บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้สิ่งที่สนใจศึกษาในการวิจัยครั้งนี้ คือ การประมาณค่าสัมประสิทธิ์การถดถอยพหุของตัวแบบการถดถอยเชิงเส้น เมื่อค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติและการแจกแจงแบบเบ้ สำหรับค่าผิดพลาดที่มีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติได้ทำการศึกษาเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay และสำหรับการแจกแจงแบบเบ้จะทำการศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยใช้การแปลงที่ใช้การยกกำลังของ Box และ Cox (1964 : 211 -243) ในการแปลงข้อมูลให้มีการแจกแจงเข้าใกล้ภาวะปกติเสียก่อน และจะทำการประมาณสัมประสิทธิ์การถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุด ในที่นี้จะศึกษาเปรียบเทียบวิธีกำลังสองน้อยที่สุด กับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay โดยจะกล่าวถึงรายละเอียดในหัวข้อต่อไปนี้

2.1 ข้อสมมติทั่วไปของตัวแบบการถดถอยเชิงเส้น

2.2 วิธีกำลังสองน้อยที่สุด (Least square method)

2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา (Ordinary least square method)

2.2.2 วิธีกำลังสองน้อยที่สุดแบบทั่วไป (Generalized least square method)

2.2.3 วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted least square method)

2.3 วิธีที่ใช้ตัวประมาณชนิด M (M-estimator method)

2.4 การแปลงข้อมูลด้วยการแปลงที่ใช้การยกกำลัง (Power Transformation)

2.5 ผลงานวิจัยที่เกี่ยวข้อง

2.1 ข้อสมมติทั่วไปของตัวแบบการถดถอยเชิงเส้น

ในการสมมติความสัมพันธ์เชิงเส้นระหว่างตัวแปร y และตัวแปรอิสระ m ตัว ของ $x_1, x_2, \ldots, x_m$ และค่าผิดพลาด ε นั่นคือ ถ้ามีตัวอย่าง n ค่าสังเกตของ y และค่า x เราสามารถเขียนสมการได้เป็น

$$(2.1.1) \qquad y_i = \beta_o + \beta_1 X_{i1} + \ldots + \beta_m X_{im} + \varepsilon_i \; ; \; i = 1, \ldots, n$$

ค่าสัมประสิทธิ์ $\underset{\sim}{\beta}$ และค่าพารามิเตอร์ของการแจกแจงของ $\underset{\sim}{\varepsilon}$ ไม่ทราบค่า ปัญหาคือเราต้องการประมาณ $\underset{\sim}{\beta}$ ที่ไม่ทราบเหล่านี้เมื่อมี n สมการของ (2.1.1) ซึ่งเราสามารถเขียนในรูปของเมตริกซ์ดังนี้

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

เมื่อ

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & x_{n1} & & x_{nm} \end{bmatrix}, \quad \underset{\sim}{\beta} = \begin{bmatrix} \beta_o \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_m \end{bmatrix}, \quad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

ซึ่งมีข้อสมมติคือ

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

(2.1.4) $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2 In$

(2.1.5) X เป็นชุดของค่าตัวเลขคงที่

(2.1.6) และมีอันดับ m<n

จากข้อสมมติ (2.1.3) $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ นั่นคือ ตัวแปร ε_i จะมีค่าคาดหวังเป็นศูนย์สำหรับทุกค่าของ i, i = 1,...,n

จากข้อสมมติ (2.1.4) จะได้ว่า

$$E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \cdots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & & E(\varepsilon_2\varepsilon_n) \\ \vdots & & & \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & & E(\varepsilon_n^2) \end{bmatrix}_{n \times n}$$

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

$$= \sigma^2 In$$

ตามข้อสมมติ (2.1.5) เมตริกซ์ X ต้องเป็นตัวเลขคงที่ หมายถึง การสุ่มตัวอย่างที่ซ้ำกันจะเกิดค่าหลายค่าในเวกเตอร์ $\underset{\sim}{y}$ ซึ่งแปรผันในเวกเตอร์ $\underset{\sim}{\varepsilon}$ และจำนวนค่าสังเกตของ X จะต้องมีจำนวนเกินจำนวนพารามิเตอร์ที่จะประมาณ ตามข้อสมมติ (2.1.6) ทั้งนี้ จะต้องไม่มีความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระด้วย เราเรียกความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระว่า multicollinearity

