หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

บทสรุปและข้อเสนอแนะ

จากการศึกษาเกี่ยวกับการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีที่มีการทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์การถดถอยโดยวิธีการแบ่งขนาดข้อมูลทั้งหมดที่มีออกเป็นสองส่วนให้มีขนาดเท่ากัน โดยข้อมูลทั้งสองส่วนนั้นเป็นอิสระต่อกันและมีค่าของตัวแปรอิสระชุดเดียวกัน แล้วใช้ข้อมูลส่วนหนึ่งในการทดสอบสมมติฐาน อีกส่วนหนึ่งในการประมาณค่านั้น เมื่อพิจารณาความคลาดเคลื่อนของการประมาณสัมประสิทธิ์การถดถอยที่เกิดขึ้นจากวิธีการดังกล่าว เพื่อให้การคำนวณสะดวกขึ้นในที่นี้จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าตัวแปร Y ด้วย Y* ที่เกิดขึ้นเป็นหลัก ผลการวิเคราะห์สรุปได้ดังนี้

1. ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ Y* ที่เกิดขึ้นจากการศึกษาครั้งนี้จะมีลักษณะเช่นเดียวกับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ Y ที่เกิดขึ้นจากการใช้วิธีการโดยทั่วไปในการทดสอบสมมติฐานและประมาณค่า กล่าวคือ $M_1(\theta_1,\lambda_0)$ สำหรับ $\lambda_0\epsilon(0,+\infty)$ นั้นจะมีค่าอยู่ระหว่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการใช้ $X_2\hat{\beta}_2$ เป็นตัวประมาณของ $Y(M_1(\theta_1,\infty))$ และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดขึ้นจากการใช้ X_2b_2 เป็นตัวประมาณของ $Y(M_1(\theta_1,0)$ เมื่อ non-centrality parameter θ_1 มีค่าเป็น 0 และเมื่อ θ_1 มีค่าเพิ่มขึ้น $M_1(\theta_1,\lambda_0)$ จะมีค่าเพิ่มขึ้นและจะมีค่าเท่ากับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ X_2b_2 ที่จุด $\theta_1=\theta_c$ เมื่อ $\theta_c=\frac{m}{2}$ ซึ่งต่างกับกรณีของ $M(\theta_1,\lambda_0)$ ที่ค่า θ_c จะมีค่าอยู่ระหว่าง $\frac{m}{4}$ กับ $\frac{m}{2}$ และ $M_1(\theta_1,\lambda_0)$ จะมีค่าสูงสุดที่จุด $\theta=\theta_{MAX}$ โดย $\theta_{MAX}>\frac{m}{2}$ จากนั้นเมื่อ θ_1 มีค่าเพิ่มขึ้น $M_1(\theta_1,\lambda_0)$ จะค่อย ๆ ลดลงและเมื่อ θ_1 มีค่ามากเข้าใกล้ค่าอนันต์ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ Y* จะมีค่าเท่ากับ $k\sigma^2$ ทั้งสองกรณี

2. ถ้าให้ D เป็นผลต่างระหว่าง $M_1(\theta_1,\lambda_1)$ กับ $M(\theta_1,\lambda_1)$ จะพบว่า D จะมีค่าเป็นลบสำหรับ non-centrality parameter ที่อยู่ในช่วง (θ_L,θ_u) และในกรณีที่ θ_1 มีค่ามากเข้าใกล้ค่าอนันต์ D จะมีค่าเป็น 0 นั่นคือไม่มีค่าแตกต่างระหว่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ Y*

