บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

สิ่งที่สนใจในการวิจัยครั้งนี้คือการประมาณค่าสัมประสิทธิ์การถดถอยพหุของตัวแบบถดถอยเชิงเส้นด้วยวิธี M-estimator โดยที่เกณฑ์ความแกร่งของ Huber เมื่อตัวประมาณสเกล (scale estimator) เปลี่ยนไป ในที่นี้ค่าผิดพลาดมีการแจกแจงแบบเบ้และมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ สำหรับค่าผิดพลาดที่มีการแจกแจงหางยาวกว่าการแจกแจงปกติได้ทำการศึกษาเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ที่ใช้ค่าประมาณสเกลแบบต่างๆ และสำหรับการแจกแจงแบบเบ้จะทำการศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยใช้การแปลงค่าที่อยู่ในรูปกำลังของ Box และ Cox (1964 หน้า 211-243) ในการแปลงข้อมูลให้มีการแจกแจงใกล้ภาวะปกติเสียก่อน และทำการประมาณสัมประสิทธิ์การถดถอยพหุด้วยวิธีที่กล่าวมาแล้วโดยจะกล่าวถึงรายละเอียดในหัวข้อต่อไปนี้

2.1 ข้อสมมุติทั่วไปของตัวแบบการถดถอยเชิงเส้น

2.2 วิธีการประมาณค่าสัมประสิทธิ์การถดถอยแบบต่างๆ

2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา (Ordinary least square method)

2.2.2 วิธีกำลังสองน้อยที่สุดแบบทั่วไป (General least square method)

2.2.3 วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted least square method)

2.2.4 วิธีประมาณค่าแบบ M (M-Estimate method)

2.3 ตัวประมาณค่าสเกล (Scale-Estimator)

2.4 การแปลงข้อมูลที่อยู่ในรูปยกกำลัง (Power Transformation)

2.5 ผลงานวิจัยที่เกี่ยวข้อง

2.1 ข้อสมมุติทั่วไปของตัวแบบการถดถอยเชิงเส้น

ความสัมพันธ์เชิงเส้นระหว่างตัวแปร $\underset{\sim}{y}$ กับตัวแปรอิสระ $X_1, X_2, \ldots, X_n$ และค่าผิดพลาด $\underset{\sim}{\varepsilon}$ เมื่อมีค่าสังเกต n ค่าของ $\underset{\sim}{y}$ และค่า X สามารถเขียนอยู่ในรูปสมการเชิงเส้นดังนี้

(2.1.1) $$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \ ; \ i = 1, 2, \ldots, n$$

ซึ่งสามารถจัดอยู่ในรูปเมตริกซ์ได้ดังนี้

(2.1.2) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_n, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}_{nx(p+1)}, \underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}_n, \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_n$$

ซึ่งมีข้อสมมติคือ

ก) $$E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$$

นั่นคือ ตัวแปร ε_i จะมีค่าคาดหวังเป็น 0 สำหรับทุกค่าของ $i = 1, 2, \ldots, n$

ข) $$E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2 I_n$$

จะได้ว่า $$E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \ldots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_1\varepsilon_2) & E(\varepsilon_2^2) & \ldots & E(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & & \vdots \\ E(\varepsilon_1\varepsilon_n) & E(\varepsilon_2\varepsilon_n) & \ldots & E(\varepsilon_n^2) \end{bmatrix}_{nxn}$$

$$
= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}_{n \times n}
$$

$$
= \sigma^2 I_n
$$

ค) X เป็นชุดของค่าตัวเลขที่คงที่ หรือ X เป็น fixed matrix หมายถึงการสุ่มตัวอย่างที่ซ้ำกันจะเกิดค่าหลายค่าในเวกเตอร์ $\underset{\sim}{y}$ ซึ่งแปรผันในเวกเตอร์ $\underset{\sim}{\varepsilon}$ และจำนวนค่าสังเกตของ X จะต้องมีจำนวนเกินจำนวนพารามิเตอร์ที่จะประมาณ

ง) X มี full rank กล่าวคือ rank(X) = p < n ทั้งนี้จะต้องไม่มีความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ ซึ่งเราเรียกความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระนี้ว่า multicollinearlity

2.2 การประมาณค่าสัมประสิทธิ์การถดถอยพหุ

ความสัมพันธ์ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เป็นความสัมพันธ์ที่แสดงถึงความผันแปรของค่า $\underset{\sim}{y}$ และเป้าหมายที่ต้องการคือตัวแบบเชิงเส้นที่แท้จริง (true linear model)

ให้ $\hat{\underset{\sim}{\beta}}$ คือค่าประมาณของ $\underset{\sim}{\beta}$ โดยที่ค่าประมาณนี้กระทำโดยการอาศัยข้อมูลจากค่าสังเกต

ให้ $\hat{y}_i$ เป็นค่าประมาณของ y_i ที่ได้จากสมการความสัมพันธ์ดังกล่าว และ $E(\underset{\sim}{y})$ จะสามารถใช้แทนความสัมพันธ์ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ ได้ดีก็ต่อเมื่อ ผลรวมของค่าความคลาดเคลื่อนมีค่าต่ำสุด

โดยหลักเกณฑ์ดังกล่าวเราสามารถหาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุดังต่อไปนี้

2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา (Ordinary Least Square Method (OLS))

วิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นวิธีนี้มีรากฐานมาจากทฤษฎีการ

ประมาณเชิงเส้น (Theory of Linear Estimation) โดย คาร์ล เฟดริก เกาส์ (Karl Friedrick Gause) ในปี ค.ศ.1777 - 1855 และ อังเดร แอนดรีวิช มาร์คอฟ (Andrei Andrewich Markov) ในปี ค.ศ.1856 - 1922 โดยมีหลักในการประมาณค่าสัมประสิทธิ์การถดถอยคือ ทำให้ผลบวกกำลังสองของค่าผิดพลาดมีค่าน้อยที่สุดซึ่งแสดงรายละเอียดได้ดังนี้

<u>นิยามที่ 2.1</u>

จากสมการ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$ จะได้ว่าตัวประมาณกำลังสองน้อยที่สุดของ $\underset{\sim}{\beta}$ คือ $\hat{\underset{\sim}{\beta}}$ ซึ่งทำให้ผลรวมกำลังสองของค่าผิดพลาด (sum square error หรือ SSE) มีค่าน้อยที่สุด

จากนิยามที่ 2.1 เราจะทำการหาตัวประมาณกำลังสองน้อยที่สุดได้ โดยกำหนด $\hat{\underset{\sim}{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)'$ เป็นเวกเตอร์ของตัวประมาณค่าของ $\underset{\sim}{\beta}$

(2.2.1.1) ดังนั้น $\underset{\sim}{y} = X\hat{\underset{\sim}{\beta}}$

โดยที่ $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์แนวตั้งของค่าผิดพลาด n ค่าของ $\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}$

จาก (2.2.1.1) ผลบวกกำลังสองของค่าผิดพลาดคือ

(2.2.1.2) $$\begin{aligned} SSE &= \hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}} \\ &= (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}) \\ &= (\underset{\sim}{y}' - \hat{\underset{\sim}{\beta}}'X')(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}) \\ &= \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'X\hat{\underset{\sim}{\beta}} - \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}} \end{aligned}$$

