยเงื่อนไข $S_{yy}(S_{xx} - \sigma_\delta^2) - S_{xy} > 0$ และ $S_{xx} - \sigma_\delta^2 > 0$ กล่าวคือข้อมูลจะต้องเป็นไปตามเงื่อนไขดังกล่าวจึงจะหาช่วงความเชื่อมั่นด้วยวิธี ACI และวิธี SCI ได้ ในขณะที่วิธี GCI นั้นจะไม่ถูกจำกัดด้วยเงื่อนไขดังกล่าว และสำหรับช่วงความเชื่อมั่นที่ได้จากวิธี AHCI พบว่าในบางสถานการณ์นั้นจะไม่สามารถหาขอบเขตของช่วงความเชื่อมั่นได้

ดังนั้น ผู้วิจัยจึงสนใจศึกษาและเสนอแนวทางในการพัฒนาวิธีการประมาณค่าพารามิเตอร์แบบช่วงสำหรับสัมประสิทธิ์การถดถอยอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปรสำหรับขนาดตัวอย่างเล็ก และเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบช่วงที่พัฒนาขึ้นกับวิธี ACI และ SCI โดยใช้การจำลองมอนติคาร์โล (Monte Carlo method) และเกณฑ์ในการพิจารณาช่วงความเชื่อมั่นที่เหมาะสมนั้นจะพิจารณาจากค่าประมาณความน่าจะเป็นคุ้มรวม และความกว้างเฉลี่ยของช่วงความเชื่อมั่น

วิธีดำเนินการวิจัย

1. ช่วงความเชื่อมั่น

1.1 ช่วงความเชื่อมั่นเชิงเส้นกำกับ

ในการวิเคราะห์การถดถอยอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปร ภายใต้ความสัมพันธ์ในลักษณะ Linear structural relationship เมื่อทราบค่า σ_δ^2 จะได้ตัวประมาณค่าพารามิเตอร์ ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum likelihood estimators : MLEs) (Fuller, 1987; Cheng and Van Ness, 1999) ดังนี้

$$\hat{\beta}_1 = \frac{S_{xy}}{(S_{xx} - \sigma_\delta^2)} \; , \; \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\mu}_x = \bar{X} \, , \; \hat{\sigma}^2 = S_{xx} - \sigma_\delta^2 \, , \; \hat{\sigma}_\varepsilon^2 = S_{yy} - \hat{\beta}_1 S_{xy}$$

เมื่อ $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ $S_{xx} = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, $S_{xy} = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$ และ $S_{yy} = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

ซึ่งตัวประมาณที่ได้ จะถูกจำกัดด้วยเงื่อนไข

$$S_{yy}(S_{xx} - \sigma_\delta^2) - S_{xy} > 0 \text{ และ } S_{xx} - \sigma_\delta^2 > 0 \tag{5}$$

พิจารณา $\hat{\beta}_1$ เมื่อ $n \to \infty$ จะได้

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{V}(\hat{\beta}_1)}} \xrightarrow{d} N(0,1) \tag{6}$$

โดยที่ $\hat{V}(\hat{\beta}_1) = (n-1)^{-1}\hat{\sigma}^{-4}(S_{xx}S_{\upsilon\upsilon} + \hat{\beta}_1^2\sigma_\delta^4)$ และ $S_{\upsilon\upsilon} = (n-2)^{-1}(n-1)(S_{yy} - 2\hat{\beta}_1 S_{xy} + \hat{\beta}_1^2 S_{xx})$

จะได้ช่วงความเชื่อมั่นวิธี ACI ดังนี้

$$\hat{\beta}_1 - Z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{\beta}_1)} < \beta_1 < \hat{\beta}_1 + Z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{\beta}_1)} \tag{7}$$

Gleser (1987) ได้เสนอให้ใช้ $t_{1-\frac{\alpha}{2},n-2}$ แทน $Z_{1-\frac{\alpha}{2}}$ เนื่องจากการแจกแจงที สามารถใช้ได้เมื่อขนาดตัวอย่างเล็ก และเมื่อขนาดตัวอย่างใหญ่การแจกแจงทีจะลู่เข้าสู่การแจกแจงปรกติมาตรฐาน

