

การตรวจเอกสาร

การตรวจเอกสารแบ่งออกเป็น 2 ส่วน คือ ส่วนของวิธีการทางสถิติและส่วนของผลงานวิจัยที่เกี่ยวข้อง

วิธีการทางสถิติ

1. วิธีการถดถอยแบบค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุด (Least Absolute Deviations Regression : LAD)

วิธีการถดถอยแบบค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุดได้พัฒนาโดย Gauss and Laplace (1787) มีหลักการ คือ ต้องการค่าประมาณของพารามิเตอร์ β_0 และ β_1 ที่ทำให้ผลรวมของค่าสัมบูรณ์ของเศษตกค้าง (sum of the absolute value of the residual; $\sum |\hat{e}_i|$) หรือ $\sum |y_i - (a + bx_i)|$ มีค่าน้อยที่สุด เมื่อ a และ b เป็นตัวประมาณค่าแบบค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุด ของพารามิเตอร์ β_0 และ β_1 ตามลำดับ

การหาเส้นการถดถอยที่ดีที่สุดโดยวิธีการถดถอยแบบค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุดนี้ไม่มีสูตรสำหรับการประมาณค่าพารามิเตอร์โดยตรง แต่มีหลักการที่สำคัญ คือ หาเส้นการถดถอยที่ดีที่สุดที่ผ่านจุดเริ่มต้นของข้อมูล (x_0, y_0) ซึ่งมีขั้นตอนดังนี้

สมมติให้ มีจุดสองจุด คือ (x_0, y_0) และ (x_i, y_i) ซึ่งเป็นจุดที่ใช้ลากเส้นผ่านจุดข้อมูลอื่นๆ โดยที่ความชันของเส้นการถดถอยที่ผ่านจุดสองจุดนี้ คือ

$$b = \frac{y_i - y_0}{x_i - x_0} \quad (1)$$

เมื่อ (x_0, y_0) หมายถึง จุดเริ่มต้นของข้อมูล

และ (x_i, y_i) หมายถึง จุดที่อยู่บนเส้นถดถอยที่ประมาณขึ้น

เรียงลำดับค่าความชันที่ได้จากค่าน้อยที่สุดไปหาค่ามากที่สุด นั่นคือ

$$b_{(1)} \leq b_{(2)} \leq \dots \leq b_{(n)} \quad (2)$$

$$\frac{y_1 - y_0}{x_1 - x_0} \leq \frac{y_2 - y_0}{x_2 - x_0} \leq \dots \leq \frac{y_n - y_0}{x_n - x_0} \quad (3)$$

จากนั้นคำนวณผลรวมสะสมของค่าสัมบูรณ์ของข้อมูลแต่ละค่าลบด้วยข้อมูลเริ่มต้นในแต่
ละรอบ ($\sum |x_i - x_1|$) มาเปรียบเทียบกับค่า $\frac{\sum |x_i - x_1|}{2}$ ถ้าค่า x_i ใด ทำให้ $\sum |x_i - x_1|$ มากกว่า
 $\frac{\sum |x_i - x_1|}{2}$ ค่า x_i นั้นจะเป็นจุดเริ่มต้น (x_0) ในการดำเนินการรอบถัดไป และดำเนินการตาม
ขั้นตอนดังกล่าวจนกระทั่งได้ข้อมูลที่ซ้ำกับจุดข้อมูลก่อนหน้าจึงหยุดขั้นตอน จะได้ค่าประมาณ
พารามิเตอร์ที่ต้องการ กล่าวคือ กำหนดให้ $T = \sum |x_i - x_0|$ และหาค่า k ที่เป็นไปตามเงื่อนไข
ต่อไปนี้

$$|x_1 - x_0| + \dots + |x_{k-1} - x_0| < \frac{1}{2}T \quad (4)$$

$$|x_1 - x_0| + \dots + |x_{k-1} - x_0| + |x_k - x_0| > \frac{1}{2}T \quad (5)$$

ดังนั้น เส้นการถดถอยที่ดีที่สุดที่ผ่านจุด (x_0, y_0) คือ $\hat{Y}_i = a + bX_i$ โดยที่

