บทที่ 5

ผลสรุปการวิเคราะห์และการอภิปรายผล

การวิจัยครั้งนี้ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อตัวประมาณสเกลเปลี่ยนไป และค่าผิดพลาดที่ใช้มีการแจกแจงแบบเบ้และการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ โดยทำการเปรียบเทียบวิธีกำลังสองน้อยที่สุด วิธี M-estimator ที่ใช้ the standard deviation the median absolute deviation และ the modified biweight A-estimator (ที่ c = 9 และ 10) เป็นตัวประมาณสเกล ซึ่งใช้เกณฑ์ความแกร่งของ Huber เมื่อ b = 2 ด้วยการพิจารณาเกณฑ์ของอัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง (RDAMSE) และอัตราส่วนผลต่างความผิดพลาดกำลังสอง (RDMSE) ในแต่ละกรณี ซึ่งสามารถสรุปผลการวิเคราะห์ได้ดังนี้

5.1 ผลสรุปของการเปรียบเทียบวิธี OLS STD MAD และ MOD (ที่ c = 9 และ 10)

จากการคำนวณหาค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธีการประมาณทั้ง 5 วิธี ซึ่งใช้เกณฑ์ความแกร่งของ Huber เมื่อ b = 2 โดยนำค่า MSE ของทั้ง 5 วิธีมาเปรียบเทียบกันด้วยเกณฑ์อัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง (RDAMSE) และอัตราส่วนผลต่างความผิดพลาดกำลังสอง (RDMSE) สามารถสรุปผลได้ดังนี้

5.1.1 เมื่อความผิดพลาดมีการแจกแจงแบบเบ้ ได้ใช้การแจกแจงแบบแกมมาไวบูลล์ และลอกนอร์มอลเป็นกรณีศึกษา โดยที่การแจกแจงแบบแกมมาและไวบูลล์กำหนดจากค่าสัมประสิทธิ์ความแปรปรวน ส่วนการแจกแจงลอกนอร์มอลกำหนดจากค่าเฉลี่ยและความแปรปรวน

1) เมื่อค่าผิดพลาดมีการแจกแจงแบบแกมมา ซึ่งพบได้ในการวิเคราะห์งานทางด้านเศรษฐศาสตร์ เช่น การกระจายรายได้ เป็นต้น ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้จะอาศัยเทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox เพื่อแปลงข้อมูลให้เข้าสู่ภาวะปกติก่อน ผลปรากฏว่าเมื่อใช้เกณฑ์การเปรียบเทียบอัตราส่วนผลต่างค่าเฉลี่ยความผิดพลาดกำลังสอง และอัตราส่วนผลต่างความผิดพลาดกำลังสอง ณ.

C.V. = 70 % ($\alpha = 2, \beta = 5$ 10 และ 15) วิธีกำลังสองน้อยที่สุด จะให้ผลในการประมาณค่าสัมประสิทธิ์การถดถอยพหุดีที่สุด รองลงมาคือวิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล และเมื่อ C.V. = 100 % ($\alpha = 1, \beta = 5$ 10 และ 15) วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล จะให้ผลในการประมาณค่าสัมประสิทธิ์การถดถอยพหุดีที่สุด รองลงมาคือวิธีกำลังสองน้อยที่สุด เมื่อใช้จำนวนตัวแปรอิสระและขนาดตัวอย่างแบบต่างๆ

2) <u>เมื่อค่าผิดพลาดมีการแจกแจงแบบไวบูลล์</u> ซึ่งพบได้ในงานทางวิศวกรรม เช่น เป็นตัวแบบสำหรับการกระจายของการเกิดข้อบกพร่อง (time to failure) ของเครื่องจักรหรืออุปกรณ์ไฟฟ้า เป็นต้น ในการประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้นั้นจะอาศัยเทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox เพื่อแปลงข้อมูลให้เข้าสู่ภาวะเสียก่อน ผลปรากฏเมื่อใช้เกณฑ์การเปรียบเทียบอัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง และอัตราส่วนผลต่างความผิดพลาดกำลังสอง เมื่อใช้ C.V. = 52 % ($\alpha = 2$, $\beta = 5$ 10 และ 15) วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล จะให้ผลในการประมาณค่าสัมประสิทธิ์การถดถอยพหุดีที่สุดแต่จะมีค่าใกล้เคียงกับวิธีกำลังสองน้อยที่สุด และเมื่อใช้ค่า C.V. = 100 % ($\alpha = 1, \beta = 5$ 10 และ 15) วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล จะให้ผลในการประมาณค่าสัมประสิทธิ์การถดถอยพหุดีที่สุด รองลงมาคือวิธีกำลังสองน้อยที่สุด เมื่อใช้จำนวนตัวแปรอิสระและขนาดตัวอย่างแบบต่างๆ

