ความแกร่งในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณสำหรับค่าผิดปกติในตัวแปรตาม

Robust Multiple Linear Regression for Outliers In Dependent Variables

กัลยา บุญหล้า[1*] และ เมทินี ชมภูสว่าง[1]
Kanlaya Boonlha[1] and Maytinee Chompoosawang[1]
[1]ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร
[1]Department of Mathematics, Faculty of Science, Naresuan University
วันที่ส่งบทความ : 4 พฤศจิกายน 2563 วันที่แก้ไขบทความ : 11 มีนาคม 2564 วันที่ตอบรับบทความ : 23 เมษายน 2564
Received: 4 November 2020, Revised: 11 March 2021, Accepted: 23 April 2021

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบการถดถอยเชิงเส้นพหุคูณเมื่อมีค่าผิดปกติในตัวแปรตาม ด้วยวิธีการประมาณค่า 3 วิธี คือ วิธีกำลังสองน้อยที่สุด (OLS) วิธี M เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Andrews (M – Andrews) และ วิธี GM เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Huber (GM – Huber) โดยจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ทำซ้ำ 1,000 ครั้งสำหรับแต่ละสถานการณ์ ด้วยโปรแกรม R เวอร์ชัน 3.6.1 สำหรับการจำลองข้อมูลกำหนดให้มีตัวแปรอิสระ 3 ตัว และกำหนดค่า $\beta_0 = 0, 1, -1$ พิจารณาที่ $\beta_1 = \beta_2 = \beta_3 = 1$ และ $\beta_1 = \beta_2 = \beta_3 = -1$ ร้อยละค่าผิดปกติเท่ากับ 0, 10, 20 และ 30 และขนาดตัวอย่าง คือ 20, 30, 50 และ 100 รวมทั้งสิ้น 120 สถานการณ์ ทั้งนี้เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ผลการวิจัยพบว่า กรณีไม่มีค่าผิดปกติ พบว่าวิธี OLS และวิธี M – Andrews ให้ค่า MSE ต่ำที่สุด ส่วนกรณีมีค่าผิดปกติทุก สถานการณ์วิธี GM - Huber ให้ค่า MSE ต่ำที่สุด ยกเว้นขนาดตัวอย่างเท่ากับ 20 และ 30 เมื่อร้อยละค่าผิดปกติเท่ากับ 30 วิธี M - Andrews และวิธี OLS ให้ค่า MSE ต่ำที่สุด

คำสำคัญ : การถดถอยเชิงเส้นพหุคูณ ค่าผิดปกติ ตัวประมาณ M-Andrews ตัวประมาณ GM-Huber

Abstract

The purpose of this study was to compare the methods of regression coefficient estimation when data contain outliers in dependent variable in the multiple linear regression models for 3 methods, ordinary least squares method (OLS), M method using

---

*ที่อยู่ติดต่อ E-mail address: kanlayab@nu.ac.th

Andrews weight function (M - Andrews) and GM method using Huber weight function (GM - Huber). Simulated data by Monte Carlo technique, repeated 1,000 times for each situation with R programming version 3.6.1. For simulated data, 3 independent variables with $\beta_0 = 0, 1, -1$ when $\beta_1 = \beta_2 = \beta_3 = 1$ and $\beta_1 = \beta_2 = \beta_3 = -1$. The percentage of outliers was 0, 10, 20 and 30, and sample size was 20, 30, 50 and 100 total 120 situations. The criterion of comparison was mean square error (MSE). The results showed that in case of no outliers on the data, OLS and M - Andrews was mostly the lowest of MSE. When data contain outliers in dependent variable the GM – Huber provided generally the lowest of MSE in all situations except the sample size is 20 and 30 when percentage of outlier equal 30, the M - Andrews and OLS was the lowest of MSE.

Keywords: Multiple Linear Regression, Outliers, M-Andrews estimator, GM-Huber estimator

1. บทนำ

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Variable) กับตัวแปรอิสระ (Independent Variable) [1] โดยการวิเคราะห์การถดถอยเชิงเส้นพหุคูณเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระตั้งแต่สองตัวขึ้นไปกับตัวแปรตามหนึ่งตัว

ในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณด้วยวิธีกำลังสองน้อยที่สุดถ้าเป็นไปตามข้อกำหนดสมมติพื้นฐานของความคลาดเคลื่อน คือ ความคลาดเคลื่อนเป็นอิสระต่อกันการมีการแจกแจงปกติค่าเฉลี่ยเท่ากับศูนย์ ความแปรปรวนคงที่ จะทำให้การประมาณค่าการถดถอยเชิงเส้นพหุคูณ ด้วยวิธีกำลังสองน้อยที่สุด มีคุณสมบัติเป็นตัวประมาณค่าเชิงเส้นที่ไม่เอนเอียงที่ดีที่สุด (Best Linear Unbiased Estimator: BLUE) แต่ในทางปฏิบัติความคลาดเคลื่อนอาจมีสาเหตุที่ทำให้ไม่เป็นไปตามข้อกำหนดสมมติพื้นฐานของการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ เช่น การเกิดค่าผิดปกติ (Outlier) ซึ่งค่าผิดปกติ คือค่าของตัวแปรตามหรือตัวแปรอิสระบางตัวที่มีค่าสูงผิดปกติหรือต่ำผิดปกติจากค่าสังเกตส่วนใหญ่ในชุดข้อมูล เมื่อเกิดค่าผิดปกติในชุดข้อมูลตัวประมาณกำลังสองน้อยที่สุดอาจจะไม่เหมาะสม เนื่องจากวิธีการนี้มีความไวต่อค่าผิดปกติ ทำให้เส้นการถดถอยที่ได้เบี่ยงเบนไปจากข้อมูลกลุ่มใหญ่ เนื่องจากเส้นสมการถดถอยที่ได้ถูกปรับทิศทางไปตามข้อมูลที่ผิดปกติส่งผลให้ความแปรปรวนของมีค่าสูงขึ้นจึงทำให้ค่าสัมประสิทธิ์การถดถอยที่ได้มีค่าไม่เหมาะสม