2.2 วิธีกำลังสองน้อยที่สุด (Least square method)

2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา (Ordinary least square method)

วิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นวิธีนี้มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of Linear Estimation) ที่เกิดขึ้นโดย คาร์ล เฟดริก เกาส์ (Karl Friedrick Gauss) ในปี ค.ศ.1777 - 1855 และอังเดร แอนดรีวิช มาร์คอฟ (Andrei Andrewich Markov) ในปี ค.ศ.1856 - 1922 โดยมีหลักในการประมาณค่าสัมประสิทธิ์คือ ทำให้ผลบวกกำลังสองของค่าผิดพลาดมีค่าน้อยที่สุดซึ่งแสดงรายละเอียดดังนี้

นิยาม 2.1 จากสมการ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}\ ,\ {}^{2}In)$ ตัวประมาณกำลังสองน้อยที่สุดของ $\underset{\sim}{\beta}$ คือ $\hat{\underset{\sim}{\beta}}$ ที่ทำให้ผลบวกกำลังสองของค่าผิดพลาด (Sum square errors) หรือ SSE มีค่าน้อยที่สุด

จากนิยาม 2.1 จะทำการหาตัวประมาณกำลังสองน้อยที่สุดได้ดังนี้

กำหนด $$\hat{\underset{\sim}{\beta}} = (\hat{\beta}_o\ ,\ \hat{\beta}_1\ ,\ \ldots,\ \hat{\beta}_m)'$$

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

ดังนั้น $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์แนวตั้งของค่าผิดพลาด n ค่าของ

จาก (2.2.1) ผลบวกกำลังสองของค่าผิดพลาดคือ

$$\begin{aligned} SSE &= \hat{\underset{\sim}{\varepsilon}}\hat{\underset{\sim}{\varepsilon}} \\ &= (y - X\hat{\underset{\sim}{\beta}})\ (y - X\hat{\underset{\sim}{\beta}}) \\ (2.2.2)\quad &= \underset{\sim}{y}\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}X\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}X\ X\hat{\underset{\sim}{\beta}} \end{aligned}$$

การหาค่าน้อยที่สุดของผลบวกกำลังสองของค่าผิดพลาดทำได้โดยการหาอนุพันธ์ (differentiate) เทียบกับ $\hat{\underset{\sim}{\beta}}$ แล้วกำหนดให้เท่ากับ $\underset{\sim}{0}$ ดังนี้

$$\begin{aligned} \frac{\partial}{\partial\hat{\underset{\sim}{\beta}}}\ (\underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}X\underset{\sim}{y}' + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}) &= \underset{\sim}{0} \\ -\ 2\ X'y + 2X'X\hat{\underset{\sim}{\beta}} &= \underset{\sim}{0} \\ (2.2.3)\quad \hat{\underset{\sim}{\beta}} &= (X'X)^{-1}\ X\underset{\sim}{y}' \end{aligned}$$

เมื่อ $(XX')^{-1}X'$ เป็นเมตริกซ์ของค่าคงที่ สมาชิกใน $\hat{\underset{\sim}{\beta}}$ เป็นฟังก์ชันเชิงเส้นของ $\underset{\sim}{y}$ นั่นคือ $\hat{\underset{\sim}{\beta}}$ เป็นตัวประมาณเชิงเส้น และแทน (2.1.2) ใน (2.2.3) เราได้

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

$$(2.2.4) \qquad = \beta + (X'X)^{-1} X'\varepsilon$$

จะได้ว่า $E(\hat{\beta}) = \beta + (X'X)^{-1} X'E(\varepsilon) = \beta$ ถ้า $E(\varepsilon) = 0$ นั่นคือ β เป็นค่าที่ไม่เอนเอียง

และ $V(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$

$$= (X'X)^{-1} X'E(\varepsilon\varepsilon') X(X'X)^{-1}$$

$$(2.2.5) \qquad = \sigma^2 (X'X)^{-1} \qquad \text{โดยใช้} \quad E(\varepsilon\varepsilon') = \sigma^2 In$$