3) <u>เมื่อความผิดพลาดมีการแจกแจงแบบลอกนอร์มอล</u> ซึ่งพบในงานวิเคราะห์ทางด้านเศรษฐศาสตร์ เช่นการวิเคราะห์การกระจายรายได้ เป็นต้น ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบเบ้ จะอาศัยเทคนิคในการแปลงข้อมูลในรูปยกกำลังของ Box และ Cox เพื่อแปลงข้อมูลให้เข้าสู่ภาวะปกติก่อน ผลปรากฏว่าเมื่อใช้เกณฑ์การเปรียบเทียบอัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง และอัตราส่วนผลต่างความผิดพลาดกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุดจะให้ผลในการประมาณค่าสัมประสิทธิ์การถดถอยพหุดีที่สุด แต่มีค่าใกล้เคียงกับวิธีที่เหลือ เมื่อใช้จำนวนตัวแปรอิสระและขนาดตัวอย่างแบบต่างๆ

5.1.2 <u>เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ</u> ได้ใช้การแจกแจงแบบปกติปลอมปนและแบบที เป็นกรณีศึกษา โดยที่การแจกแจงปกติปลอมปนสามารถกำหนดจากค่าสเกลแฟกเตอร์และเปอร์เซนต์การปลอมปน ส่วนการแจกแจงแบบทีกำหนดจากระดับความเป็นเสรี และเปรียบเทียบผลการประมาณของวิธีทั้ง 5 วิธีด้วยเกณฑ์อัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง (RDAMSE) และอัตราส่วนผลต่างความผิดพลาดกำลังสอง (RDMSE) ซึ่งให้ผลสรุปดังนี้

1) <u>เมื่อความผิดพลาดมีการแจกแจงแบบปกติปลอมปน</u> วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล สามารถประมาณค่าสัมประสิทธิ์การถดถอยพหุได้ดีที่สุด รองลงมาคือวิธีกำลังสองน้อยที่สุด วิธี M-estimator ที่ใช้ the modified biweight A-estimator ที่ c = 10 และ 9 เป็นตัวประมาณสเกล และวิธี M-estimator ที่ใช้ the standard deviation ของ ϵ_i เป็นตัวประมาณสเกล ให้ผลใกล้เคียงกันซึ่งสามารถเรียงลำดับปัจจัยที่มีอิทธิพลกับค่าผิดปกติที่เกิดขึ้นจากมากไปน้อยดังนี้

1.1) สเกลแฟกเตอร์ (Scale) คือ ณ. สเกลแฟกเตอร์ 15 10 และ 5

1.2) เปอร์เซนต์การปลอมปน (p) คือ ณ. เปอร์เซนต์การปลอมปน 30 20 10 และ 5

1.3) จำนวนตัวแปรอิสระ (m) และขนาดตัวอย่าง (n) คือ m = 3 และ n = 20 , m = 5 10 และ n = 150 100 และ 50

2) <u>เมื่อความผิดพลาดมีการแจกแจงแบบที</u> วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล สามารถประมาณค่าสัมประสิทธิ์การถดถอยพหุได้ดีที่สุด รองลงมาคือวิธีกำลังสองน้อยที่สุด และวิธี M-estimator ที่ใช้ the modified biweight A-estimator ที่ c = 9 ตามลำดับ โดยเรียงลำดับปัจจัยที่มีอิทธิพลต่อการประมาณค่าสัมประสิทธิ์การถดถอยพหุที่เกิดขึ้นจากมากไปน้อยดังนี้

2.1) ระดับความเป็นเสรี (DF) คือระดับความเป็นเสรีที่ DF = 4 และ 8

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.2) จำนวนตัวแปรอิสระ (m) และขนาดตัวอย่าง (n) คือ m = 3 5 และ 10 ตามลำดับ เมื่อขนาดตัวอย่าง n = 20