1.2 ช่วงความเชื่อมั่นแซนด์วิช

วิธี SCI เป็นช่วงความเชื่อมั่นที่พัฒนามาจากช่วงความเชื่อมั่นเชิงเส้นกำกับ เมื่อขนาดตัวอย่างใหญ่จะใช้ตัวประมาณแซนด์วิช (Sandwich estimator) ในการประมาณค่าความแปรปรวนของ $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$ ที่ได้จากวิธีภาวะน่าจะเป็นสูงสุด ซึ่งในการหาตัวประมาณดังกล่าวจะหาโดยใช้ทฤษฎีสมการการประมาณที่ไม่เอนเอียง (Unbiased estimation equation theory) เริ่มจากการกำหนดสมการการประมาณที่ไม่เอนเอียงสำหรับ $\boldsymbol{\beta}$ คือ $\sum_{i=1}^{n}\boldsymbol{\psi}(X_i, Y_i, \boldsymbol{\beta}) = \mathbf{0}$ และกำหนดให้ $\boldsymbol{\psi}(X_i, Y_i, \boldsymbol{\beta}) = \frac{\partial}{\partial\boldsymbol{\beta}}Q_i(\boldsymbol{\beta})$ เมื่อ $Q_i(\boldsymbol{\beta}) = (Y_i - \beta_0 - \beta_1 X_i)^2 - \beta_1^2\sigma_\delta^2$ ซึ่งวิธีการที่ใช้ในการแก้สมการดังกล่าวเรียกว่าการหาตัวประมาณแบบ Maximum likelihood type estimators (M – estimators) สำหรับ $\boldsymbol{\beta}$ และ $E(\boldsymbol{\psi}(X_i, Y_i, \boldsymbol{\beta})) = \mathbf{0}$, $\forall i$ และจากวิธีการดังกล่าวจะได้ตัวประมาณที่มีคุณสมบัติความคงเส้นคงวาสำหรับ $\boldsymbol{\beta}$ (Tsai, 2013)

เมื่อ $n \to \infty$ การแจกแจงของตัวสถิติ $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$ จะลู่เข้าสู่การแจกแจงปรกติ ดังนี้

$$[\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})]^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathbf{d}} \mathbf{N}_2(\mathbf{0}, \mathbf{I_2}) \tag{8}$$

โดยที่ $\hat{\mathbf{V}}\left(\hat{\boldsymbol{\beta}}\right) = n^{-1}\hat{\mathbf{A}}_n^{-1}\hat{\mathbf{B}}_n(\hat{\mathbf{A}}_n^{-1})^T$ ซึ่งตัวประมาณค่าความแปรปรวนของ $\hat{\boldsymbol{\beta}}$ นี้จะเรียกว่าตัวประมาณแซนด์วิช

เมื่อ $\hat{\mathbf{A}}_\mathbf{n} = n^{-1}\sum_{i=1}^{n}[\frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{\psi}(X_i, Y_i, \hat{\boldsymbol{\beta}})]$ และ $\hat{\mathbf{B}}_n = n^{-1}\sum_{i=1}^{n}\boldsymbol{\psi}(X_i, Y_i, \hat{\boldsymbol{\beta}})\boldsymbol{\psi}^T(X_i, Y_i, \hat{\boldsymbol{\beta}})$ จะได้ $\hat{\mathbf{A}}_n$ และ $\hat{\mathbf{B}}_n$ ดังนี้

$$\hat{\mathbf{A}}_n = \frac{1}{n}\begin{bmatrix} -n & -\sum_{i=1}^{n}X_i \\ -\sum_{i=1}^{n}X_i & -\sum_{i=1}^{n}X_i^2 + n\sigma_\delta^2 \end{bmatrix}$$

$$\hat{\mathbf{B}}_n = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 & \sum_{i=1}^{n}\left\{\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)\left[\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)X_i + \hat{\beta}_1\sigma_\delta^2\right]\right\} \\ \sum_{i=1}^{n}\left\{\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)\left[\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)X_i + \hat{\beta}_1\sigma_\delta^2\right]\right\} & \sum_{i=1}^{n}\left(\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)X_i + \hat{\beta}_1\sigma_\delta^2\right)^2 \end{bmatrix}$$