3. สำหรับ θ_L และ θ_u ซึ่งเป็นค่าต่ำสุดและสูงสุดของช่วงของค่า non-centrality parameter θ_1 ที่ทำให้ D มีค่าเป็นลบ หรือช่วงของ θ_1 ที่ทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังส่องของตัวประมาณค่าตัวแปร Y ในกรณีที่เราสนใจศึกษามีค่าต่ำกว่ากรณีที่ใช้วิธีการทั่วไปนั้น θ_L จะมีค่าอยู่ในช่วง $[0,\frac{m}{2})$ และ θ_u จะมีค่าอยู่ในช่วง $[\frac{m}{2},\infty)$ ซึ่งมีลักษณะดังต่อไปนี้คือ ในกรณีที่ m มีค่าคงที่ T - k มีค่าเพิ่มขึ้นค่าของ θ_L, θ_u จะลดลง และเมื่อ m มีค่าน้อยคือ $m \leq 24$ นั้น เมื่อ T - k มีค่าเพิ่มขึ้นเรื่อย ๆ θ_L จะมีค่าลดลงจนกระทั่งพบว่าเมื่อ θ_1 มีค่าเท่ากับ 0 ผลต่าง D จะมีค่าเป็นลบในกรณีนี้ θ_L จะมีค่าเป็น 0 สำหรับ θ_u เมื่อ T - k มีค่ามากและเพิ่มขึ้นเรื่อย ๆ นั้น θ_u จะมีค่าไม่ต่างกันมากนัก และจะมีค่าเข้าใกล้ $\frac{m}{2}$ สำหรับกรณีที่ T - k มีค่าคงที่และ m มีค่าเพิ่มขึ้นนั้น θ_L, θ_u จะมีค่าเพิ่มขึ้นเช่นเดียวกัน สำหรับค่า θ_L ในกรณีที่ $m \leq 16$ จะพบว่า θ_L มีค่าอยู่ภายในช่วง $[0, 0.5)$ เมื่อ m มีค่าเพิ่มขึ้นเท่ากับ 24 นั้น θ_L จะเพิ่มขึ้น แต่อย่างไรก็ตาม θ_L ยังคงมีค่าน้อยกว่า 1 ในกรณีที่ m มีค่าเท่ากับ 60 และ 120 จะพบว่า θ_L ยังคงมีค่าอยู่ระหว่างช่วง (0,4)

ในการศึกษานี้ได้พิจารณาค่าวิกฤต (λ_1)ของการทดสอบสมมติฐานแบบ F ที่ระดับความเชื่อมั่นเป็น 95% และ 99% เมื่อเปรียบเทียบค่า θ_L และ θ_u ที่ได้จากการใช้ λ_1 ที่ระดับความเชื่อมั่นต่างกัน จะพบว่าค่า θ_L ที่ได้จากการใช้ค่าวิกฤตที่ระดับความเชื่อมั่นสูง 99% โดยทั่วไปจะมีค่าต่ำกว่าที่ระดับความเชื่อมั่น 95% นั่นคือที่ระดับ m และ T - k ค่าเดียวกัน เมื่อ λ_1 มีค่าเพิ่มขึ้น จะพบว่า θ_L โดยทั่วไปจะมีค่าลดลง และ θ_u จะมีค่าเพิ่มขึ้น และได้ทำการคำนวณหาผลต่างดังกล่าวสำหรับกรณีที่จำนวนสัมประสิทธิ์การถดถอย (k) มีค่าเท่ากับจำนวนข้อจำกัดของค่าสัมประสิทธิ์การถดถอย (m) และจำนวนสัมประสิทธิ์ (k) เป็นสองเท่าของจำนวนข้อจำกัด (m) จะพบว่า θ_L และ θ_u ของทั้งสองกรณีนั้นมีค่าไม่ต่างกันมากนัก และยิ่งในกรณีที่ m มีค่ามาก และ T - k มีค่าน้อย ๆ นั้น θ_L จะมีค่าใกล้เคียงกันมาก สำหรับ θ_u ในกรณีที่ m มีค่าน้อย ค่า θ_u ของทั้งสองกรณีจะใกล้เคียงกันมาก

การศึกษานี้ชี้ให้เห็นว่าการประมาณค่าสัมประสิทธิ์การถดถอย ในกรณีที่มีการทดสอบสมมติฐานโดยใช้ข้อมูลต่างชุดกันทำการประมาณค่าและทดสอบสมมติฐานนั้น จะได้ตัวประมาณค่าสัมประสิทธิ์การถดถอย β ที่ดีกว่าตัวประมาณที่ได้จากการใช้ตัวอย่างชุดเดียวกันทดสอบสมมติฐาน

และประมาณค่าในช่วงที่ non - centrality parameter มีค่าอยู่ระหว่าง θ_L กับ θ_u ซึ่งมีลักษณะดังกล่าวข้างต้น ทั้งนี้การศึกษานี้มีข้อจำกัดว่าข้อมูลที่ใช้ทดสอบสมมติฐานและประมาณค่านั้นมีขนาดเท่ากัน เป็นอิสระต่อกัน และมีค่าของตัวแปรอิสระชุดเดียวกัน

สำหรับผู้ที่สนใจศึกษาในเรื่องนี้ อาจทำการวิจัยต่อไปได้โดยลดข้อจำกัดลง หรืออาจเปรียบเทียบวิธีการประมาณค่าทั้งสองวิธี โดยใช้ค่าอื่น ๆ เช่น average relative risk เป็นเกณฑ์ แล้วเปรียบเทียบผลที่ได้กับวิธีการที่ใช้ในการศึกษาครั้งนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย