

การประมาณค่าพารามิเตอร์สำหรับการถดถอยที่แกร่งด้วยวิธีภาวะน่าจะเป็นสูงสุด และวิธีประมาณค่าเอส

นิธิภัทร กมลสุข*

สำนักการศึกษาทั่วไป สถาบันการจัดการปัญญาภิวัฒน์

Received: 3 February 2020

Revised: 21 August 2020

Accepted: 24 August 2020

บทคัดย่อ

สำหรับการประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยในแบบเชิงเส้น ตัวประมาณค่าพารามิเตอร์วิธีกำลังสองน้อยที่สุดจะเป็นตัวประมาณค่าที่ไม่เอนเอียงเชิงเส้นดีที่สุด แต่ถ้ามีค่านอกเกณฑ์เกิดขึ้นกับค่าสังเกตหรือค่าสังเกตนั้นต่างจากค่าอื่นมากๆ จะส่งผลต่อค่าประมาณที่ได้ จึงต้องใช้วิธีการถดถอยที่แกร่งมาประมาณค่าพารามิเตอร์แทน โดยบทความวิชาการนี้จะนำเสนอขั้นตอนและวิธีการประมาณค่าพารามิเตอร์ที่แกร่ง 2 วิธี คือ วิธีภาวะน่าจะเป็นสูงสุดและวิธีการประมาณค่าเอส พร้อมทั้งแสดงผลการเปรียบเทียบประสิทธิภาพของตัวประมาณค่าพารามิเตอร์วิธีภาวะน่าจะเป็นสูงสุด เมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber และวิธีการประมาณค่าเอส โดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าสัมประสิทธิ์ของการตัดสินใจเป็นเกณฑ์การพิจารณา ซึ่งจากรายงานผลการวิจัยที่รวบรวมขึ้นพบว่า ค่าประมาณของพารามิเตอร์วิธีการประมาณค่าเอสจะมีประสิทธิภาพมากกว่าวิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber กรณีที่ค่าสังเกตจากตัวแปรอิสระมีค่าต่างจากค่าสังเกตอื่นๆ (High Leverage) เมื่อความคลาดเคลื่อนแจกแจงปกติ

คำสำคัญ: วิธีภาวะน่าจะเป็นสูงสุด วิธีการประมาณค่าเอส ค่านอกเกณฑ์

*ผู้ประสานงานหลัก; อีเมล: nithipatkam@gmail.com

Parameter Estimation for Robust Regression with Maximum Likelihood Estimation and S-estimation

Nithipat Kamolsuk*

Office of General Education, Panyapiwat Institute of Management

Received: 3 February 2020

Revised: 21 August 2020

Accepted: 24 August 2020

ABSTRACT

For parameters estimation or coefficients estimation in linear model, the Least Square (LS) estimator of parameters has always turned out to be the best linear unbiased estimator. However, if the observations contain outliers or high leverage, this may affect the Least Square estimates. So, an alternative approach; the so-called robust regression method, is needed to obtain a better fit of the model or more precise estimator of parameters. This article presents the procedure of robust regression methods, namely, the maximum likelihood and S-estimation methods and shows the comparison of the efficiency of the parameter resulted from the maximum likelihood method when using the Huber's function and that of the S-estimation method. The criterion for efficiency comparison was the mean square error (MSE) and the coefficient of determination. From the information collected from research reports, it was found that the parameters estimation with the S-estimation method was more effective than that of the maximum likelihood method when using the Huber's function, in the case that the observations from independent variables having high leverage when the error distribution being normal.

Keywords: maximum likelihood method, S-estimation method, outlier

*Corresponding Author; Email: nithipatkam@gmail.com

บทนำ

การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis) เป็นวิธีการหนึ่งที่ใช้ศึกษาความสัมพันธ์เชิงเส้น (Linear) ระหว่างตัวแปรอิสระ (Independent Variable) กับตัวแปรตาม (Dependent Variable) ที่แสดงอยู่ในรูปตัวแบบเชิงเส้น (Linear Model) ตามสมการที่ 1 และ 2

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i \quad (1)$$

$$= \sum_{j=0}^p \beta_j x_{ij} + e_i ; i = 1, 2, \dots, n \quad (2)$$

เมื่อ y_i แทนค่าสังเกตที่ i จากตัวแปรตาม y และ x_{ij} แทนค่าสังเกตที่ i จากตัวแปรอิสระ x_j โดยที่ β_j แทนพารามิเตอร์ที่ j และ e_i แทนความคลาดเคลื่อนสุ่ม (Random Error) n แทนจำนวนค่าสังเกตทั้งหมด และ p แทนจำนวนตัวแปรต้น เมื่อ $n > p$ โดยมีค่า $x_{i0} = 1$ สำหรับ $i = 1, 2, \dots, n$ และ $j = 0, 1, \dots, p$ ซึ่งค่าประมาณของ β_j หรือ $\hat{\beta}_j$ จะนำไปประมาณค่า y จึงได้ค่าประมาณของ y หรือ $\hat{y}$ เมื่อ $\hat{y} = \sum_{j=0}^p \hat{\beta}_j x_{ij}$ ที่มีส่วนเหลือ (Residual) แทนด้วย $\hat{e}_i$ เมื่อ $\hat{e}_i = y_i - \hat{y}_i$ (Ahmed and Maha, 2016)

การประมาณค่าพารามิเตอร์จากการวิเคราะห์การถดถอยเชิงเส้น โดยมากมักใช้วิธีกำลังสองน้อยที่สุด (Least Square: LS) ที่ใช้หลักการประมาณค่าพารามิเตอร์จากผลรวมของความคลาดเคลื่อนกำลังสองน้อยที่สุด หรือ $\min_{\hat{\beta}} \sum_{i=1}^n e_i^2$ ซึ่งค่าประมาณของพารามิเตอร์ β หรือ $\hat{\beta}$ มีสมบัติที่ไม่เอนเอียง (Unbiased) ถ้าค่าคาดหวัง (Expected Value) ของ $\hat{\beta}$ หรือ $E(\hat{\beta})$ เท่ากับ β และมีความแปรปรวนของ $\hat{\beta}$ หรือ $Cov(\hat{\beta})$ น้อยที่สุดเท่ากับ $\sigma^2 (X'X)^{-1}$ เมื่อ σ^2 แทนความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) มีค่าเท่ากับ $\frac{e'e}{n-p}$ เมื่อ e และ e' แทนเมทริกซ์และเมทริกซ์สลับเปลี่ยน (Transpose Matrix) ของส่วนเหลือตามลำดับ n แทนจำนวนค่าสังเกตทั้งหมด และ p แทนจำนวนตัวแปรต้น ด้วยเหตุนี้วิธีกำลังสองน้อยที่สุดจึงเป็นวิธีการประมาณค่าพารามิเตอร์ที่ทำให้ตัวประมาณค่ามีสมบัติเป็นตัวประมาณเชิงเส้นไม่เอนเอียงดีที่สุด (Best Linear Unbiased Estimator: BLUE) (Montgomery, Peck and Vining, 2006)