ดังนั้นเมื่อเกิดปัญหาจากการวิเคราะห์ข้อมูลด้วยวิธีกำลังสองน้อยที่สุด การวิเคราะห์การถดถอยที่มีความแกร่ง (Robust) จึงเป็นทางเลือกหนึ่งในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณ ซึ่งเป็นการวิเคราะห์การถดถอยที่ใช้หลักการลดอิทธิพลของข้อมูลที่มีค่าผิดปกติลง โดยนักสถิติหลายท่านได้ทำการคิดวิธีการประมาณค่าสัมประสิทธิ์การถดถอยที่มีความแกร่งหลายวิธี เช่น Huber [2] ได้ศึกษาวิธีการหาตัวประมาณแบบแกร่งโดยเรียกว่าวิธี M เป็นวิธีที่นำค่าฟังก์ชันที่ถูกเลือกอย่างเหมาะสมมาแทนที่ค่าความคลาดเคลื่อนกำลังสองในวิธีกำลังสองน้อยที่สุดและวิธี M ยังสามารถหาตัวประมาณค่าสัมประสิทธิ์การ

ถดถอยเชิงเส้นพหุคูณได้หลายรูปแบบขึ้นอยู่กับฟังก์ชันถ่วงน้ำหนักที่แตกต่างกันไป ฟังก์ชันถ่วงน้ำหนักที่เป็นที่นิยมได้แก่ของ Andrews (Holland and Welsch, 1977) [3] ต่อมา Simpsona and Montgomery [4] ได้พัฒนาวิธี GM มาจากวิธี M โดยทำการถ่วงน้ำหนักเพิ่มเข้าไปในฟังก์ชันของค่าความคลาดเคลื่อนที่ถูกเลือกให้เหมาะสม โดยมีหลักการคือหาตัวประมาณค่าสัมประสิทธิ์การถดถอยที่ทำให้ผลรวมของฟังก์ชันของค่าความคลาดเคลื่อนที่ถูกเลือกให้เหมาะสมมีค่าน้อยที่สุด

อรพรรณ ตันตระกูล และคณะ [5] ได้ศึกษาเรื่องการเปรียบเทียบวิธีการถดถอยที่มีค่าผิดปกติ ในตัวแปรอิสระและตัวแปรตามสำหรับตัวแบบการถดถอยเชิงเส้นแบบพหุคูณโดยทำการเปรียบเทียบวิธีประมาณค่า 3 วิธี คือ วิธีกำลังสองน้อยที่สุด วิธีกำลังสองน้อยที่สุดที่ทำการปรับแต่ง (LTS) วิธีตัวประมาณ GM - Huber เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ผลการศึกษาพบว่า วิธีตัวประมาณ GM - Huber ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดและยังพบว่าปัจจัยที่มีผลต่อประสิทธิภาพของการประมาณค่าพารามิเตอร์ ทั้ง 3 วิธี ได้แก่ องศาความเป็นอิสระ ขนาดตัวอย่าง และอัตราส่วนค่าผิดปกติของตัวแปรอิสระจะแปรผกผันกับค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง

จากการศึกษาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณ เมื่อข้อมูลมีค่าผิดปกติ ที่กล่าวมา เมื่อข้อมูลมีค่าผิดปกติในตัวแปรตามเป็นสถานการณ์ที่อาจพบได้บ่อยครั้งในการวิเคราะห์ข้อมูล ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบการถดถอยเชิงเส้นพหุคูณ เมื่อข้อมูลมีค่าผิดปกติในตัวแปรตาม โดยใช้วิธี OLS วิธี M - Andrews และวิธี GM - Huber ภายใต้สถานการณ์ขนาดตัวอย่างเป็น 20, 30, 50 และ 100 ซึ่งเป็นขนาดตัวอย่างที่อาจพบเห็นได้บ่อย ๆ และกำหนดจำนวนตัวแปรอิสระที่ใช้ในการศึกษามี 3 ตัว ค่าสัมประสิทธิ์การถดถอย $\beta_0 = 0, 1, -1$ และ $\beta_1, \beta_2, \beta_3$ ให้มีทิศทางเดียวกันและมีทิศทางตรงกันข้าม ซึ่งแตกต่างจากการศึกษาของอรพรรณ ตันตระกูล และคณะ [5] โดยศึกษาข้อมูลที่จำลองด้วยเทคนิคมอลติคาร์โล (Monte Carlo Simulation Technique) และเปรียบเทียบประสิทธิภาพของทั้ง 3 วิธีด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอย ซึ่งวิธีการประมาณค่าสัมประสิทธิ์การถดถอยวิธีใดมีค่า MSE ต่ำที่สุด ถือว่าวิธีนั้นมีประสิทธิภาพในการประมาณค่ามากที่สุด

2. วิธีการดำเนินงานวิจัย

การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Method) โดยใช้โปรแกรม R เวอร์ชั่น 3.6.1 ทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์

1. สร้างข้อมูลตัวแปรอิสระ X_1, X_2 และ X_3 มีการแจกแจงแบบเอกรูป (Uniform Distribution) ดังนี้ $X_1 \sim U(-1,1)$, $X_2 \sim U(-3,3)$ และ $X_3 \sim U(-5,5)$ กำหนดให้ตัวแปรอิสระแต่ละตัวไม่มีค่าผิดปกติ โดยค่าของตัวแปรอิสระอยู่ในช่วง $[Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$

2. กำหนดข้อมูลของความคลาดเคลื่อนให้มีการแจกแจงแบบปกติ (Normal Distribution) $\varepsilon_i \sim N(0,1)$

3. สร้างตัวแปรตาม (Y_i) โดยมีรูปแบบดังนี้

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad \text{เมื่อ } i = 1,2,\ldots,n \tag{1}$$

โดยที่ $n = 20, 30, 50, 100$ และค่าสัมประสิทธิ์การถดถอย $\beta_0 = 0, 1, -1$ และ $\beta_1, \beta_2, \beta_3$ ให้มีทิศทางเดียวกันและมีทิศทางตรงกันข้าม กำหนดให้ตัวแปรตามมีค่าผิดปกติระดับไม่รุนแรง ตามเงื่อนไขค่าผิดปกติแผนภาพกล่องคือค่าตัวแปรตามอยู่ในช่วง $[Q_1 - 3(IQR), Q_1 - 1.5(IQR)]$ หรือ $[Q_3 + 1.5(IQR), Q_3 + 3(IQR)]$ มีร้อยละค่าผิดปกติ คือ 0, 10, 20 และ 30 ของแต่ละขนาดตัวอย่าง

4. ประมาณค่าสัมประสิทธิ์การถดถอยที่ต้องการศึกษา โดยใช้วิธีตัวประมาณกำลังสองน้อยที่สุด วิธี M เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Andrews และวิธี GM เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Huber ดังนี้

4.1 วิธีกำลังสองน้อยที่สุด

วิธีกำลังสองน้อยที่สุด ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ β ดังนี้

$$\hat{\beta}_{OLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{2}$$

เมื่อ Y คือ เมทริกซ์ของตัวแปรตาม ขนาด $n \times 1$

X คือ เมทริกซ์ของตัวแปรอิสระ ขนาด $n \times (k+1)$

โดยที่ $\underset{(n\times1)}{\boldsymbol{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $\underset{(n\times p)}{\boldsymbol{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$, $\underset{(p\times1)}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$

4.2 วิธี M เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Andrews

วิธี M เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Andrews หรือวิธี M – Andrews ได้ถูกพัฒนาขึ้นในปี ค.ศ. 1964 Huber ได้ศึกษาวิธีการหาตัวประมาณแบบแกร่งที่เรียกว่า วิธี M โดยมีพื้นฐานมาจากตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator) หลักการของวิธี M คือการประมาณค่าสัมประสิทธิ์การถดถอยที่ทำให้ผลรวมของฟังก์ชัน ρ มีค่าน้อยที่สุด เมื่อ ρ เป็นฟังก์ชันของค่าคลาดเคลื่อนที่ถูกเลือกอย่างเหมาะสมเพื่อมาแทนที่ค่าคลาดเคลื่อนกำลังสอง และในปี ค.ศ. 2012 Montgomery, Peck และ Vining [6] ได้ศึกษาวิธี M เพิ่มเติมจาก Huber ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ (β) ดังนี้

$$\hat{\beta}_{MA} = \left(\boldsymbol{X}'\boldsymbol{W}^{(m)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}^{(m)}\boldsymbol{Y} \tag{3}$$

เมื่อ $\boldsymbol{Y}$ คือ เมทริกซ์ของตัวแปรตาม ขนาด $n \times 1$

$\boldsymbol{X}$ คือ เมทริกซ์ของตัวแปรอิสระ ขนาด $n \times (k+1)$

$\boldsymbol{W}^{(m)}$ คือ เมทริกซ์ทแยงมุม ขนาด $n \times n$ ที่มี w_i เป็นสมาชิกในแนวทแยงมุม โดยที่ w_i เป็นฟังก์ชันถ่วงน้ำหนักของ Andrews

สามารถเขียนสรุปขั้นตอนการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ β ดังนี้

ขั้นตอนที่ 1 คำนวณค่าสัมประสิทธิ์การถดถอยเริ่มต้นของ β จากวิธี OLS

ขั้นตอนที่ 2 หาค่า $\boldsymbol{e}_i^{(1)}$ และ $s^{(1)}$ จาก

$$\boldsymbol{e}_i^{(1)} = \boldsymbol{Y} - \boldsymbol{X}\hat{\beta}^{(0)} \qquad \text{โดยที่ } i = 1,2,\dots,n \tag{4}$$

$$s^{(1)} = \frac{median\left|e_i^{(1)} - median\left(e^{(1)}\right)\right|}{0.6745} \tag{5}$$

ขั้นตอนที่ 3 นำค่า $\boldsymbol{e}_i^{(1)}$ และ $s^{(1)}$ จากขั้นตอนที่ 2 มาคำนวณ z ในรอบที่ 1 $\boldsymbol{z_i}^{(1)} = \frac{\boldsymbol{e}_i^{(1)}}{s^{(1)}}$ ดังนี้

ฟังก์ชันถ่วงน้ำหนักของ Andrews (Montgomery, Pack & Vining, 2012) [6] กำหนดดังนี้ ให้ $w(z) = \frac{\Psi(z)}{\left(\frac{z}{c}\right)}$ จะได้ฟังก์ชัน $w(z)$ ดังนี้

$$w(z) = \begin{cases} \dfrac{sin\left(\frac{z}{c}\right)}{\left(\frac{z}{c}\right)} & ; |z| \le c\pi \\ 0 & ; |z| > c\pi \end{cases} \tag{6}$$

เมื่อ z คือค่าคลาดเคลื่อนมาตรฐานที่คำนวณได้จาก $\boldsymbol{z}_i = \frac{\boldsymbol{e}_i}{s}$ และ c มีค่าเท่ากับ 1.339

ขั้นตอนที่ 4 จากน้ำหนักของแต่ละค่าสังเกตที่ได้จากขั้นตอนที่ 3 นำมาคำนวณค่าประมาณสัมประสิทธิ์การถดถอยของ β ในรอบที่ 1 ดังนี้ $\hat{\beta}^{(1)} = \left(\boldsymbol{X}'\boldsymbol{W}^{(1)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}^{(1)}\boldsymbol{Y}$

โดยที่ $\boldsymbol{W}^{(1)}$เป็นเมทริกซ์ทแยงมุมของน้ำหนักในรอบที่ 1 ขนาด $n \times n$

$$\boldsymbol{W}^{(1)} = \begin{bmatrix} w_1^{(1)}(z) & 0 & \cdots & 0 \\ 0 & w_2^{(1)}(z) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n^{(1)}(z) \end{bmatrix}_{n\times n}$$

เมื่อ $w_i(z)$ เป็นสมาชิกในแนวทแยงมุม เมื่อ $i = 1,2,\dots,n$ โดยที่ $w_i(z)$ เป็นฟังก์ชันถ่วงน้ำหนัก

ขั้นตอนที่ 5 หาค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณสัมประสิทธิ์การถดถอยเริ่มต้นกับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ 1 นั่นคือ $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(0)}\right|$

ขั้นตอนที่ 6 ถ้า $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(0)}\right|$ จากขั้นตอนที่ 5 มีค่ามากกว่า 0.001 ให้ไปทำซ้ำรอบที่ m ในขั้นตอนที่ 7 ต่อไป แต่ถ้า $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(0)}\right|$ จากขั้นตอนที่ 5 พบว่าค่าประมาณสัมประสิทธิ์ทุกค่าน้อยกว่าหรือเท่ากับ 0.001 จะได้ค่าประมาณสัมประสิทธิ์การถดถอยจากรอบที่ 1 สิ้นสุดการคำนวณ

ขั้นตอนที่ 7 การทำซ้ำรอบที่ m เมื่อ $m = 2,3,\dots$ เพื่อหาค่า $\boldsymbol{e}_i^{(m)}$ และ $s^{(m)}$ จาก

$$\boldsymbol{e}_i^{(m)} = \boldsymbol{Y} - \boldsymbol{X}\hat{\beta}^{(m-1)} \quad \text{โดยที่ } i = 1,2,\dots,n$$

$$s^{(1)} = \frac{median\left|\boldsymbol{e}_i^{(1)} - median\left(e^{(1)}\right)\right|}{0.6745}$$

4.3 วิธี GM เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Huber

ในปี ค.ศ. 1996 Simpsona and Montgomery ได้พัฒนาวิธี GM มาจากวิธี M โดยการถ่วงน้ำหนักเพิ่มเข้าไปในฟังก์ชันของค่าคลาดเคลื่อนที่ถูกเลือกให้เหมาะสม (ρ) มีหลักการ คือ หาตัวประมาณค่าสัมประสิทธิ์การถดถอยที่ทำให้ผลรวมของฟังก์ชันค่าคลาดเคลื่อนที่ถูกเลือกให้เหมาะสม (ρ) มีค่าน้อยที่สุด และในปี ค.ศ. 2012 Montgomery, Peck และ Vining [6] วิธี GM เมื่อใช้ฟังก์ชันถ่วงน้ำหนักของ Huber ประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ β ดังนี้

$$\hat{\beta}_{GM} = \left(\boldsymbol{X}\boldsymbol{W}^{(m)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{W}^{(m)}\boldsymbol{Y} \tag{7}$$

เมื่อ $\boldsymbol{Y}$ คือ เมทริกซ์ของตัวแปรตาม ขนาด $n \times 1$

$\boldsymbol{X}$ คือ เมทริกซ์ของตัวแปรอิสระ ขนาด $n \times (k+1)$

$\boldsymbol{W}^{(m)}$ คือ เมทริกซ์ทแยงมุม ขนาด $n \times n$ ที่มี w_i เป็นสมาชิกในแนวทแยงมุม โดยที่ w_i เป็นฟังก์ชันถ่วงน้ำหนักของ Huber

สามารถเขียนสรุปขั้นตอนการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุ β ดังนี้

ขั้นตอนที่ 1 คำนวณค่าสัมประสิทธิ์การถดถอยเริ่มต้นของ $\hat{\beta}$ จากวิธี OLS

ขั้นตอนที่ 2 หาค่า $\boldsymbol{e}_i^{(1)}$, $s^{(1)}$และ $\boldsymbol{\pi}_i^{(1)}$

$$\boldsymbol{\pi}_i^{(1)} = \frac{median|z_i{}^{(1)}|}{\boldsymbol{z}_i{}^{(1)}} \quad \text{เมื่อ } \boldsymbol{z}_i{}^{(1)} = \frac{\boldsymbol{e}_i^{(1)}}{s^{(1)}} \tag{8}$$

ขั้นตอนที่ 3 นำค่า $\boldsymbol{e}_i^{(1)}$ และ $s^{(1)}$จากขั้นตอนที่ 2 มาคำนวณเพื่อหาน้ำหนักของแต่ละค่าสังเกต

ฟังก์ชันถ่วงน้ำหนักของ Huber (Holland & Welsch, 1977) [3] กำหนดดังนี้

ให้ $w(u) = \frac{\Psi(u)}{u}$ จะได้ฟังก์ชัน $w(u)$ ดังนี้

$$w(u) = \begin{cases} \frac{-c}{u} & ; u < -c \\ 1 & ; -c \le u \le c \\ \frac{c}{u} & ; u > c \end{cases} \tag{9}$$

เมื่อ u คือค่าคลาดเคลื่อนมาตรฐานที่คำนวณได้จาก $u = \frac{e_i}{\boldsymbol{\pi}_i s}$ และ c มีค่าเท่ากับ 1.345

ขั้นตอนที่ 4 จากน้ำหนักของแต่ละค่าสังเกตที่ได้จากขั้นตอนที่ 3 นำมาคำนวณค่าประมาณสัมประสิทธิ์การถดถอยของ β ในรอบที่ 1 ดังนี้ $\hat{\beta}^{(1)} = \left(\boldsymbol{X}\boldsymbol{W}^{(1)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{W}^{(1)}\boldsymbol{Y}$