เมื่อ $\hat{\beta}_1 = \dfrac{S_{xy}}{(S_{xx} - \sigma_\delta^2)}$, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$ และ $\hat{V}^*(\hat{\beta}_1) \in \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}})$ ดังนั้นจะได้ช่วงความเชื่อมั่นวิธี SCI ดังนี้

$$\hat{\beta}_1 - z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}^*(\hat{\beta}_1)} < \beta_1 < \hat{\beta}_1 + z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}^*(\hat{\beta}_1)} \tag{9}$$

โดยจะใช้ $t_{1-\frac{\alpha}{2}, n-2}$ แทน $Z_{1-\frac{\alpha}{2}}$ และวิธี SCI ถูกจำกัดด้วยเงื่อนไขในสมการที่ 5

1.3 ช่วงความเชื่อมั่นกำลังสองน้อยที่สุดปรับปรุง (Improved ordinary least square confidence interval : IOLS)

Fuller (1987) ได้กล่าวว่า ตัวประมาณค่าแบบจุดสำหรับ β_1 ในตัวแบบการถดถอยอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปรจะพิจารณาจากการหาความสัมพันธ์เชิงเส้นของตัวแปร 2 ตัว คือตัวแปรอิสระ (X) และตัวแปรตาม (Y) ในตัวแบบการถดถอยอย่างง่าย ดังนั้นเมื่อกำหนด β_{1R} คือ ค่าสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยอย่างง่าย และกำหนดให้ $\hat{\beta}_{1R} = \dfrac{S_{xy}}{S_{xx}}$ เป็นตัวประมาณของ β_{1R} ที่ได้จากวิธีกำลังสองน้อยที่สุด (Ordinary least square method)

จากข้อตกลงเบื้องต้นของตัวแปร X และ Y ในตัวแบบการถดถอยอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปร (สมการที่ 4) จะได้ว่า

$$\beta_{1R} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \frac{\beta_1\sigma^2}{\sigma_{xx}} \tag{10}$$

เมื่อ β_1 คือ ค่าสัมประสิทธิ์การถดถอยของตัวแบบเมื่อมีความคลาดเคลื่อนในตัวแปร และ $\sigma_{xx} = \sigma^2 + \sigma_\delta^2$ ทำการแก้สมการเพื่อหาค่า β_1 จะได้

$$\beta_1 = \left(\frac{\sigma_{xx}}{\sigma^2}\right)\beta_{1R} = \left(\frac{\sigma^2+\sigma_\delta^2}{\sigma^2}\right)\beta_{1R} \tag{11}$$

สำหรับการหาช่วงความเชื่อมั่นสำหรับ β_1 จะเริ่มพิจารณาจากตัวประมาณค่าแบบจุดของ β_{1R} จะได้ว่า $\frac{\hat{\beta}_{1R} - \beta_{1R}}{\sqrt{\hat{V}(\hat{\beta}_{1R})}}$ มีการแจกแจงที ด้วยองศาเสรี เท่ากับ $n - 2$ (Casella and Berger, 2002; Chiawkhun, 2007) ดังนั้นที่ระดับความเชื่อมั่น $(1-\alpha)100\%$ จะได้ช่วงความเชื่อมั่นสำหรับ β_{1R} ดังนี้

$$\hat{\beta}_{1R} - t_{1-\frac{\alpha}{2},n-2}\sqrt{\hat{V}(\hat{\beta}_{1R})} < \beta_{1R} < \hat{\beta}_{1R} + t_{1-\frac{\alpha}{2},n-2}\sqrt{\hat{V}(\hat{\beta}_{1R})} \tag{12}$$