ทั้งนี้การวิเคราะห์การถดถอยเชิงเส้นต้องมีการตรวจสอบเงื่อนไขของความคลาดเคลื่อน e ที่ประกอบด้วย e ต้องมีการแจกแจงปกติ (Normal Distribution) ที่มีค่าเฉลี่ยเป็นศูนย์ และมีความแปรปรวนคงที่เท่ากับ σ^2 โดยที่ e_i และ e_j เมื่อ $i \neq j$ และ $i, j = 1, 2, \dots, n$ ต้องเป็นอิสระกัน ซึ่งสามารถแทนเงื่อนไขของ e ด้วย $e \sim NID(0, \sigma^2)$ เมื่อ NID แทนการแจกแจงปกติและเป็นอิสระกันของ e หากพบว่า e ไม่เป็นไปตามเงื่อนไขแล้ว จะกล่าวได้ว่าการประมาณค่าพารามิเตอร์ด้วยวิธี LS ไม่ได้เป็นวิธีการที่ดีที่สุด โดยสาเหตุหนึ่งที่ทำให้ e ไม่เป็นไปตามเงื่อนไขคือ มีค่านอกเกณฑ์ (Outlier) เกิดขึ้นกับข้อมูลหรือค่าสังเกตที่นำมาวิเคราะห์ (Adedia, Adebani, Labeodan and Adeyemi, 2016) โดยอิทธิพลของค่านอกเกณฑ์นี้นอกจากจะทำให้ความคลาดเคลื่อน e ไม่เป็นไปตามเงื่อนไขแล้ว ยังส่งผลต่อค่าประมาณของค่าพารามิเตอร์วิธีกำลังสองน้อยที่สุด (Yarmohammadi and Mahmoudvand, 2010)

ดังนั้นสิ่งสำคัญสำหรับการวิเคราะห์การถดถอยเชิงเส้น นอกจากการตรวจสอบเงื่อนไขเบื้องต้นของความคลาดเคลื่อน e แล้ว ยังต้องมีการตรวจสอบค่าสังเกตที่นำมาวิเคราะห์ด้วยว่า มีค่านอกเกณฑ์หรือมีค่าสังเกตที่แตกต่างจากค่าอื่นมาก ๆ หรือไม่ ถ้าพบว่ามีค่าสังเกตเป็นค่านอกเกณฑ์หรือมีค่าสังเกตที่แตกต่างจากค่าอื่นมากๆ แล้ว การประมาณค่าพารามิเตอร์ด้วยวิธี LS ไม่ได้เป็นวิธีการที่ดีที่สุด จึงใช้การวิเคราะห์การถดถอยที่แกร่ง (Robust

Regression) มาเป็นวิธีการประมาณค่าพารามิเตอร์แทนวิธี LS ซึ่งวิธีประมาณค่าพารามิเตอร์ เมื่อมีค่านอกเกณฑ์หรือมีค่าสังเกตที่แตกต่างจากค่าอื่นมากๆ ที่นิยมใช้มีหลายวิธี การจะเลือกใช้วิธีใดต้องพิจารณาประสิทธิภาพของค่าประมาณที่ได้ โดยบทความนี้จะแสดงให้เห็นประสิทธิภาพของค่าประมาณของพารามิเตอร์ 2 วิธี ได้แก่ วิธีภูวามะน่าจะเป็นสูงสุด เมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber และวิธีการประมาณค่าเอส จากรายงานผลการวิจัยที่รวบรวมขึ้น เพื่อใช้เป็นแนวทางในการเลือกใช้วิธีการประมาณค่าพารามิเตอร์ได้อย่างเหมาะสมที่สุด

วัตถุประสงค์

1. เพื่อแสดงขั้นตอนการประมาณค่าพารามิเตอร์จากการวิเคราะห์การถดถอยที่แกร่ง 2 วิธี คือ วิธีภูวามะน่าจะเป็นสูงสุดและวิธีการประมาณค่าเอส
2. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าพารามิเตอร์จากวิธีภูวามะน่าจะเป็นสูงสุด เมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber และวิธีการประมาณค่าเอส โดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าสัมประสิทธิ์ตัวกำหนดเป็นเกณฑ์การพิจารณา

วิธีการถดถอยที่แกร่ง (Robust Regression Methods)

การประมาณค่าพารามิเตอร์จากการวิเคราะห์การถดถอยที่แกร่งมีหลายวิธี ในที่นี้ได้สรุปขั้นตอนการประมาณค่าพารามิเตอร์ที่นิยมใช้ 2 วิธี คือ วิธีภูวามะน่าจะเป็นสูงสุดและวิธีการประมาณค่าเอส และเปรียบเทียบเพื่อให้เห็นความแตกต่างของขั้นตอนการประมาณค่าพารามิเตอร์และประสิทธิภาพของค่าประมาณที่ได้ จากรายงานผลการวิจัยที่รวบรวมขึ้น เมื่อใช้ข้อมูลจริงและข้อมูลที่สร้างขึ้น (Generate) จากการจำลองสถานการณ์ (Simulation) ภายใต้งैนไขต่างๆ

วิธีภูวามะน่าจะเป็นสูงสุด (Maximum Likelihood) หรือวิธีการประมาณค่าเอ็ม (M-estimation) นำเสนอครั้งแรกโดย Huber ในปี ค.ศ. 1964 ที่เป็นการประมาณค่าพารามิเตอร์จากผลรวมของฟังก์ชันความคลาดเคลื่อนหรือ $\rho(e_i)$ ที่มีค่าน้อยที่สุด (Pitselis, 2013) ซึ่งอยู่ในรูปแบบ ดังนี้

$$\text{Min}_{\beta} \sum_{i=1}^n \rho(e_i)$$

วิธีการประมาณค่าพารามิเตอร์ทำได้โดยการแก้สมการอนุพันธ์ย่อยอันดับที่หนึ่ง (First Order Partial Derivative) ของฟังก์ชัน $\rho(e_i)$ เทียบกับค่าพารามิเตอร์ β_j เมื่อ $j = 0, 1, \dots, p$ ซึ่งจะได้ระบบสมการจำนวน k สมการ เมื่อ $k = p + 1$ ที่แสดงได้ตามสมการที่ 3

$$\sum_{i=1}^n x_{ij} \psi \left(\frac{e_i}{s} \right) = 0 \quad (3)$$

เมื่อ ψ คืออนุพันธ์ย่อยอันดับที่หนึ่งของ $\rho(e_i)$ ทั้งนี้ Montgomery (2006) ได้สรุปตัวอย่างฟังก์ชันความคลาดเคลื่อนที่พัฒนาโดยฮูเบอร์ และมีผู้พัฒนาต่อเนื่องตามตารางที่ 1

ตารางที่ 1 ฟังก์ชันความคลาดเคลื่อนและขอบเขตความคลาดเคลื่อน

ฟังก์ชัน	$\rho(e)$	$\psi(e)$	ขอบเขต
Least squares	$\frac{1}{2} e^2$	e	$ e < \infty$
Huber's t function	$\frac{1}{2} e^2$	e	$ e \leq t$
Huber's t function at $t = 2$	$ e t - \frac{1}{2}t^2$	$t \operatorname{sign}(e)$	$ e > t$
Ramsay's E_a function at $a = 0.3$	$a^{-2}[1 - \exp(-a e) \cdot (1 + a e)]$	$e \exp(-a e)$	$ e < \infty$
Andrews's wave function	$a[1 - \cos(e/a)]$	$\sin(e/a)$	$ e \leq a\pi$
Andrews's wave function at $a =$ 1.339	$2a$	0	$ e > a\pi$
Hampel's 17A function	$\frac{1}{2} e^2$	e	$ e \leq a$
$a = 1.7$ $b = 3.4$ $c = 8.5$	$a e - \frac{1}{2}a^2$	$a \sin(e)$	$a < u \leq b$
	$\frac{a(c u - \frac{1}{2}u^2)}{c - b} - (7/6)a^2$	$\frac{a \operatorname{sign}(u)(c - u)}{c - b}$	$b < u \leq c$
	$a(b + c - a)$	0	$ u > c$

เมื่อ s คือตัวประมาณค่าที่แท้จริงของการแปรผันของส่วนเหลือ (Residual Variation) โดยที่ $\hat{e}_i$ แทนส่วนเหลือที่ i และ 0.6745 เป็นค่าคงที่ ที่ทำให้ s เป็นตัวประมาณค่าที่ไม่เอนเอียงเมื่อประมาณค่าจากตัวอย่างขนาดใหญ่คำนวณได้ดังนี้

$$s = \operatorname{median}|\hat{e}_i - \operatorname{median}(\hat{e}_i)|/0.6745 \quad (4)$$

ขั้นตอนการประมาณค่าพารามิเตอร์จะใช้วิธีกำลังสองน้อยที่สุดที่ถูกถ่วงน้ำหนักซ้ำหลายรอบ (Iteratively Reweighted Least Squares: IRLS) ซึ่งเป็นการใช้ระเบียบวิธีเชิงตัวเลข (Numerical Method) มาหาคำตอบของสมการหรือค่าพารามิเตอร์ที่ต้องการทราบค่า ที่ถูกนำไปใช้อย่างแพร่หลายซึ่งพัฒนาขึ้นครั้งแรกโดย Beaton และ Tukey ในปี ค.ศ.1974 ที่ใช้หาค่าประมาณของพารามิเตอร์จากเมทริกซ์ของค่าถ่วงน้ำหนัก $\underline{W}$ ตามรูปแบบของสมการที่ 5

$$\underline{\hat{\beta}} = (\underline{X} \underline{W} \underline{X})^{-1} \underline{X} \underline{W} \underline{Y} \quad (5)$$

เมื่อ $\underline{\beta}$ แทนเวกเตอร์ค่าประมาณของพารามิเตอร์ $\underline{\beta}$

$\underline{W}$ แทนเมทริกซ์ค่าถ่วงน้ำหนักที่เป็นเมทริกซ์ทแยงมุม (Diagonal Matrix) ขนาด $n \times n$

$\underline{X}$ แทนเมทริกซ์ของค่าสังเกตจากตัวแปรอิสระ

$\underline{X}$ แทนเมทริกซ์สลับเปลี่ยนของค่าสังเกตจากตัวแปรอิสระ

$\underline{Y}$ แทนเวกเตอร์ของค่าสังเกตจากตัวแปรตาม

การหา $\underline{\beta}$ เริ่มจากค่าพารามิเตอร์ที่ประมาณจากวิธี LS แทนด้วย $\underline{\beta}^{(0)}$ จากนั้นจึงคำนวณส่วนเหลือ $\hat{e}_i$ จากค่าพารามิเตอร์ที่ประมาณจากวิธี LS และค่าประมาณของการแปรผันของส่วนเหลือ หรือ s จากสมการที่ 4 นำค่าทั้งหมดที่ได้มาคำนวณเมทริกซ์ถ่วงน้ำหนัก $\underline{W}_i^{(0)}$ ของสมาชิกในแถวที่ i ของเมทริกซ์ $\underline{X}$ หรือ $\underline{X}_i$ ดังนี้

$$\underline{W}_i^{(0)} = \begin{cases} \frac{\psi\left[\frac{(y_i - \underline{X}_i \underline{\beta}^{(0)})}{s}\right]}{\frac{(y_i - \underline{X}_i \underline{\beta}^{(0)})}{s}} & ; y_i \neq \underline{X}_i \underline{\beta}^{(0)} \\ 1 & ; y_i = \underline{X}_i \underline{\beta}^{(0)} \end{cases} \quad (6)$$

จากสมการ 5 เมื่อนำ $\underline{W}_i^{(0)}$ ไปถ่วงน้ำหนักจะได้ค่าประมาณของพารามิเตอร์ในรอบแรก หรือ $\underline{\beta}^{(1)}$ ตามสมการที่ 7

$$\underline{\beta}^{(1)} = (\underline{X} \underline{W}_i^{(0)} \underline{X})^{-1} \underline{X} \underline{W}_i^{(0)} \underline{Y} \quad (7)$$