โดยที่ $\boldsymbol{W}^{(1)}$ เป็นเมทริกซ์ทแยงมุมของน้ำหนักในรอบที่ 1 ขนาด $n \times n$

$$\boldsymbol{W}^{(1)} = \begin{bmatrix} w_1^{(1)}(u) & 0 & \cdots & 0 \\ 0 & w_2^{(1)}(u) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n^{(1)}(u) \end{bmatrix}_{n \times n}$$

เมื่อ $w_i(u)$ เป็นสมาชิกในแนวทแยงมุม เมื่อ $i = 1,2,\ldots,n$ โดยที่ $w_i(u)$ เป็นฟังก์ชันถ่วงน้ำหนัก

ขั้นตอนที่ 5 หาค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณสัมประสิทธิ์การถดถอยเริ่มต้นกับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ 1 นั่นคือ $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(0)}\right|$

ขั้นตอนที่ 6 ถ้า $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(0)}\right|$ จากขั้นตอนที่ 5 มีค่ามากกว่า 0.001 ให้ไปทำซ้ำรอบที่ m ในขั้นตอนที่ 7 ต่อไป แต่ถ้า $\left|\hat{\beta}^{(1)} - \hat{\beta}^{(0)}\right|$ จากขั้นตอนที่ 5 พบว่าค่าประมาณสัมประสิทธิ์ทุกค่าน้อยกว่าหรือเท่ากับ 0.001 จะได้ค่าประมาณสัมประสิทธิ์การถดถอยจากรอบที่ 1 สิ้นสุดการคำนวณ

ขั้นตอนที่ 7 การทำซ้ำรอบที่ m เมื่อ $m = 2,3,\ldots$ เพื่อหาค่า $\boldsymbol{e}_i^{(m)}$ และ $s^{(m)}$ จาก

$$\boldsymbol{e}_i^{(m)} = \boldsymbol{Y} - \boldsymbol{X}\hat{\beta}^{(m-1)} \quad \text{โดยที่ } i = 1,2,\ldots,n$$

$$s^{(m)} = \frac{median\left|\boldsymbol{e}_i^{(m)} - median(e^{(m)})\right|}{0.6745}$$

$$\boldsymbol{\pi}_i^{(m)} = \frac{median|z_i{}^{(m)}|}{\boldsymbol{z}_i{}^{(m)}} \quad \text{เมื่อ } \boldsymbol{z}_i{}^{(m)} = \frac{\boldsymbol{e}_i^{(m)}}{s^{(m)}}$$

5. คำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธีภายใต้สถานการณ์ต่าง ๆ โดยการทำซ้ำ 1,000 รอบ

6. เปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าในแต่ละสถานการณ์โดยกำหนดค่าสัมประสิทธิ์การถดถอยดังนี้ กรณี $\beta_0 = 1$ พิจารณาที่ $\beta_1 = \beta_2 = \beta_3 = 1$ และ $\beta_1 = \beta_2 = \beta_3 = -1$ กรณี $\beta_0 = 0$ พิจารณาที่ $\beta_1 = \beta_2 = \beta_3 = 1$ และ $\beta_1 = \beta_2 = \beta_3 = -1$ และกรณี $\beta_0 = -1$ พิจารณาที่ $\beta_1 = \beta_2 = \beta_3 = 1$ และ $\beta_1 = \beta_2 = \beta_3 = -1$

3. ผลการวิจัย

การเปรียบเทียบค่าประมาณสัมประสิทธิ์การถดถอยกรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรตาม พบว่าเมื่อ $\beta_0 = 1$ และค่า $\beta_1 = \beta_2 = \beta_3 = 1$ พบว่าค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธี เมื่อขนาดตัวอย่างเพิ่มขึ้นค่า MSE มีแนวโน้มลดลง และค่า MSE มีค่าเพิ่มขึ้นเมื่อมีร้อยละค่าผิดปกติเพิ่มขึ้น สามารถสรุปผลโดยจำแนกตามค่าสัมประสิทธิ์การถดถอยได้ดังนี้ กรณีไม่มีค่าผิดปกติ พบว่าวิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุดในทุก ๆ ขนาดตัวอย่าง กรณีร้อยละค่าผิดปกติเท่ากับ 10 และ 20 ในทุก ๆ ขนาดตัวอย่าง พบว่าวิธี GM – Huber ให้ค่า MSE ต่ำที่สุด กรณีร้อยละค่าผิดปกติเท่ากับ 30 เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุด และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 วิธี GM – Huber คือวิธีที่ให้ค่า MSE ต่ำที่สุด เมื่อให้ค่า $\beta_1 = \beta_2 = \beta_3 = -1$ พบว่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธี ส่วนใหญ่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และค่า MSE ส่วนใหญ่มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละค่าผิดปกติเพิ่มขึ้น กรณีไม่มีค่าผิดปกติ พบว่าวิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุดในทุก ๆ ขนาดตัวอย่าง กรณีร้อยละค่าผิดปกติเท่ากับ 10 และ 20 ในทุก ๆ ขนาดตัวอย่าง พบว่าวิธี GM - Huber ให้ค่า MSE ต่ำที่สุด กรณีร้อยละค่า

ผิดปกติเท่ากับ 30 เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี OLS และ วิธี M - Andrews ให้ค่า MSE ต่ำที่สุด และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 วิธี GM - Huber คือวิธีที่ให้ค่า MSE ต่ำที่สุด ดังตารางที่ 1

ตารางที่ 1. ค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธีเมื่อกำหนดค่าสัมประสิทธิ์การถดถอย $\beta_0 = 1$