เมื่อ $\hat{V}(\hat{\beta}_{1R}) = \dfrac{\mathrm{MSE}\,(\hat{Y}_R)}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \dfrac{\mathrm{MSE}\,(\hat{Y}_R)}{(n-1)S_{xx}}$, $\mathrm{MSE}\,(\hat{Y}_R) = \dfrac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_{0R} - \hat{\beta}_{1R}X_i)^2}{n-2}$

จะได้

$$\frac{S_{xy}}{S_{xx}} - t_{1-\frac{\alpha}{2},n-2}\sqrt{\frac{\mathrm{MSE}\,(\hat{Y}_R)}{(n-1)S_{xx}}} < \beta_{1R} < \frac{S_{xy}}{S_{xx}} + t_{1-\frac{\alpha}{2},n-2}\sqrt{\frac{\mathrm{MSE}\,(\hat{Y}_R)}{(n-1)S_{xx}}} \tag{13}$$

จาก $\beta_{1R} = \beta_1\left(\dfrac{\sigma^2}{\sigma^2 + \sigma_\delta^2}\right)$ จะได้ช่วงความเชื่อมั่นสำหรับ β_1 ดังนี้

$$\frac{S_{xy}}{S_{xx}} - t_{1-\frac{\alpha}{2},n-2}\sqrt{\frac{\mathrm{MSE}\,(\hat{Y}_R)}{(n-1)S_{xx}}} < \beta_1\left(\frac{\sigma^2}{\sigma^2+\sigma_\delta^2}\right) < \frac{S_{xy}}{S_{xx}} + t_{1-\frac{\alpha}{2},n-2}\sqrt{\frac{\mathrm{MSE}\,(\hat{Y}_R)}{(n-1)S_{xx}}} \tag{14}$$

ซึ่งจะประมาณ σ^2 ด้วย $\hat{\sigma}^2 = S_{xx} - \sigma_\delta^2$ จะได้ช่วงความเชื่อมั่นสำหรับ β_1 ดังนี้

$$\frac{S_{xx}}{S_{xx} - \sigma_\delta^2}\left(\frac{S_{xy}}{S_{xx}} - t_{1-\frac{\alpha}{2}, n-2}\sqrt{\frac{MSE\,(\hat{Y}_R)}{(n-1)S_{xx}}}\right) < \beta_1 < \frac{S_{xx}}{S_{xx} - \sigma_\delta^2}\left(\frac{S_{xy}}{S_{xx}} + t_{1-\frac{\alpha}{2}, n-2}\sqrt{\frac{MSE\,(\hat{Y}_R)}{(n-1)S_{xx}}}\right) \qquad (15)$$

ช่วงความเชื่อมั่นวิธี IOLS จะถูกจำกัดด้วยเงื่อนไข $S_{xx} - \sigma_\delta^2 > 0$

2. **ขอบเขตในการวิจัย** (Scope of the Research)

1. วิธีประมาณค่าพารามิเตอร์ที่ใช้ในการศึกษาในครั้งนี้ ประกอบด้วยวิธี ACI วิธี SCI และวิธี IOLS

2. กำหนดขนาดตัวอย่าง เท่ากับ 10, 15, 20 และ 30 และกำหนด σ_δ^2 เท่ากับ 0.1, 0.3, 0.5, 0.7 และ 0.9

3. กำหนดค่าพารามิเตอร์ $\beta_0 = 1$, $\beta_1 = 1$, $\mu_x = 1$, $k_\xi = \dfrac{\sigma^2}{\sigma^2 + \sigma_\delta^2}$, $\sigma^2 = 1 - \sigma_\delta^2$ และ $\sigma_\varepsilon^2 = \sigma_\delta^2$

4. จำลองข้อมูล $(\xi_i, \delta_i, \varepsilon_i)$ ที่มีการแจกแจงปรกติ 3 ตัวแปรจากสมการที่ 3 แล้วนำข้อมูลที่จำลองได้สร้างตัวแปรสุ่ม (X_i, Y_i) โดยใช้ความสัมพันธ์ $Y_i = \eta_i + \varepsilon_i$ เมื่อ $\eta_i = \beta_0 + \beta_1\xi_i$ และ $X_i = \xi_i + \delta_i$