นำ $\underline{\beta}^{(1)}$ ที่ได้ ไปเป็นค่าเริ่มต้นของการคำนวณรอบถัดไป และจะกระทำซ้ำไปจนกระทั่งลู่เข้า (Converge) สู่คำตอบ โดย (Panik, 2009, p. 292) ได้แนะนำให้หยุดคำนวณ เมื่ออัตราส่วนของการเปลี่ยนแปลงค่าพารามิเตอร์ที่ประมาณขึ้นในแต่ละรอบ หรือ $\frac{|\underline{\beta}^{(m+1)} - \underline{\beta}^{(m)}|}{\underline{\beta}^{(m)}}$ มีค่าไม่เกิน 0.001 เมื่อ $m = 0, 1, 2, \dots$ แทนรอบของการคำนวณ

วิธีการประมาณค่าเอส (S-estimation) เป็นวิธีที่พัฒนาขึ้นจากวิธีภาวะน่าจะเป็นสูงที่สุด โดย Rousseeuw และ Yohai เป็นผู้พัฒนาครั้งแรกในปี ค.ศ. 1984 และ Rousseeuw และ Leroy เป็นผู้พัฒนาต่อเนื่องในปี ค.ศ. 1987 ที่เป็นการประมาณค่าพารามิเตอร์จากค่าประมาณของการแปรผันของส่วนเหลือ หรือ $\hat{\sigma}$ น้อยที่สุด ดังนี้

$$\min_{\underline{\beta}} \hat{\sigma}_s(\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n)$$

ทั้งนี้ $\hat{\sigma}_s = s(\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n) = s$ เมื่อ s คำนวณได้จากสมการที่ 8 ดังนี้

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{\hat{e}_i}{s}\right) = K \quad (8)$$

เมื่อ K เป็นค่าคงที่และฟังก์ชันส่วนเหลือ ρ เป็นฟังก์ชันวัตถุประสงค์ (Objective Function) ของสมการที่ 8 (Pitselis, 2013, p. 235) โดย Rousseeuw and Leroy (2003, p. 142) ได้แสดงค่าคงที่ K ที่กำหนดขึ้นจากค่าร้อยละของจุดแบ่งข้อมูล (Breakdown Point) (ϵ^*) และประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ (Efficiency of Parameter) (e^*) ตามตารางที่ 2

ตารางที่ 2 ค่าคงที่ K ค่าร้อยละของจุดแบ่งข้อมูล ($\mathcal{E}^*$) และประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ (e^*)
สำหรับวิธีการประมาณค่าเอส

K	$\mathcal{E}^*$	e^*
0.1995	50%	28.7%
0.2312	45%	37.0%
0.2634	40%	46.2%
0.2957	35%	56.0%
0.3278	30%	66.1%
0.3593	25%	75.9%
0.3899	20%	84.7%
0.4194	15%	91.7%
0.4475	10%	96.6%

สำหรับขั้นตอนการประมาณค่าพารามิเตอร์โดยวิธี IRLS เริ่มจากนำค่าประมาณของพารามิเตอร์วิธี LS มาเป็นค่าเริ่มต้น (Initial Value) แทนด้วย $\hat{\beta}^{(0)}$ จากนั้นให้คำนวณส่วนเหลือจากค่าประมาณของพารามิเตอร์วิธี LS หรือ $\hat{\epsilon}_i^{(0)}$ และค่าการแปรผันของส่วนเหลือหรือ $s^{(0)}$ จาก $s = \text{median}|\hat{\epsilon}_i - \text{median}(\hat{\epsilon}_i)| / 0.6745$ เช่นเดียวกับวิธีภาวะน่าจะเป็นสูงสุด หลังจากนั้นให้คำนวณค่าส่วนเหลือมาตรฐาน (u_0) จาก $\hat{\epsilon}_i^{(0)} / s^{(0)}$ เพื่อนำไปใช้คำนวณหาค่าน้ำหนักเริ่มต้น หรือ $W_i^{(0)}$ เมื่อ $W_i^{(0)} = \psi(u_0) / u_0$ โดย $\psi(u_i)$ แทนอนุพันธ์ย่อยอันดับที่หนึ่งของ $\rho(u_i)$ หรือ $\rho'(u_i)$ ทั้งนี้หากใช้ฟังก์ชัน $\rho(u_i)$ ของ Huber มาคำนวณฟังก์ชัน $\psi(u_i)$ จะแสดงได้ ดังนี้

$$\psi(u_i) = \rho'(u_i) = \begin{cases} 1 & ; |u_i| \leq c \\ c/|u_i| & ; |u_i| > c \end{cases} \quad (9)$$

เมื่อ c เป็นค่าคงที่ที่ใช้กับฟังก์ชัน $\rho(u_i)$ ของ Huber มีค่าเท่ากับ 1.345σ โดยที่ σ แทนค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนที่ประมาณได้จากค่า s จากนั้นนำ $W_i^{(0)}$ ที่คำนวณได้ไปถ่วงน้ำหนักกับค่าสังเกตจะได้ค่าประมาณของพารามิเตอร์ในรอบแรกแทนด้วย $\hat{\beta}^{(1)}$ ตามสมการที่ 10

$$\hat{\beta}^{(1)} = (\underline{X} W_i^{(0)} \underline{X})^{-1} \underline{X} W_i^{(0)} \underline{Y} \quad (10)$$

การคำนวณในรอบถัดไปจะคำนวณค่าการแปรผันของส่วนเหลือใหม่ จากค่าถ่วงน้ำหนักที่ได้ ตามสมการที่ 11

$$s = \sqrt{\frac{1}{nK} \sum_{i=1}^n W_i r_i^2} \quad (11)$$

จากนั้นคำนวณส่วนเหลือมาตรฐาน u_i ในรอบต่อไป เพื่อนำมาทั้งหมดที่ได้มาคำนวณค่าน้ำหนัก และเมื่อนำไปถ่วงน้ำหนักกับค่าสังเกต ทำให้ได้ค่าประมาณของพารามิเตอร์ในรอบถัดไป และจะกระทำเช่นนี้ซ้ำจนกระทั่งเข้าสู่ค่าตอบเช่นเดียวกับวิธีภาวนาจะเป็นสูงสุด

ประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ (The Efficiency of the Parameter Estimator)