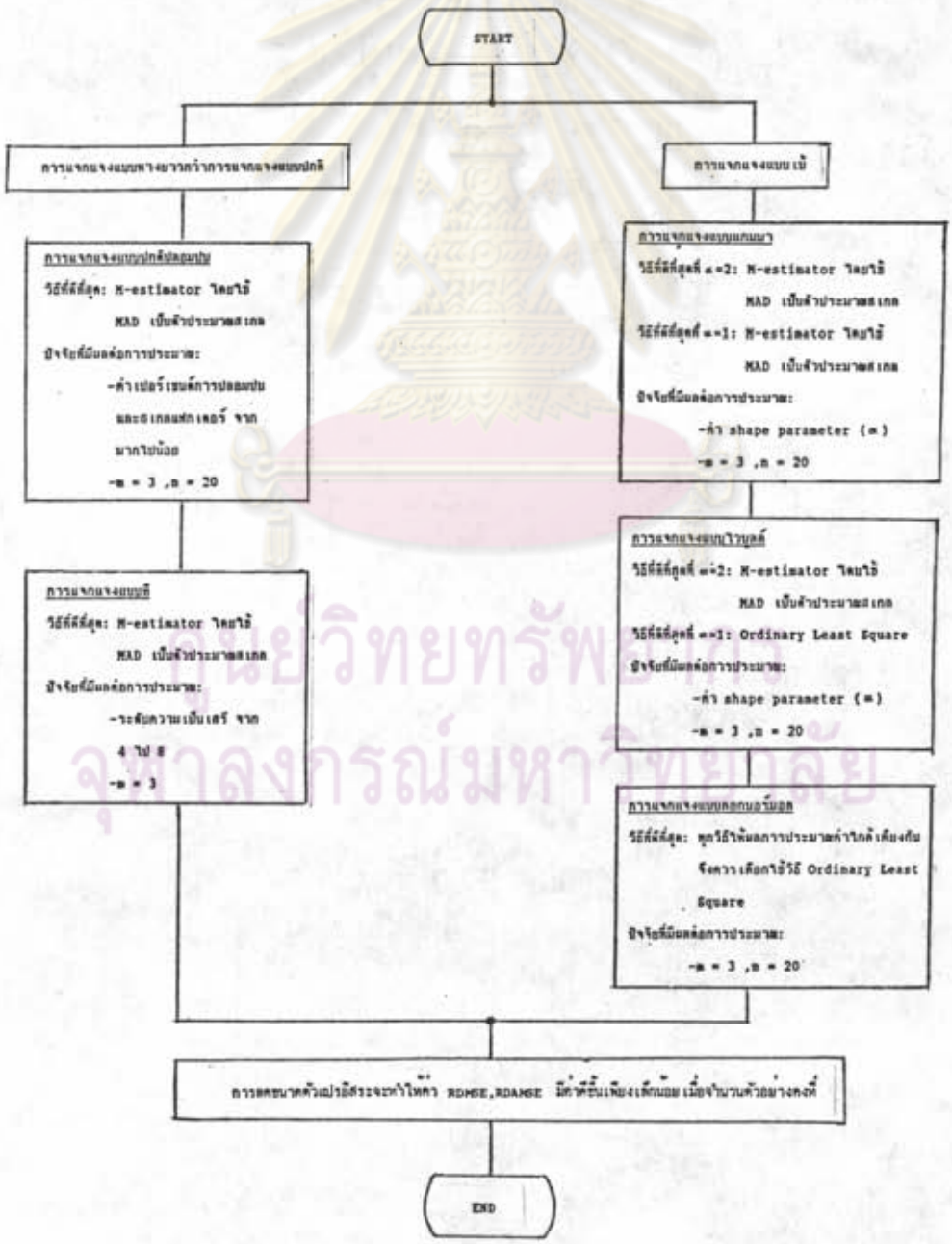
สำหรับจำนวนตัวแปรอิสระและขนาดตัวอย่างมีอิทธิพลไม่มากนักในการเลือกใช้วิธีการประมาณค่าสัมประสิทธิการถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ และการแจกแจงแบบเบ้ เนื่องจากวิธีกำลังสองน้อยที่สุดและวิธี M-estimator ที่ใช้ตัวประมาณสเกลดังกล่าวจะพิจารณาถึงค่าผิดพลาดที่เกิดขึ้นในตัวระบบเชิงเส้น โดยที่ตัวแปรอิสระไม่เกิด Multicolinearity ซึ่งกันและกัน ดังนั้นค่าผิดพลาดทั้งหมดจึงมาจากตัวระบบเชิงเส้นอย่างเดียวเท่านั้น

5.2 การอภิปรายผล

จากผลการวิเคราะห์นี้เมื่อพิจารณาถึง MSE ของวิธีการประมาณสัมประสิทธิ์การถดถอยพหุทุกวิธี เมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้โดยใช้เกณฑ์การแจกแจงแบบแกมมา ไวบูลย์ และลอกนอร์มอล เป็นกรณีศึกษา ผลปรากฏว่าการแจกแจงที่มีรูปร่างแบบเบ้ดังกล่าวหลังจากอาศัยเทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox สำหรับการแปลงข้อมูลให้เข้าสู่ภาวะปกติแล้ว การแจกแจงแบบแกมมาและไวบูลล์ที่ C.V. มี $\alpha = 1$ วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล สามารถใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุได้ดีที่สุด รองลงมาคือวิธีกำลังสองน้อยที่สุด และเมื่อ C.V. มี $\alpha = 2$ สำหรับการแจกแจงแบบแกมมา วิธีกำลังสองน้อยที่สุดจะให้ผลในการประมาณค่าสัมประสิทธิ์ดีที่สุด แต่เมื่อค่าผิดพลาดมีการแจกแจงแบบไวบูลล์ วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกลจะให้ผลในการประมาณค่าสัมประสิทธิ์ดีที่สุดและจะมีค่าใกล้เคียงกับวิธีกำลังสองน้อยที่สุด ส่วนการแจกแจงแบบลอกนอร์มอล วิธีกำลังสองน้อยที่สุดจะให้ผลในการประมาณค่าสัมประสิทธิ์การถดถอยพหุได้ดีที่สุด และเมื่อค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติโดยใช้เกณฑ์การแจกแจงแบบปกติปลอมปนและการแจกแจงแบบที เป็นกรณีศึกษา ผลปรากฏว่าการแจกแจงของค่าผิดพลาดที่มีค่าผิดปกติเกิดขึ้นในรูปแบบต่างๆ (ขึ้นกับเปอร์เซนต์การปลอมปน (p) สเกลแฟกเตอร์ (c) และระดับความเป็นเสรี (DF)) ที่มีทั้งหางยาวมากและหางยาวเล็ก

น้อย วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกลจะให้ผลในการประมาณค่าสัมประสิทธิ์การถดถอยพหุดีที่สุด และค่าเปอร์เซนต์การปลอมปน (p) สเกลแฟกเตอร์ (c) และระดับความเป็นเสรี (DF) ค่อนข้างมีอิทธิพลต่อการประมาณค่าสัมประสิทธิ์การถดถอยพหุมากกว่าจำนวนตัวแปรอิสระและขนาดตัวอย่างที่เปลี่ยนแปลง

ตารางที่ 5.2.1 แสดงผลสรุปการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุของค่าผิดพลาดของการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติและการแจกแจงแบบเบ้



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

5.3 ข้อเสนอแนะ

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อตัวประมาณสเกลเปลี่ยนไป และค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ ควรเลือกใช้วิธี M-estimator ที่ใช้ the median absolute deviation เป็นตัวประมาณสเกล เมื่อใช้เกณฑ์ความแกร่งของ Huber ณ. b = 2 เนื่องจากสามารถลดอิทธิพลของค่าผิดพลาดปกติลงได้

และในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อตัวประมาณสเกลเปลี่ยนไป และค่าผิดพลาดมีการแจกแจงแบบเบ้ควรใช้เทคนิคการแปลงข้อมูลที่อยู่ในรูปยกกำลังของ Box และ Cox ในการแปลงข้อมูลให้เข้าสู่ภาวะปกติเสียก่อน แล้วทำการตรวจสอบว่าข้อมูลหลังจากการแปลงมีการแจกแจงเป็นแบบปกติหรือไม่ ถ้ามีการแจกแจงเป็นแบบปกติจึงเลือกใช้วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุทั้ง 3 วิธี ซึ่งจะให้ผลลัพธ์ใกล้เคียงกัน ดังนั้นการใช้วิธีกำลังสองน้อยที่สุดจะทำได้สะดวกและรวดเร็วกว่าและขอเสนอแนะดังนี้

1) นอกจากวิธี M-estimator ที่ใช้ตัวประมาณสเกลดังกล่าว เมื่อใช้เกณฑ์ความแกร่งของ Huber ที่ b = 2 แล้วยังสามารถศึกษาถึงวิธี M-estimator แบบอื่นที่นิยมใช้อีกได้แก่ วิธี Huber's Proposal 2 วิธี One-step Huber วิธี simultaneous estimation of location and scale เป็นต้น รวมทั้งการใช้ตัวสเกลแบบเฉพาะเช่น Olshen's estimate ที่ใช้เกณฑ์ความแกร่ง $\rho(x) = x/(\mu + x^2)$ ซึ่งมีตัวประมาณสเกลเฉพาะ $S(F) = F^{-1}(3/4) - F^{-1}(1/4)$ เป็นต้น

2) ผู้สนใจควรจะศึกษาตัวประมาณสเกลแบบอื่นที่นิยมใช้อีก เช่น the modified sine A-estimator และ the trimmed standard deviation เป็นต้น

3) นอกจากนี้อาจศึกษาค่าผิดพลาดที่มีการแจกแจงแบบเบ้ และการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ เช่นการแจกแจงแบบเบตา (Beta distribution) เป็นต้น

4) จากข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลเชิงปริมาณ เมื่อข้อมูลมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ และการแจกแจงแบบเบ้ซึ่งใช้ตัวประมาณที่แกร่งชนิด M (M-estimator) ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ ซึ่งผู้ที่สนใจอาจศึกษาตัวประมาณค่าสัมประสิทธิ์การถดถอยพหุที่ใช้ตัวประมาณความแกร่งชนิด L (L-estimator) และตัวประมาณความแกร่งชนิด R (R-estimator)