5. ตรวจสอบข้อมูลที่จำลองได้ในข้อ 4 ว่ามีความสอดคล้องกับเงื่อนไขในสมการที่ 5 หรือไม่ เนื่องจากช่วงความเชื่อมั่นเชิงเส้นกำกับและช่วงความเชื่อมั่นแซนด์วิชถูกจำกัดด้วยเงื่อนไขดังกล่าว และช่วงความเชื่อมันที่พัฒนาถูกจำกัดด้วยเงื่อนไข $S_{xx} - \sigma_\delta^2 > 0$

6. กำหนดระดับความเชื่อมัน 90% และ 95%

7. ทำซ้ำในข้อ 1 - 6 จำนวน 1,000 ครั้ง

8. เปรียบเทียบช่วงความเชื่อมั่นจากค่าประมาณความน่าจะเป็นคุ้มรวม โดยใช้เกณฑ์ของ Bradley (1978) ที่ระดับความเชื่อมั่น 90% และ 95% ค่าประมาณความน่าจะเป็นคุ้มรวมค่าพารามิเตอร์ β_1 จะอยู่ในช่วง 0.850 – 0.950 ($0.850 \leq \hat{CP} \leq 0.950$) และ 0.925 – 0.975 ($0.925 \leq \hat{CP} \leq 0.975$) ตามลำดับ (Romano *et al.*, 2011) เมื่อค่า $\hat{CP}$ อยู่ในช่วงดังกล่าวแสดงว่า ช่วงความเชื่อมั่นมีค่า $\hat{CP}$ ที่ใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด จากนั้นพิจารณาเปรียบเทียบความกว้างเฉลี่ยของช่วงความเชื่อมั่น

ผลการวิจัยและวิจารณ์ผล

ตารางที่ 1 ค่าประมาณความน่าจะเป็นคุ้มรวมของช่วงความเชื่อมั่นทั้ง 3 วิธี ที่ระดับความเชื่อมั่น 90% และ 95%

n	k_ξ	90%			95%		
		ACI	SCI	IOLS	ACI	SCI	IOLS
n=10	0.1	0.836	0.800	0.792	0.903	0.855	0.866
	0.3	**0.895**	0.848	**0.881**	**0.942**	0.894	**0.929**
	0.5	**0.894**	0.849	**0.883**	**0.941**	0.894	**0.933**
	0.7	**0.936**	**0.898**	**0.929**	**0.972**	**0.929**	**0.965**
	0.9	**0.938**	**0.861**	**0.932**	**0.969**	**0.925**	**0.965**
n=15	0.1	0.825	0.801	0.765	0.885	0.853	0.845
	0.3	**0.897**	**0.867**	**0.875**	**0.943**	0.912	**0.932**
	0.5	**0.937**	**0.909**	**0.919**	**0.968**	**0.940**	**0.954**
	0.7	**0.945**	**0.908**	**0.930**	**0.974**	**0.941**	**0.966**
	0.9	**0.934**	**0.878**	**0.921**	**0.972**	**0.934**	**0.964**
n=20	0.1	0.826	0.819	0.780	0.905	0.878	0.866
	0.3	**0.900**	**0.867**	**0.877**	**0.941**	0.917	**0.929**
	0.5	**0.918**	**0.898**	**0.901**	**0.952**	**0.934**	**0.942**
	0.7	**0.929**	**0.896**	**0.890**	**0.961**	**0.944**	**0.954**
	0.9	**0.924**	**0.868**	**0.905**	**0.964**	**0.927**	**0.952**
n=30	0.1	0.820	0.817	0.770	0.876	0.874	0.840
	0.3	**0.904**	**0.889**	**0.882**	**0.941**	**0.931**	**0.932**
	0.5	**0.930**	**0.913**	**0.893**	**0.962**	**0.945**	**0.948**
	0.7	**0.944**	**0.926**	**0.883**	**0.972**	**0.954**	**0.946**
	0.9	**0.918**	**0.880**	**0.902**	**0.958**	**0.936**	**0.953**

หมายเหตุ : ค่า $\widehat{CP}$ ตัวหนา หมายถึง เมื่อกำหนดระดับความเชื่อมั่น 90% และ 95% ค่า $0.850 \leq \widehat{CP} \leq 0.950$ และ $0.925 \leq \widehat{CP} \leq 0.975$ ตามลำดับ

ตารางที่ 2 ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นทั้ง 3 วิธี ที่ระดับความเชื่อมั่น 90% และ 95%

n	k_ξ	90%			95%		
		ACI	SCI	IOLS	ACI	SCI	IOLS
n=10	0.1	-	-	-	-	-	-
	0.3	14.0230	-	**4.5509**	17.3898	-	**5.6435**
	0.5	6.1973	-	**3.1013**	7.6852	-	**3.8458**
	0.7	2.1963	2.0913	**1.6377**	2.7236	2.5934	**2.0309**
	0.9	0.7770	**0.6407**	0.7126	0.9635	**0.7945**	0.8837
n=15	0.1	-	-	-	-	-	-
	0.3	26.5431	19.7701	**4.5080**	32.3800	-	**5.4993**
	0.5	4.0485	6.0131	**2.1348**	4.9388	7.3354	**2.6043**
	0.7	1.7459	1.6604	**1.2191**	2.1298	2.0255	**1.4871**
	0.9	0.5324	**0.4505**	0.4987	0.6495	**0.5496**	0.6084
n=20	0.1	-	-	-	-	-	-
	0.3	5.9826	16.0659	**3.2369**	7.2482	-	**3.9217**
	0.5	2.5877	3.5098	**1.6939**	3.1351	4.2523	**2.0522**
	0.7	1.1185	1.0220	**0.9013**	1.3552	1.2382	**1.0919**
	0.9	0.4405	**0.3939**	0.4135	0.5337	**0.4772**	0.5009
n=30	0.1	-	-	-	-	-	-
	0.3	4.5928	17.3529	**2.4344**	5.5304	20.8954	**2.9314**
	0.5	2.5529	2.7024	**1.3649**	3.0741	3.2541	**1.6436**
	0.7	0.8672	0.8221	**0.7023**	1.0443	0.9899	**0.8457**
	0.9	0.3388	**0.3121**	0.3191	0.4080	**0.3758**	0.3842

หมายเหตุ : ค่า AL ตัวหนา หมายถึง ค่า AL ที่แคบที่สุดภายใต้ขนาดตัวอย่างและ k_ξ เดียวกัน ณ ระดับความเชื่อมั่น 90% และ 95%

จากตารางที่ 1 ในการเปรียบเทียบค่า $\hat{CP}$ ของช่วงความเชื่อมั่นสำหรับ β_1 ทั้ง 3 วิธี เมื่อกำหนดขนาดตัวอย่างเล็กกว่าหรือเท่ากับ 30 $(n \le 30)$ ที่ระดับความเชื่อมั่น 90% และ 95% พบว่า เมื่อ $k_\xi \ge 0.3$ ช่วงความเชื่อมั่นวิธี ACI และวิธี IOLS มีค่า $\hat{CP}$ ใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด แต่เมื่อ $k_\xi = 0.1$ พบว่าค่า $\hat{CP}$ ของทั้ง 2 วิธีข้างต้นจะมีค่าน้อยกว่าสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด ในขณะที่วิธี SCI พบว่า เมื่อ $n = 10$ และ $k_\xi \ge 0.7$ พบว่า ค่า $\hat{CP}$ จะมีค่า

ใกล้เคียงกับสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ณ ระดับความเชื่อมั่น 90% และ 95% แต่เมื่อขนาดตัวอย่างเพิ่มมากขึ้น พบว่า ณ ระดับความเชื่อมั่น 90% เมื่อ $n \geq 15$ และ $k_\xi \geq 0.3$ พบว่า ค่า $\hat{CP}$ จะมีค่าใกล้เคียงกับสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และ ณ ระดับความเชื่อมั่น 95% พบว่า เมื่อ $n = 15$ และ 20 และ $k_\xi \geq 0.5$ พบว่า ค่า $\hat{CP}$ จะมีค่าใกล้เคียงกับสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และหาก $n > 20$ ค่า $\hat{CP}$ จะมีค่าใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนดเมื่อ $k_\xi \geq 0.3$