เมื่อประมาณค่าพารามิเตอร์ได้แล้ว จะตรวจสอบประสิทธิภาพของตัวประมาณค่าพารามิเตอร์จากค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) โดยพบว่าหากค่าประมาณใดมี MSE น้อยกว่าแสดงว่าค่าประมาณนั้นมีประสิทธิภาพสูงกว่า ซึ่งจากข้อมูลของ Candan (1995) อ้างอิงในงานวิจัยของ Meral and Onur (2011) ที่ประกอบด้วยค่าสังเกตจากตัวแปรตาม y ที่แทนปริมาณสาร Sulphur Dioxide ซึ่งมีค่านอกเกณฑ์ปนอยู่ และตัวแปรอิสระ ได้แก่ x_1 แทนความกดอากาศ (Pressure) x_2 แทนอุณหภูมิต่ำสุดในแต่ละวัน (Minimum Daily Temperature) x_3 แทนค่าความชื้น (Humidity) x_4 แทนค่าความร้อนจากดวงอาทิตย์ (Sun) x_5 แทนปริมาณน้ำฝน (Rain) x_6 แทนความเร็วลม (Speed of Wind) x_7 แทนทิศทางลม (Direction of Wind) และ G แทนตัวแปรหุ่น (Dummy Variable) โดยแสดงความสัมพันธ์ของตัวแปรต่าง ๆ ตามสมการถดถอย ตามสมการที่ 12

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 G \quad (12)$$

เมื่อ $\hat{y}$ แทนค่าประมาณของปริมาณสาร Sulphur Dioxide และ β_j เมื่อ $j = 0,1,2, \dots, 8$ แทนค่าประมาณของพารามิเตอร์ หรือค่าสัมประสิทธิ์การถดถอย ทั้งนี้เมื่อใช้โปรแกรม S-Plus คำนวณค่าประมาณของพารามิเตอร์ โดยเปรียบเทียบกับระหว่างวิธี LS วิธีภาวนาจะเป็นสูงสุด และวิธีการประมาณค่าเอส ที่ใช้เมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber เมื่อไม่มีค่านอกเกณฑ์เกิดขึ้น และกำหนดให้มีค่านอกเกณฑ์กับตัวแปรหุ่น (Dummy Variable) G จำนวน 1 ค่าและ 2 ค่า จึงได้ค่าประมาณของพารามิเตอร์ที่แสดงได้ตามตารางที่ 3 ตารางที่ 4 และ 5 ตามลำดับ

ตารางที่ 3 ค่าประมาณของพารามิเตอร์ เมื่อไม่มีค่านอกเกณฑ์

พารามิเตอร์	วิธีการประมาณค่าพารามิเตอร์		
	LS	ภาวนาจะเป็นสูงสุด	ตัวประมาณค่าเอส
β_0	0.9862	1.7849	0.7626
β_1	0.0047	0.0039	0.0050
β_2	-0.0336	-0.0352	-0.0355
β_3	-0.0014	-0.0015	-0.0016
β_4	-0.0003	-0.0004	-0.0005
β_5	-0.0378	-0.0389	-0.0410
β_6	0.0093	0.0132	0.0190
β_7	-0.0398	-0.0420	-0.0414
β_8	0.3021	0.2855	0.0218

ตารางที่ 4 ค่าประมาณของพารามิเตอร์ เมื่อมีค่านอกเกณฑ์หนึ่งค่า

พารามิเตอร์	วิธีการประมาณค่าพารามิเตอร์		
	LS	ภาวจะน่าจะเป็นสูงสุด	ตัวประมาณค่าเอส
β_0	9.5637	4.3804	2.4112
β_1	-0.0036	0.0014	0.0034
β_2	-0.0336	-0.0350	-0.0355
β_3	-0.0018	-0.0019	-0.00206
β_4	0.0005	-0.0002	-0.00043
β_5	-0.0296	-0.0367	-0.0395
β_6	-0.0037	0.0087	0.01614
β_7	-0.0577	-0.0483	-0.0457
β_8	0.3342	0.2983	0.2834

ตารางที่ 5 ค่าประมาณของพารามิเตอร์ เมื่อมีค่านอกเกณฑ์สองค่า

พารามิเตอร์	วิธีการประมาณค่าพารามิเตอร์		
	LS	ภาวจะน่าจะเป็นสูงสุด	ตัวประมาณค่าเอส
β_0	11.2018	4.7731	2.3818
β_1	-0.0058	0.00081	0.0033
β_2	-0.0281	-0.0328	-0.0342
β_3	0.00198	0.0000106	-0.0012
β_4	0.00115	0.0000418	-0.0003
β_5	-0.0257	-0.03525	-0.03896
β_6	0.0043	0.01256	0.01821
β_7	-0.0401	-0.0401	-0.0417
β_8	0.3094	0.2809	0.2746

จากตารางที่ 4 และ 5 พบว่า ค่าประมาณของพารามิเตอร์เมื่อมีค่านอกเกณฑ์เกิดขึ้นจะแตกต่างจากค่าประมาณของพารามิเตอร์เมื่อไม่มีค่านอกเกณฑ์ตามตารางที่ 3 อย่างชัดเจน โดยความแตกต่างที่เห็นได้อย่างชัดเจนมากที่สุดคือ ค่าประมาณของค่าพารามิเตอร์จากวิธี LS แสดงว่าหากมีค่านอกเกณฑ์เกิดขึ้นกับค่าสังเกตแล้วจะส่งผลต่อค่าประมาณของพารามิเตอร์จากวิธี LS มาก ทั้งนี้เมื่อเปรียบเทียบความแตกต่างระหว่างค่าประมาณของพารามิเตอร์ด้วยวิธีภาวจะน่าจะเป็นสูงสุดกับวิธีการประมาณค่าเอสจากค่านอกเกณฑ์หนึ่งและสองค่าจะพบความแตกต่างกันเพียงเล็กน้อย

เมื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าพารามิเตอร์จากค่า MSE จากงานวิจัยของ Meral and Onur (2011, p.749-751) ตามตารางที่ 6 ด้วยค่าสังเกตหรือข้อมูลที่สร้างขึ้น ด้วยขนาดตัวอย่างเท่ากับ 15 จากตัวแปรอิสระ 5 ตัวแปร ที่มีการแจกแจงเอกรูป (Uniform Distribution) ในช่วง -1 ถึง 1 หรือ $U[-1,1]$ และค่าคลาดเคลื่อนแจกแจงปกติ ที่มีค่าเฉลี่ย เท่ากับ 0 และความแปรปรวน (Variance) เท่ากับ 100 โดยที่ตัวแปรตาม