ร้อยละค่าผิดปกติ	n	$\beta_1 = \beta_2 = \beta_3 = 1$			$\beta_1 = \beta_2 = \beta_3 = -1$		
		OLS	M	GM	OLS	M	GM
0	20	**<u>0.066767</u>**	**<u>0.066929</u>**	0.113333	**<u>0.072195</u>**	**<u>0.072256</u>**	0.114582
	30	**<u>0.043177</u>**	**<u>0.043176</u>**	0.064870	**<u>0.042104</u>**	**<u>0.042103</u>**	0.072312
	50	**<u>0.024876</u>**	**<u>0.024876</u>**	0.036247	**<u>0.024302</u>**	**<u>0.024301</u>**	0.036785
	100	**<u>0.011401</u>**	**<u>0.011401</u>**	0.017179	**<u>0.011227</u>**	**<u>0.011227</u>**	0.016489
10	20	0.529345	0.519986	**<u>0.117577</u>**	0.542198	0.531039	**<u>0.133198</u>**
	30	0.361440	0.354574	**<u>0.068199</u>**	0.347085	0.341916	**<u>0.070309</u>**
	50	0.305465	0.305438	**<u>0.042050</u>**	0.299713	0.299687	**<u>0.036227</u>**
	100	0.238152	0.230835	**<u>0.017407</u>**	0.229070	0.219347	**<u>0.018769</u>**
20	20	1.243954	1.244938	**<u>0.957374</u>**	1.071789	1.072732	**<u>1.002805</u>**
	30	1.029680	1.029672	**<u>0.430906</u>**	1.064499	1.064495	**<u>0.547793</u>**
	50	0.802559	0.802625	**<u>0.170410</u>**	0.766295	0.766809	**<u>0.222070</u>**
	100	0.466515	0.466500	**<u>0.024561</u>**	0.479972	0.479955	**<u>0.021783</u>**
30	20	**<u>2.049031</u>**	**<u>2.049031</u>**	2.600694	**<u>1.856078</u>**	**<u>1.856083</u>**	2.429670
	30	**<u>1.431463</u>**	**<u>1.431459</u>**	1.534473	**<u>1.375594</u>**	**<u>1.375595</u>**	1.479890
	50	1.088064	1.088059	**<u>1.006115</u>**	1.040458	1.040452	**<u>0.953072</u>**
	100	0.902618	0.902613	**<u>0.742990</u>**	0.866856	0.866849	**<u>0.727119</u>**

เมื่อ $\beta_0 = 0$ และค่า $\beta_1 = \beta_2 = \beta_3 = 1$ พบว่าค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธี เมื่อขนาดตัวอย่างเพิ่มขึ้นค่า MSE มีแนวโน้มลดลง และค่า MSE มีค่าเพิ่มขึ้นเมื่อมีร้อยละค่าผิดปกติเพิ่มขึ้น สามารถสรุปผลโดยจำแนกตามค่าสัมประสิทธิ์การถดถอยได้ดังนี้ กรณีไม่มีค่าผิดปกติ พบว่าวิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุดในทุก ๆ ขนาดตัวอย่าง กรณีร้อยละค่าผิดปกติเท่ากับ 10 และ 20 ในทุก ๆ ขนาดตัวอย่าง พบว่าวิธี GM – Huber ให้ค่า MSE ต่ำที่สุด กรณีร้อยละค่าผิดปกติเท่ากับ 30 เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุด และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 วิธี GM – Huber คือวิธีที่ให้ค่า MSE ต่ำที่สุด เมื่อให้ค่า $\beta_1 = \beta_2 = \beta_3 = -1$ พบว่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธี ส่วนใหญ่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และค่า MSE ส่วนใหญ่มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละค่าผิดปกติเพิ่มขึ้น กรณีไม่มีค่าผิดปกติ พบว่าวิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุดในทุก ๆ ขนาดตัวอย่าง กรณีร้อยละค่าผิดปกติเท่ากับ 10 และ 20 ในทุก ๆ ขนาดตัวอย่าง พบว่าวิธี GM -Huber ให้ค่า MSE ต่ำที่สุด กรณีร้อยละค่าผิดปกติเท่ากับ 30 เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี OLS และ วิธี M

- Andrews ให้ค่า MSE ต่ำที่สุด และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 วิธี GM - Huber คือวิธีที่ให้ค่า MSE ต่ำที่สุด ดังตารางที่ 2

ตารางที่ 2. ค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง3 วิธีเมื่อกำหนดค่าสัมประสิทธิ์การถดถอย $\beta_0 = 0$

ร้อยละค่าผิดปกติ	n	$\beta_1 = \beta_2 = \beta_3 = 1$			$\beta_1 = \beta_2 = \beta_3 = -1$		
		OLS	M	GM	OLS	M	GM
0	20	**0.069201**	**0.069201**	0.110733	**0.067027**	**0.067027**	0.116400
	30	**0.041631**	**0.041631**	0.064580	**0.042145**	**0.042144**	0.066199
	50	**0.023631**	**0.023631**	0.035323	**0.023231**	**0.023267**	0.035279
	100	**0.011603**	**0.011603**	0.016487	**0.011488**	**0.011488**	0.018104
10	20	0.467442	0.459037	**0.129323**	0.514948	0.472033	**0.126785**
	30	0.354619	0.348299	**0.072576**	0.378037	0.376294	**0.075072**
	50	0.225315	0.221567	**0.042749**	0.217710	0.211197	**0.036859**
	100	0.301878	0.301852	**0.019513**	0.297513	0.297487	**0.018526**
20	20	1.242811	1.240178	**1.008344**	1.099016	1.098550	**1.017844**
	30	1.115554	1.115549	**0.514394**	1.109434	1.109430	**0.436105**
	50	0.762997	0.764195	**0.191437**	0.808809	0.808787	**0.207584**
	100	0.478150	0.478134	**0.023562**	0.481996	0.481981	**0.022437**
30	20	**1.998436**	**1.998430**	2.198531	**1.891598**	**1.891601**	2.366435
	30	**1.388362**	**1.388361**	1.518344	**1.427966**	**1.427963**	1.507408
	50	1.079583	1.079579	**0.970077**	1.098654	1.098650	**1.015448**
	100	0.866270	0.866263	**0.696401**	0.872232	0.872225	**0.725620**