ในการพิจารณาค่า AL ของช่วงความเชื่อมั่นสำหรับ β_1 ทั้ง 3 วิธี เมื่อกำหนดขนาดตัวอย่างเล็กกว่าหรือเท่ากับ 30 $(n \leq 30)$ ที่ระดับความเชื่อมั่น 90% และ 95% จะเปรียบเทียบความกว้างเฉลี่ยเมื่อ ค่า $\hat{CP}$ มีค่าใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด จากตารางที่ 2 พบว่า เมื่อ $k_\xi = 0.3, 0.5$ และ 0.7 วิธี IOLS จะมีค่า AL ที่แคบกว่าวิธี ACI และวิธี SCI ในขณะที่เมื่อ $k_\xi = 0.9$ พบว่า วิธี SCI จะเป็นวิธีที่มีค่า AL ที่แคบที่สุด แต่เมื่อขนาดตัวอย่างเพิ่มมากขึ้นและ $k_\xi = 0.9$ จะพบว่าค่า AL ของวิธี IOLS จะมีค่าใกล้เคียงกับวิธี SCI และหากพิจารณาเปรียบเทียบช่วงความเชื่อมั่นระหว่างวิธี ACI และวิธี SCI พบว่า เมื่อ $k_\xi \leq 0.5$ วิธี ACI มีความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่แคบกว่าวิธี SCI ยกเว้นกรณี $n = 15$, $k_\xi = 0.3$ ที่ระดับความเชื่อมั่น 90% แต่เมื่อ $k_\xi > 0.5$ วิธี SCI จะมีความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่แคบกว่าวิธี ACI ซึ่งสอดคล้องกับงานวิจัยของ Tsai (2013) อีกทั้งช่วงความเชื่อมั่นทั้ง 3 วิธีข้างต้นยังถูกจำกัดด้วยเงื่อนไข ดังนี้ สำหรับการสร้างช่วงความเชื่อมั่นวิธี ACI และ SCI ข้อมูลจะต้องเป็นไปตามเงื่อนไข $S_{yy}(S_{xx} - \sigma_\delta^2) - S_{xy} > 0$ และ $S_{xx} - \sigma_\delta^2 > 0$ ในขณะที่วิธี IOLS นั้นข้อมูลจะต้องเป็นไปตามเงื่อนไข $S_{xx} - \sigma_\delta^2 > 0$