แจกแจงบีตา (Beta Distribution) ที่มีรูปแบบและพารามิเตอร์เป็น Beta (5, 3, $\sqrt{6}$, 0, 0) เมื่อกำหนดค่านอกเกณฑ์จากตัวแปรอิสระเท่ากับ 0 1 และ 2 ค่า

ตารางที่ 6 ค่า MSE ของค่าประมาณพารามิเตอร์ภายใต้ค่านอกเกณฑ์ต่างๆ

จำนวนค่านอกเกณฑ์	วิธีการประมาณค่าพารามิเตอร์		
	LS	ภาวะน่าจะเป็นสูงสุด	ตัวประมาณค่าเอส
0	0.062710	0.067337	0.134558
1	0.09231048	0.07472932	0.1315891
2	0.1183422	0.09996891	0.02261035

จากตารางที่ 6 พบว่าค่า MSE ของค่าประมาณพารามิเตอร์ด้วยวิธี LS เมื่อไม่มีค่านอกเกณฑ์เกิดขึ้น (ค่านอกเกณฑ์เท่ากับศูนย์) มีค่าน้อยที่สุด แสดงว่าค่าประมาณของพารามิเตอร์นี้ที่มีประสิทธิภาพมากกว่าวิธีภาวะน่าจะเป็นสูงสุดและวิธีตัวประมาณค่าเอส ซึ่งบ่งชี้ได้ว่าวิธี LS เหมาะสมที่จะนำมาใช้ประมาณค่าพารามิเตอร์กรณีที่ไม่มีความนอกเกณฑ์เกิดขึ้น แต่ถ้าในกรณีที่ค่านอกเกณฑ์เกิดขึ้นกับตัวแปรอิสระค่าประมาณของพารามิเตอร์วิธี LS จะให้ค่า MSE มากกว่าสองวิธีที่เหลือแสดงว่า ค่าประมาณของพารามิเตอร์ด้วยวิธี LS จะมีประสิทธิภาพต่ำกว่าวิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber และวิธีตัวประมาณค่าเอส จึงควรใช้วิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber หรือวิธีตัวประมาณค่าเอสแทนวิธี LS

ทั้งนี้เมื่อเปรียบเทียบประสิทธิภาพของค่าประมาณพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber กับวิธีตัวประมาณค่าเอสพบว่า ถ้ามีค่านอกเกณฑ์เกิดขึ้นกับตัวแปรอิสระในระดับมากแล้วประสิทธิภาพของค่าพารามิเตอร์วิธีตัวประมาณค่าเอสจะมากกว่าวิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber เนื่องจากมีค่า MSE น้อยกว่า ซึ่งสอดคล้องกับผลวิจัยที่เปรียบเทียบค่า MSE ของตัวประมาณค่าพารามิเตอร์วิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber กับวิธีตัวประมาณค่าเอสที่ความแปรปรวนของความคลาดเคลื่อน หรือ σ^2 เท่ากับ 10 1 และ 0.01 เมื่อมีค่านอกเกณฑ์จำนวนต่างๆ ตามตารางที่ 7

ตารางที่ 7 ค่า MSE ของวิธีภาวะน่าจะเป็นสูงสุดกับวิธีตัวประมาณค่าเอสที่ σ^2 และค่านอกเกณฑ์จำนวนต่างๆ

วิธีประมาณค่า	$\sigma^2 = 10$			$\sigma^2 = 1$			$\sigma^2 = 0.01$		
	จำนวนค่านอกเกณฑ์								
	0	1	2	0	1	2	0	1	2
วิธีภาวะน่าจะเป็นสูงสุด	0.067338	0.074678	0.103733	0.067338	0.074724	0.010495	0.067338	0.076120	0.125205
วิธีตัวประมาณค่าเอส	**	**	**	**	**	**	**	**	**

หมายเหตุ ค่า ** แสดงว่า MSE มีค่าเข้าใกล้ศูนย์มาก หรือมีค่าน้อยมากที่ไม่สามารถแสดงทศนิยมได้ทั้งหมด

นอกจากนี้ยังพบว่า ประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ด้วยวิธีตัวประมาณค่าเอสมีมากกว่าวิธีภาวจะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber ในกรณีที่ค่าสังเกตจากตัวแปรอิสระมีค่าต่างจากค่าสังเกตอื่นมากๆ (High Leverage) ตามที่ปรากฏในงานวิจัยต่างๆ ดังนี้ Almetwally and Almohgy (2018) ประมาณค่าพารามิเตอร์จากค่าสังเกตหรือข้อมูลจากตัวแปรอิสระที่สร้างขึ้นให้แจกแจงเอกรูป โดยกำหนดค่าสังเกตจากตัวแปรอิสระมีค่าต่างจากค่าอื่นมากๆ ภายใต้ความคลาดเคลื่อนแจกแจงปกติ ที่ความแปรปรวนต่างๆ พบว่าจากการจำลองสถานการณ์วิธีมอนติคาร์โล (Monte Carlo Method) จำนวน 1,500 รอบ ตัวประมาณค่าพารามิเตอร์ด้วยวิธีตัวประมาณค่าเอสมีประสิทธิภาพมากกว่าวิธีภาวจะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber โดยพิจารณาจากค่า MSE ในกรณีที่มีค่านอกเกณฑ์ร้อยละ 30 ขนาดตัวอย่างเท่ากับ 50 100 และ 150 และจำนวนพารามิเตอร์เท่ากับ 3 และ 6 ตามตารางที่ 8

ตารางที่ 8 ค่า MSE ของวิธีภาวจะน่าจะเป็นสูงสุดกับวิธีตัวประมาณค่าเอส ที่ขนาดตัวอย่าง จำนวนพารามิเตอร์ และความแปรปรวนของความคลาดเคลื่อน (σ^2) ต่างๆ

วิธีประมาณ ค่าพารามิเตอร์	จำนวนพารามิเตอร์ (p)											
	$p = 3$						$p = 6$					
	$\sigma^2 = 1$			$\sigma^2 = 25$			$\sigma^2 = 1$			$\sigma^2 = 25$		
	ขนาดตัวอย่าง											
	50	100	150	50	100	150	50	100	150	50	100	150
วิธีภาวะน่าจะเป็นสูงสุด	11.43	194.61	19.08	285.63	4865.34	477.07	32.79	272.79	69.30	819.84	6819.64	4232.57
วิธีตัวประมาณค่าเอส	0.46	0.26	0.26	11.51	6.42	6.37	1.41	0.81	0.75	35.33	20.30	18.65