เมื่อให้ค่า $\beta_0 = -1$ และค่า $\beta_1 = \beta_2 = \beta_3 = 1$ เป็นบวก พบว่าค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธี เมื่อขนาดตัวอย่างเพิ่มขึ้นค่า MSE มีแนวโน้มลดลง และค่า MSE มีค่าเพิ่มขึ้นเมื่อร้อยละค่าผิดปกติเพิ่มขึ้น สามารถสรุปผลโดยจำแนกตามค่าสัมประสิทธิ์การถดถอยได้ดังนี้ กรณีไม่มีค่าผิดปกติ พบว่าวิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุดในทุก ๆ ขนาดตัวอย่าง กรณีร้อยละค่าผิดปกติเท่ากับ 10 และ 20 ในทุก ๆ ขนาดตัวอย่าง พบว่าวิธี GM – Huber ให้ค่า MSE ต่ำที่สุด กรณีร้อยละค่าผิดปกติเท่ากับ 30 เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุด และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 วิธี GM – Huber คือวิธีที่ให้ค่า MSE ต่ำที่สุด เมื่อให้ค่า $\beta_1 = \beta_2 = \beta_3 = -1$ พบว่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธีส่วนใหญ่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และค่า MSE ส่วนใหญ่มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละค่าผิดปกติเพิ่มขึ้น กรณีไม่มีค่าผิดปกติ พบว่าวิธี OLS และวิธี M - Andrews ให้ค่า MSE ต่ำที่สุดในทุก ๆ ขนาดตัวอย่าง กรณีร้อยละค่าผิดปกติเท่ากับ 10 และ 20 ในทุก ๆ ขนาดตัวอย่าง พบว่าวิธี GM -Huber ให้ค่า MSE ต่ำที่สุด กรณีร้อยละค่าผิดปกติเท่ากับ 30 เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี OLS และ

วิธี M - Andrews ให้ค่า MSE ต่ำที่สุด และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 วิธี GM - Huber คือวิธีที่ให้ค่า MSE ต่ำที่สุด ดังตารางที่ 3

ตารางที่ 3. ค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธีเมื่อกำหนดค่าสัมประสิทธิ์การถดถอย $\beta_0 = -1$

ร้อยละค่าผิดปกติ	n	$\beta_1 = \beta_2 = \beta_3 = 1$			$\beta_1 = \beta_2 = \beta_3 = -1$		
		OLS	M	GM	OLS	M	GM
0	20	**0.069876**	**0.070123**	0.113891	**0.089571**	**0.089572**	0.130980
	30	**0.040957**	**0.040936**	0.065159	**0.037987**	**0.037987**	0.060508
	50	**0.022028**	**0.022028**	0.036525	**0.027964**	**0.027963**	0.039458
	100	**0.011856**	**0.011856**	0.015207	**0.009763**	**0.009763**	0.016076
10	20	0.512953	0.485771	**0.121817**	0.551542	0.526322	**0.130615**
	30	0.338372	0.341225	**0.066269**	0.394028	0.392896	**0.067563**
	50	0.314024	0.313997	**0.036790**	0.298488	0.298462	**0.040749**
	100	0.231399	0.228455	**0.018935**	0.232326	0.228786	**0.017454**
20	20	1.135874	1.139252	**0.989495**	1.257873	1.257291	**0.913181**
	30	1.079605	1.079600	**0.458736**	1.059807	1.059799	**0.409257**
	50	0.799969	0.799943	**0.172714**	0.753825	0.753803	**0.210105**
	100	0.484581	0.484565	**0.024058**	0.463364	0.463348	**0.023338**
30	20	**1.893377**	**1.896091**	2.476929	**1.903709**	**1.903706**	2.533547
	30	**1.388697**	**1.388696**	1.495888	**1.426019**	**1.426015**	1.515736
	50	1.095877	1.095871	**1.004219**	1.085146	1.085293	**0.986348**
	100	0.884633	0.884627	**0.724331**	0.865358	0.865351	**0.702633**

4. สรุปผลการวิจัย

จากการศึกษาวิจัยเชิงทดลองเพื่อการศึกษาและเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณ เมื่อข้อมูลมีค่าผิดปกติที่ตัวแปรตาม โดยผู้วิจัยได้ทำการเปรียบเทียบวิธีการประมาณค่า 3 วิธี ซึ่งได้แก่ วิธี OLS วิธี M – Andrews วิธี GM – Huber โดยใช้ค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยต่ำที่สุดเป็นเกณฑ์ในการเลือกวิธีการประมาณ พบว่า กรณีข้อมูลไม่มีค่าผิดปกติในตัวแปรตาม เมื่อขนาดตัวอย่างเท่ากับ 20, 30, 50 และ 100 ทุก ๆ ค่าสัมประสิทธิ์การถดถอยที่กำหนด ผลจากการวิจัยพบว่าวิธี OLS และวิธี M – Andrews เป็นวิธีที่เหมาะสมสำหรับใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ กรณีข้อมูลมีค่าผิดปกติที่ตัวแปรตาม สำหรับกรณีเมื่อค่าสัมประสิทธิ์การถดถอย $\beta_0 = 1$ ในทุก ๆ สถานการณ์ เมื่อพิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยพบว่าวิธี GM – Huber ประมาณค่าสัมประสิทธิ์การถดถอยได้ดีที่สุดในเกือบทุกสถานการณ์ ยกเว้น ร้อยละค่าผิดปกติเท่ากับ 30 ขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี OLS และวิธี M - Andrews เช่นเดียวกับกรณีค่าสัมประสิทธิ์การถดถอย $\beta_0 = 0$ และ $\beta_0 = -1$ ดังตารางที่ 4