สรุปผลการวิจัย

การศึกษาพัฒนาและเปรียบช่วงความเชื่อมั่นสำหรับ β_1 เมื่อ $n \leq 30$ ในการวิเคราะห์การถดถอยอย่างง่ายเมื่อมีความคลาดเคลื่อนในตัวแปร ภายใต้ข้อสมมติทราบค่า σ_δ^2 ซึ่งช่วงความเชื่อมั่นที่พัฒนาคือ ช่วงความเชื่อมั่นกำลังสองน้อยที่สุดปรับปรุง (IOLS) เปรียบเทียบกับช่วงความเชื่อมั่นเชิงเส้นกำกับ (ACI) และช่วงความเชื่อมันแซนด์วิช (SCI) โดยการจำลองข้อมูลให้สอดคล้องกับเงื่อนไข $S_{yy}(S_{xx} - \sigma_\delta^2) - S_{xy} > 0$ และ $S_{xx} - \sigma_\delta^2 > 0$ พบว่า เมื่อ $0.3 \leq k_\xi \leq 0.7$ ช่วงความเชื่อมั่นวิธี IOLS เป็นช่วงความเชื่อมั่นที่มีค่า $\hat{CP}$ ใกล้เคียงกับสัมประสิทธ์ของความเชื่อมั่นที่กำหนดและมีค่า AL แคบที่สุด ในขณะที่เมื่อ $k_\xi > 0.7$ ช่วงความเชื่อมั่นวิธี SCI เป็นช่วงความเชื่อมั่นที่มีค่า $\hat{CP}$ ใกล้เคียงกับสัมประสิทธิ์ของความเชื่อมั่นที่กำหนดและมีค่า AL แคบที่สุด แต่เมื่อ $k_\xi < 0.3$ พบว่าค่า $\hat{CP}$ ของช่วงความเชื่อมั่นทั้ง 3 วิธีมีค่าน้อยกว่าสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด เนื่องจากเมื่อค่า k_ξ มีค่าน้อยแสดงว่าเกิดความผิดพลาดในตัวแปรอิสระ X สูง (ค่า σ_δ^2 มีค่าสูง) ดังนั้นในการประมาณค่าพารามิเตอร์จึงเกิดความคลาดเคลื่อนสูงส่งผลให้ค่า $\hat{CP}$ ของช่วงความเชื่อมั่นทั้ง 3 วิธีมีค่าน้อยกว่าสัมประสิทธิ์ของความเชื่อมั่นที่กำหนด ซึ่งในการศึกษาในครั้งถัดไปควรหาแนวทางในการพัฒนาช่วงความเชื่อมั่นในกรณี $n \leq 30$ และ $k_\xi < 0.3$ เพื่อให้ได้ช่วงความเชื่อมั่นที่มีความเหมาะสมสำหรับใช้ในการประมาณค่า β_1 หรือหาแนวทางในการประมาณค่าแบบช่วงในกรณีที่ข้อมูลมีค่าผิดปกติ หรือมีค่าสูญหาย

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณ โครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่ให้ทุนสนับสนุนการศึกษาและการทำวิจัย และผู้ทรงคุณวุฒิทุกท่านที่ตรวจประเมินบทความ

เอกสารอ้างอิง

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, *31*,144–152.

Casella G, Berger RL. (2002). *Statistical Inference*. Pacific Grove, CA: Duxbury.

Cheng CL, Van Ness JW. (1999). *Statistical regression with measurement error*. London: Arnold.

Chiawkhun, P. (2007). *Regression Analysis*. Chiang Mai: Science and Technology Service Center, Faculty of Science, Chiang Mai University. (in Thai)

Dachodomphant, V. (2003). A comparison on multiple regression coefficient estimation methods with errors in independent variables. A thesis submitted in partial fulfillment of requirements of the degree of master of science in statistics, Department of statistics, Faculty of commerce and accountancy, Chulalongkorn university. (in Thai)

Fuller WA. (1987). *Measurement error models*. New York: Wiley.

Gleser LJ, Hwang JT. (1987). The Nonexistence of $100(1-\alpha)\%$ Confidence Sets of Finite Expected Diameter in Errors-in-Variables and Related Models. *Annals of Statistics*, *15*(4), 1351-1362.

Huwang, L. (1996). Asymptotically Honest Confidence Sets for Structural Errors-In-Variables Models. The *Annals of Statistics*, *24*(4), 1536-1546

Kiranandana, S. (2002). *Statistical Inference : basis theory. Bangkok:* Chulalongkorn University. (in Thai)

Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied linear statistical models*. New York: McGraw-Hill Higher Education.

Romano, J. L., Kromrey, J. D., Owens, C. M., & Scott, H. M. (2011). Confidence Interval Methods for Coefficient Alpha on the Basis of Discrete, Ordinal Response Items: Which One, if Any, is the Best?. *The Journal of Experimental Education*, *79*(4), 382-403.

Tsai J-R. (2010). Generalized confidence interval for the slope in linear measurement error model. *Journal of Statistical Computation and Simulation*, *80*(8), 927-936.

Tsai J-R. (2013). Interval estimation for fitting straight line when both variables are subject to error. *Computational Statistics*, *28*(1), 219-240.

Weerahandi, S. (1993). Generalized Confidence Intervals. *Journal of the American Statistical Association*, *88*(423), 899-905.