Yu and Yao(2017) ประมาณค่าพารามิเตอร์จากข้อมูลหรือค่าสังเกตจากตัวแปรอิสระที่สร้างขึ้นให้แจกแจงปกติมาตรฐาน (Standard Normal Distribution) โดยกำหนดค่าสังเกตจากตัวแปรอิสระมีค่าต่างจากค่าอื่นมากๆ ภายใต้ความคลาดเคลื่อนแจกแจงปกติมาตรฐานเช่นเดียวกัน พบว่าจากการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล จำนวน 200 รอบ ตัวประมาณค่าพารามิเตอร์ด้วยวิธีตัวประมาณค่าเอสมีประสิทธิภาพมากกว่าวิธีภาวจะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber โดยพิจารณาจากค่า MSE ในกรณีที่มีค่านอกเกณฑ์ร้อยละ 10 ขนาดตัวอย่างเท่ากับ 20 และ 100 และจำนวนพารามิเตอร์เท่ากับ 2 และ 4 ตามตารางที่ 9

ตารางที่ 9 ค่า MSE ของวิธีภาวจะน่าจะเป็นสูงสุดกับวิธีตัวประมาณค่าเอส ที่ขนาดตัวอย่าง และจำนวนพารามิเตอร์

วิธีประมาณค่าพารามิเตอร์	จำนวนพารามิเตอร์ (p)			
	$p = 2$		$p = 4$	
	ขนาดตัวอย่าง			
	20	100	20	100
วิธีภาวจะน่าจะเป็นสูงสุด	13.74	13.62	0.85	0.16
วิธีตัวประมาณค่าเอส	0.17	0.03	0.19	0.03

Alma (2011) ประมาณค่าพารามิเตอร์จากค่าสังเกตหรือข้อมูลจากตัวแปรอิสระที่สร้างขึ้นให้แจกแจงปกติมาตรฐาน โดยกำหนดค่าสังเกตจากตัวแปรอิสระร้อยละ 5 ที่มีค่าต่างจากค่าอื่นมากๆ เมื่อมีค่านอกเกณฑ์เกิดขึ้นกับค่าสังเกตจากตัวแปรตามร้อยละ 10 และร้อยละ 15 ภายใต้ความคลาดเคลื่อนแจกแจงปกติมาตรฐาน ขนาดตัวอย่างเท่ากับ 30 และจำนวนพารามิเตอร์เท่ากับ 3 4 และ 6 พบว่า จากการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล จำนวน 1,000 รอบ ค่าสัมประสิทธิ์ของการตัดสินใจ (Coefficient of Determination) หรือ R^2 ที่ได้จากการถดถอยที่ค่าพารามิเตอร์ถูกประมาณค่าด้วยวิธีตัวประมาณค่าเอสมีมากกว่าสมการถดถอยที่ค่าพารามิเตอร์ถูกประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด เมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber ซึ่งแสดงว่าตัวแปรอิสระจากสมการถดถอยที่ค่าพารามิเตอร์ถูกประมาณด้วยวิธีตัวประมาณค่าเอส สามารถอธิบายการเปลี่ยนแปลงของตัวแปรตามได้ดีกว่า ที่แสดงค่า R^2 ตามตารางที่ 10 เมื่อมีตัวแปรอิสระที่ต่างจากค่าอื่นมากๆ ร้อยละ 5 ร้อยละของค่านอกเกณฑ์ และจำนวนพารามิเตอร์ต่างๆ ดังนี้

ตารางที่ 10 ค่า R^2 ที่ร้อยละของค่านอกเกณฑ์ และจำนวนพารามิเตอร์ต่าง ๆ

วิธีประมาณ ค่าพารามิเตอร์	จำนวนพารามิเตอร์ (p)					
	$p = 3$		$p = 4$		$p = 6$	
	ร้อยละของค่านอกเกณฑ์					
	10	15	10	15	10	15
วิธีภาวะน่าจะเป็นสูงสุด	0.829	0.789	0.839	0.797	0.839	0.797
วิธีตัวประมาณค่าเอส	0.995	0.995	0.997	0.995	0.998	0.998

Shafiq, Amir and Zafakali (2017) ประมาณค่าพารามิเตอร์เมื่อใช้ค่าสังเกตจากตัวแปรตามและตัวแปรอิสระที่ใช้ค่าสังเกตจาก Ahmed and Shafiq (2013) โดยที่ตัวแปรตามแทนระดับไตรกลีเซอไรด์ (Triglyceride) หน่วยมิลลิกรัมต่อเดซิลิตร (mg/dl) ในผู้ป่วย มีตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตาม 6 ตัว ได้แก่ น้ำหนักตัว (Weight) หน่วยกิโลกรัม (kg) ปริมาณคอเลสเตอรอล (Cholesterol) รวมทั้งหมดในร่างกายหน่วย (mg/dl) ร้อยละโปรคอนเวอร์ติน (Proconvertin) ระดับคอเลสเตอรอล HDL (HDL- Cholesterol) รวมหน่วย (mg/dl) ขนาดสะโพก (Hip) หน่วยเซนติเมตร (cm) การมีลิพิด (Lipid) โดยที่ค่าสังเกตจากตัวแปรอิสระมีค่าแตกต่างจากค่าอื่นมากๆ จำนวน 4 ค่าจากค่าสังเกตทั้งหมด 39 ค่า คิดเป็นร้อยละ 10.2 ที่ความคลาดเคลื่อนแจกแจงปกติ พบว่า R^2 ที่ได้จากการถดถอยที่ค่าพารามิเตอร์ถูกประมาณด้วยวิธีตัวประมาณค่าเอสมีมากกว่าสมการถดถอยที่ค่าพารามิเตอร์ถูกประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด เมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber ตามตารางที่ 11

ตารางที่ 11 ค่า R^2 จากวิธีตัวประมาณค่าเอสและวิธีภาวะน่าจะเป็นสูงสุด

วิธีประมาณค่าพารามิเตอร์	R^2
วิธีภาวะน่าจะเป็นสูงสุด	0.466
วิธีตัวประมาณค่าเอส	0.523