ตารางที่ 4. วิธีการประมาณค่าสัมประสิทธิ์การถดถอยที่เหมาะสม

n	กรณีไม่มีค่าผิดปกติ	ร้อยละการเกิดค่าผิดปกติ		
		10	20	30
20	OLS และ M -Andrews	GM - Huber	GM - Huber	OLS และ M -Andrews
30	OLS และ M -Andrews	GM - Huber	GM - Huber	OLS และ M -Andrews
50	OLS และ M -Andrews	GM - Huber	GM - Huber	GM - Huber
100	OLS และ M -Andrews	GM - Huber	GM - Huber	GM - Huber

5. อภิปรายผลการวิจัย

จากการศึกษาครั้งนี้พบว่าปัจจัยที่มีผลต่อการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธี OLS วิธี M - Andrews และวิธี GM – Huber ประกอบด้วย ขนาดตัวอย่างมีผลต่อค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธี โดยค่า MSE ส่วนใหญ่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ร้อยละค่าผิดปกติมีผลต่อค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยของทั้ง 3 วิธี โดยค่า MSE ส่วนใหญ่มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละการเกิดค่าผิดปกติเพิ่มขึ้น จากการศึกษาการวิเคราะห์การถดถอยเชิงเส้นพหุคูณเพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยทั้ง 3 วิธี โดยกำหนดค่าสัมประสิทธิ์การถดถอยแตกต่างกัน พบว่าวิธีกำลังสองน้อยที่สุดเป็นวิธีการประมาณค่าสัมประสิทธิ์การถดถอยที่เหมาะสมที่สุด เมื่อข้อมูลไม่มีค่าผิดปกติหรือข้อมูลมีคุณสมบัติตามข้อสมมติ สอดคล้องกับการศึกษาของอรพรรณ ตันตระกูล และคณะ [5] ในกรณีที่ข้อมูลมีค่าผิดปกติเกิดขึ้นที่ตัวแปรตาม จะพบว่าวิธี GM – Huber สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้เหมาะสมที่สุดในเกือบทุกสถานการณ์ แต่ยังมีบางสถานการณ์คือ ขนาดตัวอย่างเท่ากับ 20 และ 30 วิธี M – Andrews และวิธี OLS สามารถประมาณค่าสัมประสิทธิ์การถดถอยได้เหมาะสมที่สุด สอดคล้องกับงานวิจัยของกฤตพร ธิตะจารี [7] ที่กล่าวว่า วิธีตัวประมาณ GM – Huber ให้ค่า AMSE ต่ำที่สุดเมื่อข้อมูลมีค่าผิดปกติในตัวแปรตามในกรณีขนาดตัวอย่างเท่ากับ 20

6. ข้อเสนอแนะจากการวิจัย

จากการศึกษาผู้วิจัยมีข้อเสนอแนะว่าควรมีการศึกษาเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณ เมื่อมีค่าผิดปกติในตัวแปรตาม นอกเหนือจากที่ใช้ในการศึกษาในครั้งนี้ เช่น วิธีประมาณ modified GM6 หรือวิธี M เมื่อใช้ฟังก์ชันถ่วงน้ำหนัก Welsch หรือศึกษากรณีความคลาดเคลื่อนแจกแจงแบบอื่น ๆ เช่น การแจกแจงแบบทีและการแจกแจงแบบเอกซ์โพเนนเชียล เป็นต้น

เอกสารอ้างอิง (References)

[1] วิชิต หล่อจิระชุณห์กุล และจิราวัลย์ จิตรเวช. 2548. เทคนิคการพยากรณ์. พิมพ์ครั้งที่ 3, กรุงเทพฯ : โครงการส่งเสริมและพัฒนาเอกสารวิชาการสถาบันพัฒนาบริหารศาสตร์. [ViChit Lorchirachoonkul and Jirawan Jitthavech. 2005. 3rd ed, Forecasting Techniques, Bangkok: Project for promotion and development of academic documents National Institute of Development Administration. (in Thai)]

[2] Huber, P.J. 1964. Robust Estimation of a Location Parameter**.** *Annals of Mathematical Statistics*, 35(1), 73-101.

[3] Holland, P.W. and Welsch, R.E. 1977. Robust Regression Using Iteratively Reweight Least – Squares. *Communication in Statistics – Theory and Methods*, 6(9), 813-827.

[4] Simpsona, J.R. and Montgomery, D.C. 1996. A Biased-Robust Regression Technique for the Combined Outlier-Multicolinearity Problem. *Journal of Statistics Computation and Simulation*, 56(1), 1-22.

[5] อรพรรณ ตันตระกูล, ประสิทธิ์ พยัคฆพงษ์ และบุญอ้อม โฉมที. 2555. การเปรียบเทียบวิธีการถดถอยที่มีความแกร่งสำหรับตัวแบบการถดถอยเชิงเส้นแบบพหุ. การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิตศึกษา, ครั้งที่ 13, มหาวิทยาลัยขอนแก่น, ขอนแก่น, 291-300. [Orapan Tantrakul, Prasit Payakkapong and Boonorm Chomtee. 2012. Comparison of Robust Regression Methods in Multiple Linear Regression Model. The National Graduate Research Conference, 13th, Khon Kaen University, Khon Kaen, 291-300. (in Thai)]

[6] Montgomery D.C., Peck, E.A. and Vining, G.G. 2012. Introduction to Linear Regression Analysis, 5th Edition. New York: John Wiley and Sons.

[7] กฤตพร ธิตะจารี, จุฑาภรณ์ สินสมบูรณ์ทอง และธิดาพร ศุภภากร. 2561. การเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับตัวแบบการถดถอยเชิงเส้นพหุคูณเมื่อข้อมูลมีค่านอกเกณฑ์ในตัวแปรตาม. *วารสารวิทยาศาสตร์บูรพา*, 23(2), 820-838. [Kritaporn Thitacharee, Juthaphorn Sinsomboonthong and Thidaporn Supapakorn. 2018. Efficiency Comparison of Regression Coefficient Estimation Methods for Multiple Linear Regression Model when Data Contain Outliers in Dependent Variable. *Burapha Science Journal*, 23(2) 820-838. (in Thai)]