(2.2.1.3) $$= \underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}$$

จะได้ว่า $$\frac{\partial \varepsilon'\varepsilon}{\partial \hat{\underset{\sim}{\beta}}} = \frac{\partial}{\partial \hat{\underset{\sim}{\beta}}}(\underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}) = \underset{\sim}{0}$$

$$-2X'\underset{\sim}{y} + 2X'X\hat{\underset{\sim}{\beta}} = \underset{\sim}{0}$$

$$(X'X)\hat{\underset{\sim}{\beta}} = X'\underset{\sim}{y}$$

(2.2.1.4) $\therefore$ $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$. เพราะ X มี full rank

เมื่อ $(X'X)^{-1}X'$ เป็นเมตริกซ์ของค่าคงที่ และสมาชิกใน $\hat{\underset{\sim}{\beta}}$ เป็นฟังก์ชันเชิงเส้นของ $\underset{\sim}{y}$

ดังนั้น $$\begin{aligned} \hat{\underset{\sim}{\beta}} &= (X'X)^{-1}X'(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}) \\ &= (X'X)^{-1}(X'X)\underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon} \end{aligned}$$

(2.2.1.5) $$= \underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon}$$

$$E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta} + (X'X)^{-1}X'E(\underset{\sim}{\varepsilon})$$

$$= \underset{\sim}{\beta}$$

นั่นคือ $\hat{\underset{\sim}{\beta}}$ เป็นตัวประมาณที่ไม่เอนเอียงซึ่งมี

$$Var(\hat{\underset{\sim}{\beta}}) = E(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})$$

ให้ $L = (X'X)^{-1}X'$ และ $rank(X) = p$

จาก (2.2.1.4) จะได้ว่า $\hat{\underset{\sim}{\beta}} = L\underset{\sim}{y}$

ดังนั้น
$$Cov(\hat{\underset{\sim}{\beta}}) = Cov(L'\hat{\underset{\sim}{y}})$$
$$= L\ Cov(\hat{\underset{\sim}{y}})\ L'$$
$$= L\sigma^2 pL'$$
$$= \sigma^2\ LpL'$$
$$= \sigma^2\ (X'X)^{-1} \qquad (2.2.1.6)$$

เมื่อ $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2 I_n$

ให้ $\hat{\underset{\sim}{\beta}}^*$ เป็นตัวประมาณที่ไม่เอนเอียงอีกตัวหนึ่งของ $\underset{\sim}{\beta}$ โดยที่ $\hat{\underset{\sim}{\beta}}^* \neq \hat{\underset{\sim}{\beta}}$

กำหนดให้
$$\hat{\underset{\sim}{\beta}}^* = \{(X'X)^{-1}X' + C\}\underset{\sim}{y}$$
$$= (X'X)^{-1}X'\underset{\sim}{y} + C\underset{\sim}{y}$$
$$= \hat{\underset{\sim}{\beta}} + C\underset{\sim}{y}$$

เมื่อ C คือเมตริกซ์ใดๆ ซึ่งมีขนาด k x n

กำหนด $M = (X'X)^{-1}X'$

จะได้ว่า $\hat{\underset{\sim}{\beta}} = M\underset{\sim}{y}$

กล่าวคือ $\hat{\beta}_j = \sum_i^n m_{ji}y_i$; $j = 1, 2, \ldots, k$ ซึ่ง

แสดงว่า $\hat{\underset{\sim}{\beta}}$ เป็นฟังก์ชันเชิงเส้นของ $\underset{\sim}{y}$

ในทำนองเดียวกัน $\hat{\underset{\sim}{\beta}}^* = \hat{\underset{\sim}{\beta}} + C\underset{\sim}{y}$

$$\hat{\beta}_j^* = \sum_i^n (m_{ji} + c_{ij})y_i$$

กล่าวคือ $\hat{\underset{\sim}{\beta}}^*$ ก็เป็นฟังก์ชันเชิงเส้นของ $\underset{\sim}{y}$ เช่นเดียวกัน

จาก (2.2.1.4) และ $\hat{\underset{\sim}{\beta}}^* = \underset{\sim}{\beta} + C\underset{\sim}{y}$

$$\text{จะได้ว่า} \quad \hat{\underset{\sim}{\beta}}^{*} = (X'X)^{-1}X'(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}) + C(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon})$$

$$= \underset{\sim}{\beta} + (X'X)^{-1}X' + CX\underset{\sim}{\beta} + C\underset{\sim}{\varepsilon}$$

$$= \underset{\sim}{\beta} + CX\underset{\sim}{\beta} + \{(X'X)^{-1}X' + C\}\underset{\sim}{\varepsilon}$$

ดังนั้น

$$E(\hat{\underset{\sim}{\beta}}^{*}) = \underset{\sim}{\beta} + CX\underset{\sim}{\beta} + \underset{\sim}{0}$$

$$= \underset{\sim}{\beta} + CX\underset{\sim}{\beta}$$

$$\neq \underset{\sim}{\beta}$$

กล่าวคือ $\hat{\underset{\sim}{\beta}}^{*}$ เป็นตัวประมาณที่ไม่เอนเอียงของ $\underset{\sim}{\beta}$ ได้ก็ต่อเมื่อ $CX = \underset{\sim}{0}$ และความแปรปรวนหาได้โดย

$$Var(\hat{\underset{\sim}{\beta}}^{*}) = \sigma^2\{(X'X)^{-1} + CC'\}$$

จาก (2.2.1.6) จะได้ว่า

$$Var(\hat{\underset{\sim}{\beta}}^{*}) - Var(\hat{\underset{\sim}{\beta}}) = CC'$$

ซึ่งเป็นเมตริกซ์กึ่งบวกแน่นอน (positive semidefinite matrix) แสดงว่า $\hat{\underset{\sim}{\beta}}$ มีความแปรปรวนต่ำที่สุด (minimum variance) กล่าวคือ $\hat{\underset{\sim}{\beta}}$ เป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (best linear unbiased estimator (BLUE)) ของ $\underset{\sim}{\beta}$

จาก (2.2.1.2) เราสามารถเขียนผลบวกกำลังสองของความผิดพลาดได้ใหม่ดังนี้

$$\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{y}'\underset{\sim}{y} - \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y}$$

2.2.2 <u>วิธีกำลังสองน้อยที่สุดแบบทั่วไป</u> (General Least Square Method (GLS))

ในทางปฏิบัติจะได้ว่า $E(\varepsilon_i \varepsilon_j) \neq 0$ ทุกค่าของ $i \neq j$ และ $E(\varepsilon_i \varepsilon_j) \neq \sigma^2$ สำหรับทุกค่าของ $i = j$ เช่นในตัวแบบ autoregressive หรือในกรณีเกิดปัญหาเรื่อง

heterocedasticity ในสถานการณ์เช่นนี้ $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2\Omega$ โดยที่ Ω เป็น symmetric positive definite matrix ขนาด n x n และ $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2\Omega)$

เนื่องจาก Ω เป็น symmetric positive definite matrix เราจึงหาเมตริกซ์ P ซึ่งทำให้ $P'P = \Omega$ ได้

จะได้ว่า $P^{-1}\Omega(P'^{-1}) = I_n$

(2.2.2.1) $$\Omega^{-1} = (P')^{-1}P^{-1}$$

ดังนั้นเราสามารถปรับสมการ (2.2.2.1) ให้อยู่ในรูปดังนี้

$$P^{-1}\underset{\sim}{y} = P^{-1}X\underset{\sim}{\beta} + P^{-1}$$

$$\underset{\sim}{y}^* = X^*\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}^*$$

เมื่อ $\underset{\sim}{y}^* = P^{-1}\underset{\sim}{y}$

$X^* = P^{-1}X$

และ $\underset{\sim}{\varepsilon}^* = P^{-1}\underset{\sim}{\varepsilon}$

จะได้ว่า
$$\begin{aligned} E(\underset{\sim}{\varepsilon}^*\underset{\sim}{\varepsilon}^{*\prime}) &= E\{P^{-1}\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}'(P')^{-1}\} \\ &= P^{-1}E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}')(P')^{-1} \\ &= \sigma^2 P^{-1}\Omega(P')^{-1} \\ &= \sigma^2 I_n \end{aligned}$$

จาก $E(\underset{\sim}{\varepsilon}^*) = \underset{\sim}{0}$ และ $E(\underset{\sim}{\varepsilon}^*\underset{\sim}{\varepsilon}^{*\prime}) = \sigma^2 I_n$ ซึ่งสามารถประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ได้โดยวิธี OLS เช่นเดียวกันกับ (2.2.1.4) และสามารถเขียนอยู่ในรูปของ

$$\begin{aligned} \hat{\underset{\sim}{\beta}}_{GLS} &= (X^{*\prime}X^*)^{-1}X^{*\prime}\underset{\sim}{y}^* \\ &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\underset{\sim}{y}) \end{aligned}$$

และความแปรปรวนของตัวประมาณจะมีค่าเป็น

$$\begin{aligned} Var(\underset{\sim}{\beta}_{GLS}) &= \sigma^2(X^{*\prime}X^*)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1} \end{aligned}$$

ในขณะที่ตัวประมาณกำลังสองน้อยที่สุดแบบธรรมดา (ordinary least square estimator

(OLS)) คือ

$$\hat{\underset{\sim}{\beta}}_{OLS} = (X'X)^{-1}X'\underset{\sim}{y}$$
$$= \underset{\sim}{\beta} + (X'X)^{-1}X'\underset{\sim}{\varepsilon}$$

ซึ่งเป็นตัวประมาณที่ไม่เอนเอียง เมื่อ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และความแปรปรวนหาได้โดย

$$Var(\hat{\underset{\sim}{\beta}}_{OLS}) = \sigma^2(X'X)^{-1}X'E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}')X(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

ดังนั้นตัวประมาณ GLS จะมีประสิทธิภาพมากกว่าตัวประมาณ OLS ในความหมายที่ว่า $Var(\hat{\underset{\sim}{\beta}}_{OLS})$ - $Var(\hat{\underset{\sim}{\beta}}_{GLS})$ เป็นเมตริกซ์กึ่งบวกแน่นอน (positive semi-definite matrix) ในกรณีที่ $\Omega \neq I_n$

2.2.3 <u>วิธีกำลังสองแบบถ่วงน้ำหนัก</u> (Weighted Least Square Method (WLS))

เป็นกรณีหนึ่งของ GLS เมื่อค่าผิดพลาด $\underset{\sim}{\varepsilon}$ ไม่มีความสัมพันธ์กันโดยที่ความแปรปรวนไม่เท่ากัน และเมตริกซ์ของความแปรปรวนร่วมของ $\underset{\sim}{\varepsilon}$ เขียนได้โดย

$$E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2\Omega$$
$$= \sigma^2 \begin{bmatrix} 1/w_1 & 0 & \cdots & 0 \\ 0 & 1/w_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/w_n \end{bmatrix}_{n \times n}$$

เมื่อ $W = \Omega^{-1}$ และ Ω เป็นเมตริกซ์ทแยงมุม (diagonal matrix) จะได้ว่า W เป็นเมตริกซ์ทแยงมุมด้วย โดยมีเส้นทแยงมุมเป็น $w_1, w_2, w_3, \ldots, w_n$ และสมการของ weighted least square หรือ WLS อยู่ในรูปของ

$$(X'WX)\hat{\underset{\sim}{\beta}} = X'W\underset{\sim}{y}$$
$$\hat{\underset{\sim}{\beta}}_{WLS} = (X'WX)^{-1}(X'W\underset{\sim}{y})$$

ซึ่งเราเรียก w_i ว่าเป็น weight โดยที่ค่าสังเกตที่มี w_i น้อยจะมีความแปรปรวนมากกว่าค่าสังเกตที่มี w_i มาก

2.2.4 วิธีของตัวประมาณชนิด M (M-Estimator Method)

ปี ค.ศ.1964 P.J. Huber ได้ศึกษาฟังก์ชันความผิดพลาดซึ่งเรียกว่าตัวประมาณชนิด M (M-estimator) ตัวประมาณชนิด M จะประมาณพารามิเตอร์ของค่าน้อยที่สุดของ $\sum_{i}^{n} \rho(\varepsilon_i/S)$ เมื่อ ε_i เป็นค่าผิดพลาดของค่าสังเกตตัวที่ i และ S เป็นตัวประมาณของการกระจายของตัวอย่างของ ε_i

ในตัวแบบการถดถอยเชิงเส้นเราสามารถหาตัวประมาณชนิด M จาก

$$(2.2.4.1) \qquad \min_{\underset{\sim}{\beta}} \sum_{i}^{n} (\varepsilon_i/S) = \min_{\underset{\sim}{\beta}} \sum_{i}^{n} \rho\{(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/S\}$$

เมื่อ S เป็นตัวประมาณที่แกร่งของสเกล

ในปี ค.ศ. 1969 Huber ได้เสนอแนวความคิดที่เรียกว่า Huber's Proposal 2 โดยที่ได้ทำการศึกษาค่าประมาณตำแหน่งและค่าประมาณสเกลไปพร้อมๆ กัน (simultaneaus estimation of location and scale) ซึ่งอยู่ในรูปของสมการภาวะน่าจะเป็น (likelihood equation) ดังนี้

$$(2.2.4.2) \qquad \Sigma x_{ij}\, \psi\{(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/S\} = 0$$

$$(2.2.4.3) \qquad \Sigma \chi\{(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/S\} = 0$$

เมื่อกำหนดค่า $\chi = \varepsilon_i \psi - 1$

ในปี ค.ศ. 1971 P.J. Bickel ได้ศึกษาต่อจากผลงานของ Huber's Proposal 2 และทำการเสนอวิธีที่เรียกว่า One Step Huber โดยการกำหนดค่าเริ่มต้นได้แก่ ตัวประมาณตำแหน่งลำดับที่ 1 ซึ่งอาจใช้ค่าเฉลี่ย หรือ ค่ามัธยฐาน (mean or median)

และตัวประมาณที่แกร่งของสเกล โดยสามารถหาค่าประมาณพารามิเตอร์ตำแหน่งลำดับถัดไปได้ โดยที่สามารถเขียนอยู่ในรูปของ

$$(2.2.4.4) \qquad T^{(m+1)} = T^{(m)} + \frac{(1/n)S^{(0)}\Sigma\psi\{(y_i - T^{(m)})/S^{(0)}\}}{(1/n)\ \Sigma\psi'\{(y_i - T^{(m)})/S^{(0)}\}}$$

เมื่อกำหนด $T^{(0)} = \text{med}(y_i)$ หรือ $E(y_i)$

และ $S^{(0)}$ = ตัวประมาณที่แกร่งของสเกล

ดังนั้น $T^{(m)}$ คือ ค่าประมาณตำแหน่งลำดับที่ m + 1

ต่อมาในปี ค.ศ. 1977 Huber ได้เสนอรูปแบบของค่าน้อยที่สุดของฟังก์ชันความผิดพลาดสำหรับ β และ สเกล ซึ่งอยู่ในรูปของ

$$(2.2.4.5) \qquad Q(\beta, \sigma) = (1/n)\ \Sigma\rho\ \{(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/\sigma\}\sigma$$

เมื่อ ρ คือฟังก์ชันที่เหมาะสม โดยที่ $\rho > 0$, $\rho(0) = 0$

เมื่อเราหาอนุพันธ์บางส่วนเมื่อเทียบกับ $\underset{\sim}{\beta}$ และ σ จากสมการที่ (2.2.4.2) และ (2.2.4.3) จะสามารถหาค่าประมาณของ $\underset{\sim}{\beta}$ และสเกล (scale) โดยการ iteration สมการ (2.2.4.2) และ (2.2.4.3) พร้อมๆ กัน เรียกว่า simultaneous estimation of location and scale ซึ่งในทางปฏิบัติจะกระทำได้ยาก

โดยทั่วไปค่า S ที่ประมาณขึ้นมาจะถูกเลือกไม่ให้มีอิทธิพลกับความคลาดเคลื่อนที่มีขนาดใหญ่ ดังนั้นฟังก์ชันความผิดพลาดต่างๆ จะมีค่าความคลาดเคลื่อนน้อยที่สุดเมื่อกำหนดค่า S คงที่

และวิธีการประมาณค่าสัมประสิทธิ์การถดถอยส่วนใหญ่จะใช้วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักซ้ำหลายๆ รอบ (iteratively reweighted least square) ของ Beaton และ Tukey (1964) โดยจะต้องประมาณ $\underset{\sim}{\beta}$ และ S เราสามารถเขียนสมการ m สมการของ (2.2.4.5) ได้ดังนี้

$$(2.2.4.6)\quad \sum_{i}^{n} x_{ij}\psi\,[(y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})/s] \;=\; \frac{\sum_{i}^{n} x_{ij}\psi[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s]\;[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/S]}{[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/S]} \;=\; \underset{\sim}{0}$$

$$\sum x_{ij} W_{io}(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta}) \;=\; 0 \qquad ;\; j = 1, 2, \ldots, m$$

$$W_{io} = \begin{cases} \dfrac{\psi\,[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/S]}{(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/S} & ;\; y_i \neq \underset{\sim}{X}_i'\hat{\underset{\sim}{\beta}}_0 \\ 1 & ;\; y_i = \underset{\sim}{X}_i'\hat{\underset{\sim}{\beta}}_0 \end{cases}$$

สมการเหล่านี้อาจเขียนในรูปเมตริกซ์ได้ดังนี้

$$(2.2.4.7)\qquad X'W_0X\,\hat{\underset{\sim}{\beta}} = X'W_0\underset{\sim}{y}$$

เมื่อ W เป็นเมตริกซ์ทแยงมุมขนาด n x n ของน้ำหนัก (weight) และมีสมการตามเส้นทแยงมุมเป็น $w_{10}, w_{20}, \ldots, w_{n0}$

ดังนั้นตัวประมาณ $\hat{\underset{\sim}{\beta}}$ ที่หาได้ในครั้งแรกจะมีค่าเป็น

$$(2.2.4.8)\qquad \hat{\underset{\sim}{\beta}} = (X'W_0X)^{-1}X'W_0\underset{\sim}{y}$$

และจะทำครั้งถัดไปโดยคำนวณน้ำหนักใหม่และใช้ $\hat{\underset{\sim}{\beta}}_1$ แทน $\hat{\underset{\sim}{\beta}}_o$ จะกระทำซ้ำๆ กันจนกระทั่งได้ค่า $\hat{\underset{\sim}{\beta}}$ ที่ค่อนข้างคงที่ Holland และ Welsch (1977) กล่าวว่ากระบวนการกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักทำซ้ำๆ กันนี้ต้องใช้โปรแกรมคอมพิวเตอร์ของวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักวิธีมาตรฐานเท่านั้น

2.3 <u>ตัวประมาณสเกล</u> (Scale Estimator)

ในหลายๆลักษณะค่าสเกลจะถูกพิจารณาให้เป็นตัวแปรรบกวน (nuisance parameter)

ความแกร่งของระบบ ตัวประมาณความแกร่งทางตำแหน่ง (robust location estimator) หลายตัวต้องการค่าสเกล เพื่อที่จะทำให้ค่าประมาณตำแหน่งมีประสิทธิภาพมากขึ้น ตัวประมาณสเกลจึงถูกนำมาพิจารณา เพื่อหาค่าที่เหมาะสมสำหรับตัวประมาณตำแหน่งที่แกร่ง ในที่นี้ได้เสนอตัวประมาณสเกลเพื่อทำการศึกษา 3 วิธี กล่าวคือ

2.3.1) Standard Deviation of Location (SDL)

โดยกำหนดให้ y เป็นเวกเตอร์ค่าสังเกตของตัวแปรตามขนาด n x 1 แล้วจะได้ $\bar{y}$ เป็นค่าเฉลี่ยเลขคณิต ดังนั้นค่า standard deviation ของตำแหน่งคือ

(2.3.1.1) $$S = \{ Var(\bar{y}) \}^{1/2}$$

หรือ $$= \{ \sum_{i}^{n} (y_i - \bar{y})^2 / (n-1) \}^{1/2}$$

2.3.2) The Median Absolute Deviation (MAD)

เป็นค่าสเกลพื้นฐานและจะถูกปรับด้วยค่าคงที่ 0.6745 ซึ่งจะทำให้ S เป็นตัวประมาณที่ไม่เอนเอียงโดยประมาณของ σ เมื่อ n มีขนาดใหญ่และการแจกแจงของค่าผิดพลาดเป็นแบบปกติ ดังนั้น S จะอยู่ในรูปของ

(2.3.2.1) $$S = med \mid \varepsilon_i - med(\varepsilon_i) \mid / 0.6745$$

2.3.3) The Modified Biweight A-Estimator (MBA)

ตัวประมาณชนิด A ของสเกล (A-estimator of scale) ถูกกำหนดจากค่าความแปรปรวนของตัวประมาณความแกร่งของตำแหน่งเมื่อ n มีค่ามาก (asymptotic variance of M-estimator of location) โดยที่ตัวประมาณความแกร่งชนิด M ของตำแหน่งจะให้ค่าประมาณสเกล S และค่าคงที่ c ซึ่งมากกว่าศูนย์ นั่นคือ

(2.3.3.1) $$n\{Var(T_n)\}^{1/2} \rightarrow A \quad (T,F)^{1/2}$$

เมื่อ T_n คือตัวประมาณค่าแหน่ง

และ A (T,F) คือสมการ asymptotic variance ของ M-estimator ที่ขึ้นกับฟังก์ชัน ψ และ ค่าสังเกตที่ได้จากการแจกแจงแบบ F

ให้ $u_i = \{(y_i - T)/cS_0\}$

(2.3.3.2) จะได้ว่า $$S^2_{\psi,c} = n/(n-1)\ n(cS_0)^2\{\sum_i^n \psi^2(u_i)/[\sum_i^n \psi(u_i)]^2\}$$

เมื่อ S คือ A-estimator ของสเกล

T คือ ค่ามัธยฐานของค่าสังเกต

S_0 คือ ค่า MAD ของค่าสังเกต

และถ้า $\psi(u) = u$

จะได้ว่า A-estimator ของสเกลจะถูกลดรูปลงเป็น sample standard deviation

และถ้าให้ $\psi(u) = uw_{bi}(u)$

เมื่อ $w_{bi}(u)$ คือฟังก์ชันถ่วงน้ำหนักแบบ biweight โดยที่

(2.3.3.3) $$w_{bi}(u) = \begin{cases} (1 - u^2) & ; |u| < 1 \\ 0 & ; \text{อื่นๆ} \end{cases}$$

ดังนั้นจาก (2.3.3.2) และ (2.3.3.3) จะได้สมการที่ปรับแล้วดังนี้

(2.3.3.4) $$S_{bi,c} = \frac{n}{(n-1)^{1/2}} \; \frac{[\sum_{|u_i|<1} (y_i - T)^2 (1 - u_i^2)^4]^{1/2}}{|\sum_{|u_i|<1} (1 - u_i^2)(1 - 5u_i^2)|}$$

จาก (2.3.3.4) ถ้าให้ $\ _{bi}(u) = w_{bi}(u)$ จะได้สมการดังนี้

$$(2.3.3.5) \quad S_{bi,c} = \frac{ncS_0}{(n-1)^{1/2}} \; \frac{\left[\sum_{|u_i|<1} (u_i w_{bi}(u_i))^2\right]^{1/2}}{\left|\sum_{|u_i|<1} w_{bi}(u_i) + w'_{bi}(u_i)u_i\right|}$$

และถ้าให้ $w'_{bi}(u_i)u_i$ มีค่าใกล้ 0 ดังนั้นจะได้สมการดังนี้

$$(2.3.3.6) \quad S_{mb,c} = \frac{n}{(n-1)^{1/2}} \; \frac{\left[\sum_{|u_i|<1} ((y_i - T)^2 w_{bi}(u_i))^2\right]^{1/2}}{\left|\sum_{|u_i|<1} w_{bi}(u_i)\right|}$$

ซึ่งเราเรียกตัวประมาณสเกล $S_{mb,c}$ นี้ว่า ตัวประมาณชนิด A แบบปรับปรุงการถ่วงน้ำหนักซ้ำ 2 ครั้ง (the modified biweight A-estimator)

2.4 <u>การแปลงข้อมูลที่อยู่ในรูปยกกำลัง</u> (Power Transformation)

การแปลงข้อมูลที่อยู่ในรูปยกกำลัง (Power Transformation) ได้รวมการแปลงซึ่งใช้ลอกการิทึมในวงศ์ (family) ของการแปลง การแปลงข้อมูลอาจทำให้ข้อสมมติของความเป็นไปได้สำหรับแปลงข้อมูลที่ใช้ศึกษาการแจกแจงของความผิดพลาดเป็นแบบเบ้

Tukey (ค.ศ.1957 หน้า 602-632) เป็นท่านแรกที่ศึกษารายละเอียดเกี่ยวกับการแปลงข้อมูลได้แนะนำว่า การแปลงข้อมูลจะช่วยให้ตัวแบบมีการแจกแจงเข้าใกล้เชิงเส้นโดยที่ข้อมูลจะมีการกระจายเท่ากันและเป็นปกติมากขึ้น

Box และ Cox (ค.ศ.1962 หน้า 211-243) ได้พิจารณาการแปลงในตัวแปรตามและสามารถทำได้ในตัวแปรอิสระด้วย ซึ่งสามารถเขียนอยู่ในรูปของ

$$y_i^{\lambda} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & ; \lambda = 0 \\ \ln y_i & ; \lambda = 0 \end{cases}$$

จึงเป็นข้อดีของการแปลง และเมื่อเปรียบเทียบกับการแปลงอย่างง่าย y_i^{λ} ซึ่งต่อเนื่องเมื่อ $\lambda = 0$ ดังนั้น

$$L_t \underset{\lambda \to 0}{} \frac{y^{\lambda}-1}{\lambda} = \ln y_i$$

Box และ Cox จึงพิจารณาตัวแบบของการถดถอยในรูปสมการเมตริกซ์ดังรูป

$$\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{\varepsilon} \sim I_n(\underset{\sim}{0}, \sigma^2)$ ดังนั้นฟังก์ชันความหนาแน่นร่วมของค่าสังเกต $y_1, y_2, \ldots, y_n$ คือ

$$\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i}^{n} (y_i^{\lambda} - \underset{\sim}{X}_i' \underset{\sim}{\beta})^2\right] \cdot J$$

โดยที่ J เป็นจาร์โคเบียนของการแปลงจากตัวแปร y_i เป็น $y_i^{(\lambda)}$ ดังรูป

$$J = \prod_{i}^{n} \left| \frac{dy_i^{\lambda}}{dy_i} \right|$$

$$= \prod_{i}^{n} y_i^{\lambda-1}$$

จะได้ว่า ฟังก์ชัน log-likelihood อยู่ในรูปของ

$$\ln L = -\frac{n}{2} \ln 2\pi - n \ln\sigma^2 - \frac{1}{2\sigma^2} \sum_{i}^{n} [\, y_i^{\lambda} - \underset{\sim}{X}_i' \underset{\sim}{\beta}\,]^2 + (\lambda - 1) \sum_{i}^{n} \ln y_i$$

และถ้าหาร y_i แต่ละตัวด้วย geometric mean ของ y แล้ว จะได้ $\sum_i^n \ln y_i = 0$ ดังนั้นเทอมสุดท้ายจะหายไป และค่ามากที่สุดของ L เท่ากับ $(n/2) \ln \sigma^2_{\varepsilon|\lambda}$ เมื่อ $\sigma^2_{\varepsilon|\lambda}$ เป็นผลรวมกำลังสองของความผิดพลาดจากการถดถอยของ $y^{(\lambda)}$ บน x_i จะได้ว่าขั้นตอนต่างๆ มีดังนี้

1. หาร y แต่ละค่าด้วยค่ากลางเรขาคณิตของ y
2. สำหรับค่าของ λ ทำการถดถอย $y^{(\lambda)}$ บน x_i และคำนวณผลรวมกำลังสองของค่าผิดพลาด
3. เลือกค่า λ สำหรับ $\sigma^2_{\varepsilon|\lambda}$ ที่น้อยที่สุด จะได้ว่า $\hat{\lambda}$ เป็น maximum likelihood ของ λ

ค่าผิดพลาดมาตรฐานของสัมประสิทธิ์การถดถอยสำหรับค่า λ ใดๆสามารถหาได้จากโปรแกรมการถดถอย ซึ่งเป็นค่าผิดพลาดมาตรฐานที่มีเงื่อนไข บนค่าสมมติ λ ค่าผิดพลาดมาตรฐาน หรือช่วงความเชื่อมั่นสำหรับ $\hat{\lambda}$ สามารถหาได้จากการหาส่วนกลับของ likelihood ratio test statistics และเมื่อต้องการที่จะทดสอบสมมติฐาน $\lambda = \lambda_0$ จะได้ว่า

$$\theta = \frac{\max L(\lambda_0, \underset{\sim}{\beta}_0, \sigma^2)}{\max L(\lambda, \underset{\sim}{\beta}, \sigma^2)}$$

$$= \left[\frac{\hat{\sigma}^2_{\varepsilon|\lambda}}{\sigma^2_{\varepsilon|\lambda}}\right]^{\frac{n}{2}}$$

และ $-2 \ln \sim x^2_{(1)}$ ดังนั้น ถ้าต้องการ 95 เปอร์เซนต์ของช่วงความเชื่อมั่นสำหรับ $\hat{\lambda}$ จะต้องพิจารณาทุกๆ ค่าของ λ ซึ่ง $-2 \ln \theta$ มีค่าน้อยกว่า 3.84 เพราะว่า $\Pr(-2 \ln \theta < 3.84) = 0.95$ จะได้ว่า

$$n \ln \hat{\sigma}_{\varepsilon|\lambda} - n \ln \hat{\sigma}_{\varepsilon|\lambda} < 3.84$$

เมื่อประเมินค่า $\hat{\sigma}_{\varepsilon|\lambda}$ สำหรับ λ ที่แตกต่างกันในการหา ML ของตัวประมาณ $\hat{\lambda}$ แล้วจะสามารถหาพิสัย (range) ของ λ สำหรับช่วงความเชื่อมั่น 95 เปอร์เซนต์นี้ได้

ในกรณีพิเศษสามารถแปลงค่าตัวแปรอิสระดังนี้

$$x^{(\lambda)} = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} & ; \lambda \neq 0 \\ \ln x & ; \lambda = 0 \end{cases}$$

และในบางกรณีจะสามารถแปลงค่าตัวแปรตาม $\underset{\sim}{y}$ และตัวแปรอิสระ X โดยใช้ตัวเดียวกันบน $\underset{\sim}{y}$ และตัวแปรอิสระทุกตัว ดังนี้คือ

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & ; \lambda \neq 0 \\ \ln y_i & ; \lambda = 0 \end{cases}$$

และ

$$x_i^{(\lambda)} = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & ; \lambda \neq 0 \\ \ln x_i & ; \lambda = 0 \end{cases}$$

ถ้า $\lambda = 1$ และ $\lambda = 0$ จะได้รูปแบบของสมการเชิงเส้นและสมการลอกการิทึมดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

และ

$$\ln y_i = \alpha_0 + \alpha_1 \ln x_{i1} + \alpha_2 \ln x_{i2} + \ldots + \alpha_k \ln x_{ik} + \varepsilon_i$$

Raymond J. Carroll (ค.ศ.1978) ได้ศึกษาถึงวิธีการที่แกร่งสำหรับการทดสอบการแปลงข้อมูลที่ลู่เข้าสู่ภาวะปกติโดยประมาณ ดังนั้นค่าผิดพลาดอาจเข้าใกล้ภาวะปกติเท่านั้น Hinkley (ค.ศ.1975) ได้กล่าวว่ามันเป็นข้อบกพร่องที่ไม่พิจารณาความเป็นไปได้ถึงเรื่องนี้ Carroll (ค.ศ.1980) จึงพิจารณาวิธีการใหม่โดยอาศัยความรู้พื้นฐานการแปลงข้อมูลของ Box และ Cox กับวิธีการที่แกร่ง กล่าวคือ จะแทนทฤษฎีในภาวะน่าจะเป็นเดิมเมื่อการแจกแจงเป็นแบบปกติด้วยความหนาแน่นเป็นแบบปกติที่มีศูนย์กลาง และแบบ Normal Central and Exponential Tail density

$$L(\underset{\sim}{\beta}, \sigma, \lambda) = \sigma^{n} \sum_{i}^{n} \exp\{-\rho\ [\frac{(y_i^{\lambda} - \underset{\sim}{x}_i' \underset{\sim}{\beta})}{\sigma}] + (\lambda - 1) \ln y_i\}$$

สำหรับ λ ที่คงที่จะกระทำเช่นเดียวกับปัญหาที่เกี่ยวกับตำแหน่ง โดยการหาค่าเริ่มต้นของค่า σ ซึ่งทำให้สมการข้างล่างมีค่ามากที่สุด

$$\sum_{i}^{n} \psi\{(y_i^{\lambda} - \underset{\sim}{x}_i' \underset{\sim}{\beta}) / \sigma\}\ x_i = 0$$

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ จะใช้วิธี M-estimator เพื่อจะประมาณค่า $\underset{\sim}{\hat{\beta}}$ และ $\hat{\lambda}$ ซึ่งจะให้ค่า $\hat{\sigma}^2_{\varepsilon|\lambda}$ ที่มีค่าน้อยที่สุด

2.5 ผลงานวิจัยที่เกี่ยวข้อง

ผู้ที่ศึกษาเกี่ยวกับเรื่องวิธีการที่มีความแกร่งสำหรับการถดถอยเชิงพหุและการพิจารณาสเกลที่แกร่งนั้นมีไม่มากนัก แต่มีนักวิจัยบางท่านที่ได้ศึกษาเกี่ยวกับวิธีการต่างๆ ที่ใช้ในการประมาณสัมประสิทธิ์การถดถอยเชิงเส้น และได้เสนอผลงานวิจัยที่เกี่ยวข้องซึ่งน่าสนใจ พร้อมทั้งข้อสรุปต่างๆ ดังนี้

ก) Forsythe (ค.ศ.1972 หน้า 159-166) ได้ศึกษาตัวประมาณที่แกร่งของสัมประสิทธิ์การถดถอยเชิงเส้น โดยวิธีหาค่าน้อยที่สุดของส่วนเบี่ยงเบนยกกำลังที่ p (p^{th} power deviation หรือ Lp-norm regression) โดยที่พารามิเตอร์ของตัวแบบถูกเลือกเพื่อหาค่าน้อยที่สุดของ $\sum_{i}^{n} |\varepsilon_i|^p$; $(1 < p < 2)$ สำหรับการคำนวณหาสัมประสิทธิ์การถดถอยพหุ เมื่อ $1 < p < 2$ จะทำให้เป็นโปรแกรมนอนลิเนียร์ (nonlinear) Forsythe ได้ศึกษากระบวนการนี้สำหรับการถดถอยเชิงเส้นด้วยการจำลองแบบมอนติคาร์โล โดยใช้การจำลองแบบการแจกแจงปกติปลอมปนหลายๆ แบบ เขาสังเกตว่า $p = 1.5$ เป็นการเลือกที่ผสมผสานดีในการที่จะนำมาใช้แทนวิธีกำลังน้อยที่สุด เมื่อค่าผิดพลาดมีการแจกแจงไม่เป็นแบบปกติ และเมื่อการแจกแจงของค่าผิดพลาดเป็นแบบปกติการใช้ $p = 1.5$ ทำให้ผลลัพธ์ที่ได้จากการประมาณมีประสิทธิภาพ 90

เปอร์เซนต์เมื่อเทียบกับวิธีกำลังสองน้อยที่สุด

ข) Andrew (ค.ศ.1974 หน้า 523-615) ได้ศึกษาวิธีการที่แกร่งสำหรับการถดถอยเชิงเส้นซึ่งมีฟังก์ชัน ϕ เป็นแบบ sine โดยพิจารณาตัวแบบการถดถอยเชิงเส้นเมื่อ $\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยเชิงเส้นที่ไม่ทราบค่า X เป็นเวกเตอร์ของแถวตัวแปรอิสระ σ เป็นสเกลพารามิเตอร์ที่ไม่ทราบค่า และ ε_i เป็นค่าผิดพลาด ดังนั้นการประมาณสัมประสิทธิ์การถดถอยเชิงเส้น เป็นการหาค่ามากที่สุดของ

$$\sum_{i}^{n} \phi \{ r_i(\underset{\sim}{\beta}) / s(\underset{\sim}{\beta}) \}$$

เมื่อ $r_i(\underset{\sim}{\beta}) = y_i - \underset{\sim}{X}_i' \underset{\sim}{\beta}$ และ $S(\underset{\sim}{\beta}) = \text{median}\{|r_i(\underset{\sim}{\beta})|\}$
โดยที่ฟังก์ชัน ϕ อยู่ในรูปของ

$$\rho(Z) = \begin{cases} C[\,1 + \cos(Z/C)\,] & ; |Z| < C\pi \\ 0 & ; |Z| > C\pi \end{cases}$$

Daniel และ Wood (ค.ศ.1971) ได้ทำการทดลองโดยการนำข้อมูลของ Brownee (ค.ศ.1985.sec.13.12) เมื่อพิจารณาจากตัวอย่าง 21 ค่า และตัวแปรอิสระ 3 ตัวซึ่งมีค่าผิดพลาดที่ใหญ่ผิดปกติ ในค่าสังเกตที่ 1, 3, 4 และ 21 หาสมการการประมาณได้ 4 สมการ โดยที่สมการที่ 1 ใช้วิธีกำลังสองน้อยที่สุด สมการที่ 2 ใช้วิธีกำลังสองน้อยที่สุดแต่ตัดค่าผิดปกติออกไป สมการที่ 3 ใช้วิธีที่แกร่งของ Andrew สมการที่ 4 ใช้วิธีที่แกร่งของ Andrew แต่ตัดค่าผิดปกติออกไป ผลปรากฏว่า สมการที่ 3 และ 4 ให้ผลเหมือนกัน และให้ค่า standard error ของการประมาณ $\underset{\sim}{\beta}$ ต่ำที่สุด สมการที่ 2 ให้ค่า standard error น้อยกว่าสมการที่ 1

ค) David A. Lax (ค.ศ.1985 หน้า 736-741) ได้ศึกษาตัวประมาณที่แกร่งของสเกลเมื่อค่าสังเกตที่จำกัดมีการแจกแจงแบบสมมาตรหางยาว โดยเสนอผลงานการจำลองการศึกษาของตัวประมาณที่แกร่งของสเกลเมื่อค่าสังเกตมีการแจกแจงแบบสมมาตรหางยาว และทำ

การทดสอบวงศ์ของค่าประมาณ เมื่อใช้ขนาดตัวอย่าง n = 20 จากการแจกแจงหลายๆ รูปแบบในที่นี้วงศ์ของตัวประมาณชนิด A (the family of A-estimators) ที่มีค่าความแปรปรวนโดยประมาณของตัวประมาณค่าแหน่งชนิด M เมื่อ n มีค่ามาก จะมีความแกร่งที่มากกว่า the sample standard deviation, MAD, trimmed standard deviation, และตัวประมาณชนิด M ของค่าสเกล ซึ่งความสำเร็จอย่างสูงของตัวประมาณชนิด A ที่ใช้ฟังก์ชันถ่วงน้ำหนักแบบ biweight จะเป็นพื้นฐานของการหาค่าประมาณค่าแหน่งที่มีความแกร่งสูง

Lax (1975a,b) ได้ทำการศึกษาตัวประมาณสเกล 150 ชนิด โดยที่ 17 ชนิดจากตัวประมาณสเกลที่ทำการศึกษาได้ถูกเลือกขึ้นมา เพราะเป็นตัวประมาณสเกลทั่วไปที่ใช้กัน และให้ผลจากการจำลองศึกษาได้อย่างมีประสิทธิภาพสูง

Tukey (1982) ได้เสนอรูปแบบการแจกแจงแบบสมมาตรหางยาวในการศึกษาเกี่ยวกับตัวประมาณสเกล คือ

1. การแจกแจงแบบปกติ (Normal distribution) ที่มีค่าเฉลี่ย 0 และค่าความแปรปรวน 1
2. การแจกแจงแบบสแลช (Slash distribution) ที่ทำการเลือกค่าสังเกตจากประชากรที่มีการแจกแจงในรูป Z/W เมื่อ $Z \sim N(0,1)$ และ $W \sim U(0,1)$ โดยที่ Z และ W เป็นอิสระกัน และการแจกแจงในรูปแบบนี้มีความคงเส้นคงวากับการแจกแจงแบบหางยาวของคอยช์ซี (Cauchy distribution)
3. การแจกแจงแบบวันไวล์ (One-Wile distribution) ที่ทำการเลือกค่าสังเกตขนาด n = 20 โดยที่ 19 ค่ามาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และค่าความแปรปรวน 1 ส่วนอีกค่ามาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และมีค่าความแปรปรวน 100

Lax (1975b) ได้เสนอให้ใช้สมการทดสอบประสิทธิภาพของตัวประมาณสเกลที่ได้จากการแจกแจงทั้ง 3 ดังนี้

$$E_i = 100 \times V_{min}/V$$

เมื่อ E_i คือค่าประสิทธิภาพของลอกของค่าความแปรปรวนจากการแจกแจงที่ i

V_{min} คือค่าลอกของความแปรปรวนที่ทราบค่าที่น้อยที่สุดของตัวประมาณสเกล ที่ได้จากการทำซ้ำในการแจกแจงนั้นๆ

และ V คือค่าความแปรปรวนของตัวประมาณสเกลนั้น

หลังจากนั้นนำค่า E_i ของการแจกแจงทั้งสามมาเปรียบเทียบหาค่าที่น้อยที่สุด เรียกว่า Triefficiency

ผลปรากฏว่ามีตัวประมาณสเกล 7 ชนิดได้แก่ the biwight A-estimator ที่ c = 9 และ c = 10 the modified biweight A-estimator the modified sine A-estimator the iterated Huber M-estimator ที่ b = 1.4 the sample standard deviation และ the trimmed standard deviation ถูกกำหนดให้ไม่มีความสำคัญ และมีเพียง 3 ชนิดเท่านั้นได้แก่ the biweight A-estimator ที่ c = 9 และ c = 10 และ the modified sine A-estimator ที่ให้ค่า Triefficiency สูงที่สุดประมาณ 82% ในขณะที่ตัวประมาณแบบ the sample standard deviation ให้ค่าประสิทธิภาพสูง เมื่อค่าสังเกตมีการแจกแจงแบบปกติ the trimmed standard deviation ให้ค่าประสิทธิภาพสูง เมื่อค่าสังเกตมีการแจกแจงแบบ One-Wile และสูญเสียประสิทธิภาพเพียงเล็กน้อย เมื่อค่าสังเกตมีการแจกแจงแบบปกติ แต่สูญเสียประสิทธิภาพมาก เมื่อค่าสังเกตมีการแจกแจงแบบ Cauchy หรือ Slash MAD และ Gaussian Skip ให้ผลดี เมื่อค่าสังเกตมีการแจกแจงแบบ Cauchy หรือ Slash แต่จะมีประสิทธิภาพต่ำ เมื่อค่าสังเกตมีการแจกแจงแบบปกติ และ One-Wile the biweight A-estimator ที่ c = 9 และ c = 10 ให้ผลดีกว่า MAD เมื่อค่าสังเกตมีการแจกแจงแบบปกติ และ Slash อย่างไรก็ตามตัวประมาณสเกลแบบ A-estimator ยังคงให้ค่าประมาณที่มีความแกร่งดีกว่าตัวประมาณสเกลอื่นๆ เมื่อค่าสังเกตมีการแจกแจงแบบสมมาตรหางยาว เพราะ สามารถนำมาใช้ได้เลยโดยไม่ต้องระวังเรื่องการเป็นค่าบวก ซึ่งไม่เหมือนกับ ตัวประมาณสเกลแบบ M-estimator

ง) Mervry J. Silvapulle (1985 หน้า 1490-1497) ได้ศึกษาถึงพฤติกรรมของ

ตัวประมาณความแกร่งของสมการถดถอยและตัวแปรสเกลที่กำหนดเมตริกซ์ X คงที่เมื่อ n มีค่ามาก ซึ่งโดยทั่วไปแล้วการศึกษาในเรื่องตัวประมาณชนิด M ค่าสเกล S ที่ประมาณขึ้นมาจะถูกเลือกให้ไม่มีอิทธิพลกับความคลาดเคลื่อนที่มีขนาดใหญ่ และฟังก์ชันการสูญเสียที่กำหนดให้มีค่าน้อยที่สุดเมื่อให้ค่าสเกลคงที่อยู่ในรูปสมการดังนี้

$$Q(\beta, \sigma) = \Sigma\rho\{(y_i - X_i\beta)/\sigma\}\sigma + A_n\sigma$$

ที่ $\rho > 0$, $\rho(0) = 0$, $|t|^{-1}\rho(t) \rightarrow k$ ขณะที่ $|t| \rightarrow \infty$ สำหรับ $k > 0$ และ $\{A_n\}$ คือ อันดับของค่าคงที่ที่เหมาะสม

ผลปรากฏว่าการเลือกค่าสเกล S ขึ้นกับปัญหาที่รองลงไปที่ได้จากการเลือก $n^{-1}A_n$ ในตัวประมาณแบบ HD (Huber และ Dutter Estimator) โดยที่ค่า S อยู่ในช่วง $(0, \infty)$ และฟังก์ชั่นการสูญเสียของ β จะครอบคลุมขอบเขตจากตัวประมาณที่แกร่งอย่างแท้จริง (the extremely robust minimum absolute deviation estimator) ถึงตัวประมาณที่ไม่แกร่งอย่างแท้จริงแบบวิธีกำลังสองน้อยที่สุด (the extremely nonrobust least square estimator) เหมือนกับ $n^{-1}A_n$ ที่มีช่วงจาก $\max\{\chi(\infty)$, $\chi(-\infty)\}$ ถึง 0 ซึ่งตัวประมาณแบบ HD ของ $\underset{\sim}{\beta}$ ก็ให้ค่าครอบคลุมในช่วงเดียวกัน

Huber (1977) ได้สนับสนุนแนวความคิดที่ว่า ค่า A_n จะมีค่าเป็น $(n - p)E(\chi)$ ที่นำมาพิจารณาฟังก์ชันการแจกแจงว่าใกล้เคียงกับการแจกแจงของความคลาดเคลื่อน และกล่าวต่อว่าค่าความคลาดเคลื่อนนั้นควรมีการแจกแจงใกล้เคียงการแจกแจงแบบปกติ นอกจากนี้ผลงานของ Heathcote และ Silvapulle (1981) มีส่วนสำคัญในการสนับสนุนสมการตัวประมาณ HD สำหรับตำแหน่งและสเกลด้วย $A_n = nE_\psi(\chi)$ โดยการหาค่าน้อยที่สุดของฟังก์ชันการสูญเสียที่เหมือนกับเกณฑ์ของ Cramer-von Mises ณ.E_ψ คือค่าคาดหวังของการแจกแจง $(\psi + 1)/2$ ดังนั้น $A_n = nE_\psi(\chi)$ อาจเป็นทางเลือกที่เป็นไปได้ที่ดีสำหรับตัวแบบการถดถอย จากผลงานที่ผ่านมาเมื่อกลุ่มของข้อมูลแตกต่างกันและเลือก $\psi = \{2(\text{normal})-1\}$, $\{2(\text{logistic})-1\}$ ที่ $A_n = nE_\psi(\chi)$ ซึ่งทั้งฟังก์ชั่นการสูญเสีย และวิธี HD จะนำไปสู่คำตอบที่เหมาะสมภายใต้ค่าความแปรปรวนของตัวแปรการถดถอยของฟังก์ชันความสูญเสียเมื่อ n มีค่ามาก

ในการประมาณ $(\hat{\beta}, \hat{\sigma})$ โครงสร้างของ Huber และ Dutter (1974) ปฏิบัติได้ดีในสถานการณ์ที่แตกต่างกัน และพบว่าเมื่อไรก็ตามที่ค่าเริ่มต้นต่างจาก $(\hat{\beta}, \hat{\sigma})$ มาก เวลากระทำซ้ำเรื่อยๆ จะได้ค่าที่เข้าใกล้ $(\hat{\beta}, \hat{\sigma})$ อย่างรวดเร็ว

จ) ปราณี รัตนัง (2530) ได้ศึกษาการประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบเบ้และมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ โดยการเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay และเกณฑ์การเปรียบเทียบที่ใช้คือ ค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง

ผลจากการศึกษาคือ

1. กรณีค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ เช่นการแจกแจงแบบปกติปลอมปน และการแจกแจงแบบที จะได้ว่าสเกลแฟกเตอร์ เปอร์เซนต์ของการปลอมปนจะมีอิทธิพลจากมากไปน้อยที่ทำให้วิธี M-estimator ซึ่งใช้เกณฑ์ของ Ramsay ดีกว่าวิธีกำลังสองน้อยที่สุด ระดับความเป็นอิสระก็มีอิทธิพลที่ทำให้ได้ข้อสรุปเช่นเดียวกัน

2. กรณีค่าผิดพลาดมีการแจกแจงแบบเบ้ ได้แก่การแจกแจงแบบลอกนอร์มอลแกมมาและไวบูลล์ จะได้ว่าวิธีกำลังสองน้อยที่สุดและวิธี M-estimator สามารถใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุใกล้เคียงกัน เมื่อใช้เทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox ในการแปลงข้อมูลให้เข้าสู่ภาวะปกติ