สรุป

การประมาณค่าพารามิเตอร์ หรือสัมประสิทธิ์การถดถอยในแบบการถดถอยเชิงเส้นนั้นจะใช้วิธีกำลังสองน้อยที่สุด หรือวิธี LS เมื่อไม่มีค่านอกเกณฑ์เกิดขึ้นกับค่าสังเกต เพราะทำให้ได้ตัวประมาณค่าที่มีสมบัติไม่เอนเอียงเชิงเส้นดีที่สุด หรือ BLUE และพบว่าตัวประมาณค่าพารามิเตอร์มีประสิทธิภาพมากกว่าการใช้วิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber และวิธีตัวประมาณค่าเอส เพราะมีค่า MSE น้อยกว่า แต่เมื่อมีค่านอกเกณฑ์เกิดขึ้นต้องใช้วิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber และวิธีตัวประมาณค่าเอสแทนวิธี LS

ทั้งนี้ค่าประมาณของพารามิเตอร์วิธีการประมาณค่าเอสมีประสิทธิภาพมากกว่าวิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber กรณีที่ค่าสังเกตจากตัวแปรอิสระมีค่าต่างจากค่าสังเกตอื่นมากๆ (High Leverage) เมื่อความคลาดเคลื่อนแจกแจงปกติ โดยค่าสัมประสิทธิ์ของการตัดสินใจหรือ R^2 จากสมการถดถอยที่ค่าพารามิเตอร์ถูกประมาณด้วยวิธีตัวประมาณค่าเอสมีค่ามากกว่าสมการถดถอยที่ค่าพารามิเตอร์ถูกประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber แสดงว่าตัวแปรอิสระจากสมการถดถอยที่ค่าพารามิเตอร์ถูกประมาณด้วยวิธีตัวประมาณค่าเอส สามารถอธิบายการเปลี่ยนแปลงของตัวแปรตามได้ดีกว่าวิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber ดังนั้นในการวิเคราะห์การถดถอยในแบบเชิงเส้นสิ่งสำคัญก่อนที่จะเลือกใช้วิธีประมาณค่าพารามิเตอร์แบบใด ควรมีการตรวจสอบค่าสังเกตหรือข้อมูลที่นำมาใช้ก่อนว่ามีค่านอกเกณฑ์หรือไม่ ตลอดจนควรพิจารณาว่ามีค่าสังเกตจากตัวแปรอิสระที่ต่างจากค่าสังเกตอื่นมากๆ หรือไม่ เพื่อเลือกใช้วิธีประมาณค่าพารามิเตอร์ได้อย่างเหมาะสมที่ทำให้ได้ค่าประมาณมีประสิทธิภาพมากที่สุด

ข้อเสนอแนะ

1. การประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยในแบบเชิงเส้น จำเป็นต้องตรวจสอบว่ามีค่านอกเกณฑ์เกิดขึ้นกับค่าสังเกต หรือข้อมูลที่นำมาวิเคราะห์หรือไม่ ซึ่งหากพบว่ามีค่านอกเกณฑ์เกิดขึ้นแล้วการประมาณค่าพารามิเตอร์วิธีภาวะน่าจะเป็นสูงสุดเมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber และวิธีตัวประมาณค่าเอส จะให้ตัวประมาณค่าพารามิเตอร์ที่มีประสิทธิภาพมากกว่าวิธี LS
2. เมื่อพบว่ามีค่าสังเกตจากตัวแปรอิสระที่มีค่าต่างจากค่าอื่นมากๆ (High Leverage) ภายใต้อาการคลาดเคลื่อนแจกแจงปกติแล้ว อาจพิจารณาเลือกวิธีตัวประมาณค่าเอสมาทดแทนวิธีภาวะน่าจะเป็นสูงสุด เมื่อใช้ฟังก์ชันความคลาดเคลื่อนของ Huber เนื่องจากให้ค่าประมาณของพารามิเตอร์ที่มีประสิทธิภาพมากกว่า

เอกสารอ้างอิง

- Adedia, D., Adebajji, A., Labeodan, M., and Adeyemi, S. (2016). Ordinary Least Squares and Robust Estimators in Linear Regression: Impacts of Outliers, Error and Response Contaminations. *Mathematics & Computer Science*, 13(4), 1-11.
- Ahmad, W. M. A. W., and Shafiq, M. (2013). High Density Lipoprotein Cholesterol Predicts Triglycerides Level in Three Distinct phases of Blood Pressure. *Basic and Applied Research*, 10(1), 38-46.

- Ahmed, M. G., and Maha, E. Q. (2016). Regression Estimation in the Presence of Outliers: A Comparative Study. *Probability and Statistics*, 5(3), 65-72.
- Alma, O. G. (2011). Comparison of Robust Regression Method. *Contemp. Math. Science*, 6(9), 409-421.
- Almetwally, E. M., and Almohgy, H. M. (2018). Comparison Between M-estimation, S-estimation, and MM estimation Methods of Robust Estimation with Application and Simulation. *Mathematical Archive*, 9(11), 55-63.
- Candan, M. (1995). *Robust Estimators in Linear Regression Analysis*. Thesis of the Degree of Master of Science Program in Statistics. Ankara, Turkey: Hacettepe University.
- Meral, C., and Onur, T. (2011). The Comparing of S-estimator and M-estimator in Linear Regression. *Science*, 24(4), 747-752.
- Montgomery, D. C., Peck, E. A. and Vining, G. G. (2006). *Introductions to Linear Regression Analysis*. 4th ed. New York: John Wiley & Sons.
- Panik, M. (2009). *Regression Modeling Methods, Theory, and Computation with SAS*. New York: Taylor & Francis Group.
- Pitselis, G. (2013). A review on robust estimators applied to regression credibility. *Computational and Applied Mathematics*, 239(8), 231-249.
- Rousseeuw, P. J., and Leroy, A. M. (2003). *Robust regression and outlier detection*. New York: John Wiley & sons.
- Shafiq, M., Amir, W. M., and Zafakali, N. S. (2017). Algorithm for Comparison of Robust Regression Methods In Multiple Linear Regression By Weighting Least Square Regression (SAS). *Modern Applied Statistical Methods*, 16(2), 490-505.
- Yarmohammadi, M., and Mahmoudvand, R. (2010). The effect of outliers on robust and resistant coefficient of determination in the linear regression models. *Academic Research*, 2(3), 133-138.
- Yu, C. and Yao, W. (2017). Robust Linear Regression: A Review and Comparison. *Communication in Statistics*, 46(8), 6261-6282.