$$b = \frac{y_k - y_0}{x_k - x_0} \quad (6)$$

$$a = y_0 - bx_0 \quad (7)$$

เส้นการถดถอยที่ได้จากวิธีนี้มีข้อดี คือ ถ้าค่าของข้อมูลบางค่าเกิดการเปลี่ยนแปลง จะไม่
ส่งผลกระทบต่อค่าประมาณพารามิเตอร์ที่ได้ Wimberly (1999) สรุปว่าการประมาณค่าพารามิเตอร์
โดยวิธีค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุดมีความไวต่อค่าผิดปกติจุด leverage ตัวประมาณค่าพารามิ-
เตอร์มีความน่าเชื่อถือเนื่องจากการนำขั้นตอนการหาเส้นการถดถอยข้างต้นมาทำซ้ำจนได้
ค่าประมาณพารามิเตอร์ที่มีค่าคงที่ แต่การประมาณค่าพารามิเตอร์โดยวิธีนี้มีข้อเสีย คือ เสียเวลาใน
การคำนวณ

ตัวสถิติทดสอบสำหรับทดสอบความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระของ
เส้นการถดถอย คือ

$$t = \frac{b}{\text{est.SD}(b)} \quad (8)$$

$$\text{เมื่อ } \text{est.SD}(b) = \frac{\hat{\tau}}{\sqrt{\sum (x_i - \bar{x})^2}} \quad \text{และ} \quad \hat{\tau} = \frac{\sqrt{m} [\hat{e}_{(k_2)} - \hat{e}_{(k_1)}]}{4}$$

เมื่อ $\hat{e}_{(k_1)}$ และ $\hat{e}_{(k_2)}$ หมายถึง เศษตกค้างลำดับที่ k_1 และ k_2 ตามลำดับ โดยที่ k_1 คือ จำนวนเต็มที่มีค่าเข้าใกล้ค่า $\frac{m+1}{2} - \sqrt{m}$, k_2 คือ จำนวนเต็มที่มีค่าเข้าใกล้ค่า $\frac{m+1}{2} + \sqrt{m}$ และ $m = n-2$ เมื่อ n คือ จำนวนของเศษตกค้างที่ไม่เป็นศูนย์

ในการทดสอบจะสรุปว่า ตัวแปรตามและตัวแปรอิสระไม่มีความสัมพันธ์กัน ที่ระดับนัยสำคัญ α เมื่อ $|t| > t_{\frac{\alpha}{2}, df}$ โดยที่ $t_{\frac{\alpha}{2}}$ คือ ควอนไทล์ที่ $\frac{\alpha}{2}$ ของการแจกแจงแบบ t ที่มีจำนวนองศาอิสระ (degree of freedom : df) เท่ากับ $n-2$

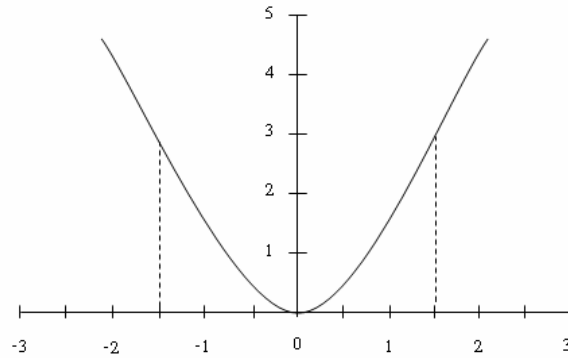
2. วิธีการถดถอยแบบเอ็ม (M Regression : M)

การถดถอยแบบเอ็มเป็นวิธีที่มีความแกร่ง (robust) ที่ได้แนวคิดจากการประมาณค่าเอ็ม (M estimation) ของ Huber (1964) ซึ่งมีชื่อว่า Huber M estimation โดยมีหลักการ คือ ต้องการค่าประมาณของพารามิเตอร์ที่ทำให้ผลรวมของฟังก์ชันของเศษตกค้าง ($\sum p(\hat{e}_i)$) มีค่าน้อยที่สุด

โดยประยุกต์ข้อดีของวิธีกำลังสองน้อยที่สุด และวิธีค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุดเข้าด้วยกัน จึงกำหนดให้ $p(e) = e^2$ เมื่อ e มีค่าเข้าใกล้ 0 และ $p(e) = |e|$ เมื่อ e มีค่าออกจาก 0 ดังนั้น

$$p(e) = \begin{cases} e^2 & ; -k \leq e \leq k \\ 2k|e| - k^2 & ; e < -k \text{ หรือ } e > k \end{cases} \quad (9)$$

เมื่อ $k = 1.5\hat{\sigma}$ โดยที่ $\hat{\sigma}$ เป็นตัวประมาณความเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนของประชากร กราฟของฟังก์ชัน $p(e)$ แสดงได้ดังภาพที่ 1



ภาพที่ 1 กราฟของฟังก์ชัน $p(e)$ เมื่อ $k = 1.5\hat{\sigma}$

ในการประมาณพารามิเตอร์ β_0 และ β_1 ด้วยการถดถอยแบบเอม จะได้ตัวประมาณค่าพารามิเตอร์ a และ b ที่ทำให้ $\sum p(y_i - (a + bx_i))$ มีค่าต่ำสุด โดยใช้การหาอนุพันธ์ของ a และ b โดยกำหนดให้เท่ากับ 0 ดังนี้

$$\sum p'(y_i - (a + bx_i)) = 0 \quad (10)$$

และ
$$\sum x_i p'(y_i - (a + bx_i)) = 0 \quad (11)$$

จากนั้นทำการปรับค่า y_i เป็นค่า $\hat{y}_i$ จะได้

$$\hat{y}_i = a + bx_i + \hat{e}_i \quad (12)$$

เมื่อ $\hat{e}_i$ คือ เศษตกค้างที่ได้รับการปรับค่าจากช่วง $[-1.5\hat{\sigma}, 1.5\hat{\sigma}]$ กล่าวคือ กรณีที่ $\hat{e}_i = -1.5\hat{\sigma}$ จะได้ค่า $e_i < -1.5\hat{\sigma}$ และกรณีที่ $\hat{e}_i = 1.5\hat{\sigma}$ จะได้ค่า $e_i > 1.5\hat{\sigma}$

จากการปรับค่าดังกล่าวนี้ ไม่มีผลต่อการเปลี่ยนแปลงค่าของ a และ b ในสมการ (10) และ (11) ดังนั้น

$$p(\hat{y}_i - (a + bx_i)) = [\hat{y}_i - (a + bx_i)]^2 \quad (13)$$

จากสมการที่ (13) ทำให้ผลรวมของฟังก์ชันของเศษตกค้าง ($\sum p(\hat{e}_i)$) หรือ $\sum [\hat{y}_i - (a + bx_i)]^2$ มีค่าต่ำสุด

Birkes and Dodge (1993) สรุปว่า วิธีการถดถอยแบบเอมมีข้อดีคือ มีคุณสมบัติความแข็งแกร่งกว่าวิธีกำลังสองน้อยที่สุด ดังนั้น เมื่อความคลาดเคลื่อนมีการแจกแจงไม่เป็นแบบปกติ สามารถนำวิธีการถดถอยแบบเอมมาใช้แทนวิธีกำลังสองน้อยที่สุดได้ แต่ข้อเสียของการประมาณค่าพารามิเตอร์ด้วยวิธีนี้ คือ มีการนำขั้นตอนดังกล่าวมาทำซ้ำจนกว่าจะได้ ค่าประมาณพารามิเตอร์ที่คงที่ ซึ่งเป็นการเสียเวลาในการคำนวณ Wimberly (1999) กล่าวว่า วิธีการถดถอยแบบเอมมีความไวต่อค่าผิดปกติและจุด leverage น้อยกว่าวิธีกำลังสองน้อยที่สุด

ตัวสถิติทดสอบสำหรับทดสอบความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระของเส้นการถดถอย คือ

$$F = \frac{STR_{\text{reduced}} - STR_{\text{full}}}{(p - q)\hat{\lambda}} \quad (14)$$

เมื่อ STR_{reduced} หมายถึง ผลรวมของความผันแปรของแยกค้ำแต่ละค่าที่ประมาณขึ้นจากค่าเฉลี่ยที่ลดรูป และ STR_{full} หมายถึง ผลรวมของความผันแปรของแยกค้ำแต่ละค่าที่ประมาณขึ้นจากค่าเฉลี่ยทั้งหมด

$$\text{ดังนั้น } STR = \sum p(\hat{e}_i) \text{ และ } \hat{\lambda} = \frac{\left(\frac{n}{m}\right) \sum \hat{e}_i^2}{n - p - 1}$$

เมื่อ p หมายถึง จำนวนของตัวแปรอิสระ (X) ทั้งหมด ในที่นี้เป็นสมการถดถอยเชิงเส้นอย่างง่ายจึงมีตัวแปรอิสระ 1 ตัว ดังนั้น $p = 1$ และ q หมายถึง จำนวนของตัวแปรอิสระ (X) ที่ไม่ได้นำมาทดสอบ ในที่นี้ไม่มีจำนวนตัวแปรอิสระเหลืออยู่ในสมการ ดังนั้น $q = 0$

ในการทดสอบจะสรุปว่า ตัวแปรตามและตัวแปรอิสระไม่มีความสัมพันธ์กัน ที่ระดับนัยสำคัญ α เมื่อ $F \geq F_{\alpha}$ หรือ $|t| \geq t_{\frac{\alpha}{2}, df}$ โดยที่ $|t| = \sqrt{F}$ เมื่อ $t_{\frac{\alpha}{2}}$ คือ ควอนไทล์ที่ $\frac{\alpha}{2}$ ของการแจกแจงแบบ t ที่มีจำนวนองศาอิสระ (degree of freedom : df) เท่ากับ $n-2$

3. วิธีการถดถอยนอนพารามетริกซ์โดยการจัดตำแหน่งของตัวเลข (Nonparametric Regression using The Ranks of Numbers : NP)

วิธีการถดถอยนอนพารามетริกซ์โดยการจัดตำแหน่งของตัวเลขเสนอโดย Härdle (1990) มีหลักการ คือ ต้องการหาค่าประมาณของพารามิเตอร์ที่ทำให้สัมประสิทธิ์ของเศษตกค้างที่ได้จากผลรวมของอันดับของค่าสัมบูรณ์ของเศษตกค้างคูณกับค่าสัมบูรณ์ของเศษตกค้าง $(\sum \text{rank}(|\hat{\epsilon}_i|)|\hat{\epsilon}_i|)$ หรือ $\sum \left[\text{rank}(\hat{\epsilon}_i) - \frac{n+1}{2} \right] \cdot \hat{\epsilon}_i$ มีค่าต่ำสุด เมื่อ n คือขนาดตัวอย่าง และ $\hat{\epsilon}_i$ คือ เศษตกค้างที่ประมาณได้ จาก $\sum \left[\text{rank}(\hat{\epsilon}_i) - \frac{n+1}{2} \right] \cdot \hat{\epsilon}_i$ ถ้าความคลาดเคลื่อนมีการแจกแจงเป็นแบบสมมาตร จะได้ว่า

$$\text{rank}(|\hat{\epsilon}_i|) = 2 \left| \text{rank}(\hat{\epsilon}_i) - \frac{n+1}{2} \right| \quad (15)$$

เนื่องจาก $\frac{n+1}{2}$ เป็นอันดับของมัธยฐานของเศษตกค้างที่เข้าใกล้ 0 และค่าสัมบูรณ์ของเศษตกค้างทางด้านลบเหมือนกับเศษตกค้างทางด้านบวก และถ้ามัธยฐานของเศษตกค้างเป็น 0 จะได้ว่า

$$\sum \left[\text{rank}(\hat{\epsilon}_i) - \frac{n+1}{2} \right] |\hat{\epsilon}_i| = \sum \left[\text{rank}(\hat{\epsilon}_i) - \frac{n+1}{2} \right] \cdot \hat{\epsilon}_i \quad (16)$$

แทนค่า $\hat{\epsilon}_i = y_i - (a + bx_i)$ ในสมการที่ (16) และเนื่องจาก a เป็นค่าคงที่ จึงไม่มีผลต่อการเปลี่ยนแปลงค่าและการเปลี่ยนแปลงอันดับของเศษตกค้าง ดังนั้น

$$\sum \left[\text{rank}(y_i - a - bx_i) - \frac{n+1}{2} \right] |y_i - a - bx_i| = \sum \left[\text{rank}(y_i - bx_i) - \frac{n+1}{2} \right] (y_i - bx_i) \quad (17)$$

$$\text{rank}(y_i - a - bx_i) = \text{rank}(y_i - bx_i) \quad (18)$$

และผลรวมของอันดับ n อันดับ คือ $\frac{n(n+1)}{2}$ จะได้

$$\sum \left[\text{rank}(y_i - bx_i) - \frac{n+1}{2} \right] \cdot a = 0 \quad (19)$$

ดังนั้น ตัวประมาณค่าพารามิเตอร์ b ที่เป็นผลรวมต่ำสุด คือ

$$\sum \left[\text{rank}(y_i - bx_i) - \frac{n+1}{2} \right] (y_i - bx_i) \quad (20)$$

การประมาณค่าพารามิเตอร์ด้วยวิธีการถดถอยนอนพาราเมตริกซ์โดยการจัดตำแหน่งของตัวเลข มีขั้นตอนดังนี้

สมมติให้จุด (x_i, y_i) และ (x_j, y_j) เป็นจุดที่ใช้ลากเส้นเชื่อมจุดข้อมูลต่างๆ และ b_{ij} เป็นการถ่วงน้ำหนักมัธยฐานของความชัน จะได้

$$b_{ij} = \frac{y_i - y_j}{x_i - x_j} \quad ; i < j \quad (21)$$

โดยความชันที่ได้มีส่วนการถ่วงน้ำหนักระหว่างจุดทั้งสอง คือ

$$w_{ij} = \frac{|x_i - x_j|}{\sum |x_i - x_j|} \quad (22)$$

และ a คือ มัธยฐานเดิมของ $y_i - bx_i$

ดังนั้น ค่าประมาณพารามิเตอร์ที่ได้สามารถเขียนในรูปแบบของสมการถดถอยเชิงเส้นอย่างง่าย คือ $\hat{Y}_i = a + bX_i$

Birkes and Dodge (1993) สรุปว่า วิธีการถดถอยนอนพาราเมตริกซ์โดยการจัดตำแหน่งของตัวเลขมีข้อดีคือ สามารถนำมาใช้ประมาณเส้นการถดถอยได้ดีในกรณีที่ความคลาดเคลื่อนมีการแจกแจงไม่เป็นแบบปกติ เนื่องจากมีการให้อันดับแทนค่าของเศษตกค้างจริง แต่วิธีนี้มีข้อเสียคือ จำนวนผลรวมของอันดับทั้งหมดที่ใช้ในการหาค่าพารามิเตอร์มีจำนวนมาก เพราะได้มาจาก

$$\frac{n(n+1)}{2}$$

ตัวสถิติทดสอบสำหรับทดสอบความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระของ
เส้นการถดถอย คือ

$$t = \frac{U}{SD(U)} \quad (23)$$

เมื่อ $U = \sum \left[\text{rank}(y_i) - \frac{n+1}{2} \right] \cdot x_i$ หรือ $U = \sum (x_i - \bar{x})(r_i - \bar{r})$ โดยที่ r คือ อันดับที่ทำกร

แทนที่ข้อมูล และ $SD(U) = \sqrt{\frac{n(n+1)}{12} \sum (x_i - \bar{x})^2}$

ในการทดสอบจะสรุปว่า ตัวแปรตามและตัวแปรอิสระไม่มีความสัมพันธ์กัน ที่ระดับ
นัยสำคัญ α เมื่อ $|t| > t_{\frac{\alpha}{2}, df}$ โดยที่ $t_{\frac{\alpha}{2}}$ คือ ควอนไทล์ที่ $\frac{\alpha}{2}$ ของการแจกแจงแบบ t ที่มี

จำนวนองศาอิสระ (degree of freedom : df) เท่ากับ $n-2$

4. การวัดประสิทธิภาพของวิธีการประมาณเส้นการถดถอย

ความผิดพลาดของการประมาณเส้นการถดถอย ขึ้นอยู่กับความคลาดเคลื่อนของการ
ประมาณซึ่งคำนวณจากผลต่างของค่าจริงกับค่าประมาณ นั่นคือ $e_i = y_i - \hat{y}_i$ เมื่อค่า y_i แทนค่า
จากข้อมูลจริง และ $\hat{y}_i$ แทนค่าที่ประมาณได้จากสมการถดถอย ความคลาดเคลื่อนของการ
ประมาณจะมีค่ามากถ้าค่าประมาณห่างจากค่าจริงมาก ดังนั้นการเปรียบเทียบค่าวัดประสิทธิภาพของ
วิธีการประมาณเส้นการถดถอย สามารถพิจารณาได้จากค่าดังต่อไปนี้

1. ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (mean square error : MSE) ซึ่งใน
การศึกษาครั้งนี้ผู้วิจัยจะทำการจำลองข้อมูลแต่ละกรณี กรณีละ 1,000 ครั้ง ดังนั้น

$$MSE = \frac{\sum_{i=1}^{1000} e_i^2}{1000} = \frac{\sum_{i=1}^{1000} (y_i - \hat{y}_i)^2}{1000}$$

2. ค่าเฉลี่ยสัมบูรณ์ของความแตกต่างของค่าประมาณเส้นการถดถอยกับเส้นการถดถอยที่
แท้จริง (mean absolute deviation : MAD) ซึ่งในการศึกษาครั้งนี้ผู้วิจัยจะทำการจำลองข้อมูลแต่ละ
กรณี กรณีละ 1,000 ครั้ง ดังนั้น

$$MAD = \frac{\sum_{i=1}^{1000} |e_i|}{1000} = \frac{\sum_{i=1}^{1000} |y_i - \hat{y}_i|}{1000}$$

เกณฑ์การตัดสิน คือ เส้นการถดถอยที่มีความผิดพลาดต่ำสุด นั่นคือ เส้นการถดถอยที่มีค่า MSE และ MAD ต่ำที่สุด จะเป็นเส้นการถดถอยที่มีประสิทธิภาพดีที่สุด

ผลงานวิจัยที่เกี่ยวข้อง

ปริยารัตน์ (2534) ศึกษาเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์จากสมการถดถอยเชิงเส้นอย่างง่าย ที่ประมาณค่าพารามิเตอร์ด้วยวิธีค่าสัมบูรณ์ต่ำสุดและวิธีกำลังสองน้อยที่สุดเกณฑ์ในการเปรียบเทียบ คือ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการพยากรณ์ กำหนดให้ความคลาดเคลื่อนมีการแจกแจงแบบปกติมาตรฐาน (standard normal distribution) แบบสม่ำเสมอ (uniform distribution) แบบลาปลาซ (laplace distribution) แบบปกติปนอมปน (contaminated normal distribution) ผลจากการศึกษาสรุปได้ดังนี้ กรณีที่ค่าคลาดเคลื่อนมีการแจกแจงแบบปกติ และแบบสม่ำเสมอ ค่าพยากรณ์ที่ได้จากการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดมีความคลาดเคลื่อนน้อยกว่าค่าพยากรณ์ที่ได้จากการประมาณค่าพารามิเตอร์ด้วยวิธีค่าสัมบูรณ์ต่ำสุด และกรณีที่ค่าคลาดเคลื่อนมีการแจกแจงแบบหางยาว ค่าพยากรณ์ที่ได้จากการประมาณค่าพารามิเตอร์ด้วยวิธีค่าสัมบูรณ์ต่ำสุดมีความคลาดเคลื่อนน้อยกว่าค่าพยากรณ์ที่ได้จากการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด

พัชนี (2536) ศึกษาประสิทธิภาพของตัวประมาณพารามิเตอร์ β_1 และ β_0 โดยวิธีการทางนอนพารามेटริกซ์เปรียบเทียบกับวิธีกำลังสองน้อยที่สุด ภายใต้สถานการณ์เมื่อข้อมูลมีการแจกแจงแบบปกติและเมื่อข้อมูลมีค่าผิดปกติ สรุปผลการศึกษาได้ดังนี้ ในการเปรียบเทียบตัวประมาณพารามิเตอร์ β_1 เมื่อข้อมูลมีการแจกแจงแบบปกติและเมื่อข้อมูลมีค่าผิดปกติเล็กน้อยในกลุ่มตัวอย่างขนาดเล็ก วิธีกำลังสองน้อยที่สุดมีประสิทธิภาพมากที่สุด และถ้าข้อมูลมีค่าผิดปกติมากขึ้น และเมื่อกลุ่มตัวอย่างขนาดใหญ่ วิธีการทางนอนพารามेटริกซ์จะมีประสิทธิภาพสูงกว่า และในกรณีการเปรียบเทียบตัวประมาณพารามิเตอร์ β_0 เมื่อข้อมูลมีการแจกแจงแบบปกติและเมื่อข้อมูลมีค่าผิดปกติเล็กน้อย ทั้งในกรณีกลุ่มตัวอย่างขนาดเล็กและขนาดใหญ่ วิธีกำลังสองน้อยที่สุดมีประสิทธิภาพสูงกว่า เมื่อข้อมูลมีค่าผิดปกติมากขึ้นและกลุ่มตัวอย่างขนาดใหญ่ขึ้น วิธีการทางนอนพารามेटริกซ์จะมีประสิทธิภาพสูงกว่าวิธีกำลังสองน้อยที่สุด

จิตรวี (2538) ศึกษาเปรียบเทียบการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ โดยวิธีกำลังสองน้อยที่สุด วิธีตัวประมาณแบบเอม และวิธีตัวประมาณ Bounded-Influence โดยใช้เกณฑ์ความแกร่งของ Huber (1964) และ Tukey (1974) ผลจากการศึกษาสรุปได้ดังนี้ ในกรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรตาม และกรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม วิธีตัวประมาณแบบเอม มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ต่ำกว่าวิธีอื่น และกรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม ณ ตำแหน่งเดียวกัน วิธีประมาณ Bounded-Influence มีค่า MSE ต่ำที่สุด และเมื่อขนาดของค่าผิดปกติของตัวแปรตามมีขนาดใหญ่ พบว่า วิธีประมาณ Bounded-Influence เมื่อใช้เกณฑ์ความแกร่งของ Tukey มีค่า MSE ต่ำที่สุด และเมื่อขนาดของค่าผิดปกติของตัวแปรตามมีขนาดเล็ก พบว่า วิธีประมาณ Bounded-Influence เมื่อใช้เกณฑ์ความแกร่งของ Huber มีค่า MSE ต่ำที่สุด

จันทร์เพ็ญ (2540) ศึกษาเปรียบเทียบการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อความคลาดเคลื่อนเกิดอัตสหสัมพันธ์ (autocorrelation) และมีค่าผิดปกติ (outlier) โดยวิธีกำลังสองน้อยที่สุด (ordinary least squares method) วิธีค่าสัมบูรณ์ต่ำสุด (least absolute value method) วิธีกำลังสองน้อยที่สุดแบบใช้การแปลงของเพรสและวินส์เทน (OLS using Prais-Winsten transformation method) วิธีการหาค่าพยากรณ์ร่วม (combine forecast method) และวิธีค่าสัมบูรณ์ต่ำสุดแบบใช้การแปลงของเพรสและวินส์เทน (LAV using Prais-Winsten transformation method) ผลจากการศึกษาสรุปได้ดังนี้ กรณีความคลาดเคลื่อนมีการแจกแจงแบบปกติ เมื่อระดับอัตสหสัมพันธ์เท่ากับ 0.1 วิธีกำลังสองน้อยที่สุดจะให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการพยากรณ์ (root mean squared forecast errors : RMSFE) ต่ำสุด และที่ระดับอัตสหสัมพันธ์เท่ากับ 0.3, 0.5, 0.7 และ 0.9 วิธีกำลังสองน้อยที่สุดแบบใช้การแปลงของเพรสและวินส์เทน จะให้ค่า RMSFE ต่ำสุด และกรณีความคลาดเคลื่อนมีการแจกแจงที่แสดงค่าผิดปกติ เมื่อระดับอัตสหสัมพันธ์เท่ากับ 0.1 วิธีค่าสัมบูรณ์ต่ำสุดจะให้ค่า RMSFE ต่ำสุด แต่เมื่อระดับอัตสหสัมพันธ์เท่ากับ 0.3, 0.5, 0.7 และ 0.9 วิธีค่าสัมบูรณ์ต่ำสุดแบบใช้การแปลงของเพรสและวินส์เทน จะให้ค่า RMSFE ต่ำสุด

สุภัทรา (2541) ศึกษาเปรียบเทียบตัวประมาณที่แกร่งสำหรับการถดถอยเชิงเส้นอย่างง่าย โดยใช้ฟังก์ชันความแกร่งของ Huber, Hampel, Andrew และ Ramsay เมื่อความคลาดเคลื่อนมีการแจกแจงไม่เป็นแบบปกติ ผลจากการศึกษาสรุปได้ดังนี้ กรณีความคลาดเคลื่อนมีการแจกแจงแบบที่ตัวประมาณแบบเอม ที่ใช้ฟังก์ชันความแกร่ง Andrew, Ramsay และ Huber จะมีประสิทธิภาพสูงสุด กรณีความคลาดเคลื่อนมีการแจกแจงแบบลอกนอร์มอล ตัวประมาณแบบเอม ที่ใช้ฟังก์ชัน

ความแกร่ง Andrew จะมีประสิทธิภาพสูงสุด และเมื่อความคลาดเคลื่อนมีการแจกแจงแบบแกมมา ตัวประมาณแบบเอ็ม ที่ใช้ฟังก์ชันความแกร่ง Hampel จะมีประสิทธิภาพสูงสุด

ประภาพร (2544) ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ 3 วิธี คือ วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณ Bounded-Influence ของ Agard และ Birch (1993) และวิธีตัวประมาณ Bounded-Influence ของ Sommer และ Huggins (1996) ผลจากการศึกษาสรุปได้ดังนี้ ประสิทธิภาพของตัวประมาณด้วยวิธีกำลังสองน้อยที่สุด และวิธีตัวประมาณ Bounded-Influence ของ Agard และ Birch มีแนวโน้มเพิ่มมากขึ้นเมื่อขนาดตัวอย่างเพิ่มมากขึ้นและมีแนวโน้มลดลงเมื่อค่าอัตราส่วนการปลอมปน และระดับความผิดปกติในความคลาดเคลื่อนเพิ่มขึ้น ปัจจัยที่มีผลต่อวิธีตัวประมาณ Bounded-Influence ของ Sommer และ Huggins ได้แก่ ขนาดตัวอย่าง อัตราส่วนการปลอมปนและระดับความผิดปกติในตัวแปรอิสระ นอกจากนี้พบว่า การแจกแจงของความคลาดเคลื่อนไม่มีผลต่อประสิทธิภาพของตัวประมาณ Bounded-Influence ของ Sommer และ Huggins

Taylor (1993) ได้ทำการปรับปรุงความเหมาะสมของเส้นโค้งด้วยวิธีค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุดโดยเปรียบเทียบกับวิธีกำลังสองน้อยที่สุด และใช้เกณฑ์ในการเปรียบเทียบ คือ สัมประสิทธิ์ของตัวกำหนด (coefficient of determination : r^2) ผลจากการศึกษาคือ วิธีกำลังสองน้อยที่สุดจะให้การถดถอยที่ดีกว่าวิธีค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุดในกรณีที่ไม่มีค่าผิดปกติ สำหรับกรณีที่ข้อมูลมีค่าผิดปกติ วิธีค่าสัมบูรณ์ของการเบี่ยงเบนต่ำสุดจะให้การถดถอยที่ดีกว่าวิธีกำลังสองน้อยที่สุด

Wimberly (1999) ได้ศึกษาเปรียบเทียบสมการถดถอยเชิงเส้นอย่างง่าย 3 วิธี คือ วิธีกำลังสองน้อยที่สุด วิธีค่าสัมบูรณ์ต่ำสุด และวิธี Huber-M ผลจากการศึกษาสรุปได้ดังนี้ วิธีกำลังสองน้อยที่สุด มีความไวต่อค่าผิดปกติและจุด leverage วิธีค่าสัมบูรณ์ต่ำสุด พบว่า มีความไวต่อค่าผิดปกติแต่น้อยกว่าวิธีกำลังสองน้อยที่สุด และมีความไวต่อจุด leverage ส่วนวิธี Huber-M ให้ผลเช่นเดียวกับวิธีกำลังสองน้อยที่สุด ถ้าความคลาดเคลื่อนมีขนาดเล็ก แต่ถ้าความคลาดเคลื่อนมีขนาดใหญ่วิธี Huber-M จะให้ผลเช่นเดียวกับกับวิธีค่าสัมบูรณ์ต่ำสุด ข้อเสียของวิธี Huber-M คือ มีความไวต่อค่าผิดปกติและจุด leverage