ถ้า β^* เป็นตัวประมาณเชิงเส้นตัวอื่นใดของ β ที่แตกต่างจาก $\hat{\beta}$ เราจะแสดงว่า $V(\beta^*) \geqslant V(\hat{\beta})$ สำหรับอสมการนี้จะแสดงโดย $V(\beta^*) - V(\hat{\beta})$ เป็นเมตริกซ์กึ่งบวกแน่นอน (positive semidefinite matrix) กล่าวคือสมาชิกทแยงมุมของ $V(\beta^*) - V(\hat{\beta})$ เป็นค่าที่มากกว่า 0 ทุกค่า นั่นคือ $V(\beta^*) \geqslant V(\hat{\beta})$ สำหรับ $i = 0,1,\ldots,m$

ตัวประมาณเชิงเส้นใด ๆ β^* สามารถเขียนได้เป็น

$$\beta^* = [(X'X)^{-1} X' + C] y$$

$$= \hat{\beta} + C y$$

$$= \beta + CX\beta + [(X'X)^{-} X' + C] \varepsilon$$

ถ้า β^* เป็นตัวประมาณไม่เอนเอียง $E(\beta^*) = \beta$ และความสัมพันธ์นี้เป็นจริงสำหรับค่าที่เป็นไปได้ของ β เราจะได้ $CX = 0$

$$V(\beta^*) = E(\beta^* - \beta)(\beta^* - \beta)'$$

$$= [(X'X)^{-1} X' + C] E(\varepsilon\varepsilon') (X'X)^{-1} X' + C$$

$$= \sigma^2 [(X'X)^{-1} + CC']$$

เทอมอื่นหายไปเพราะว่าเงื่อนไข $CX = 0$ นั่นคือ $V(\beta^*) - V(\hat{\beta}) = \sigma^2 CC'$ ซึ่งเป็นเมตริกซ์กึ่งบวกแน่นอน (positive semidefinite matrix) จากการพิสูจน์นี้จะได้ว่า

เป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (BLUE = Best linear unbiased estimator) ของ β

ส่วนผลบวกกำลังสองของค่าผิดพลาดสามารถเขียนได้เป็น

$$\hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{\beta})'(y - X\hat{\beta})$$

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

$$= y'y - \hat{\beta}X'y$$

2.2.2 <u>วิธีกำลังสองน้อยที่สุดแบบทั่วไป</u> (Generalized least square method)

กรณี $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}')$ เป็นเมตริกซ์บวกแน่นอน (positive definite) Ω ซึ่งแทน ในสมการ (2.2.5) แล้ว จากตัวแบบ (2.1.2) สามารถแปลงให้อยู่ในรูปตัวแบบ

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

เมื่อ $\underset{\sim}{y}^* = \Omega^{-\frac{1}{2}}\underset{\sim}{y}$

และ $X^* = \Omega^{-\frac{1}{2}} X$

และ $\underset{\sim}{\varepsilon}^* = \Omega^{-\frac{1}{2}} \underset{\sim}{\varepsilon}$

จะได้ว่า $E(\underset{\sim}{\varepsilon}^*\underset{\sim}{\varepsilon}^{*\prime}) = \Omega^{-\frac{1}{2}}E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}')\,\Omega^{-\frac{1}{2}} = I_n$ ในทำนองเดียวกันกับหัวข้อ 2.2.1 ค่า BLUE ของ $\underset{\sim}{\beta}$ จะเขียนได้เป็น

$$(2.2.6) \quad \underset{\sim}{\beta}_{GLS} = (X^{*\prime}X^*)^{-1}X^{*\prime}\underset{\sim}{y} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\underset{\sim}{y}$$

โดยที่ GLS แทน Generalized least squares

และความแปรปรวนของตัวประมาณจะมีค่าเป็น

$$(2.2.7) \quad V(\hat{\underset{\sim}{\beta}}_{GLS}) = \sigma^2(X^{*\prime}X^*)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}$$

ในกรณีนี้ตัวประมาณกำลังสองน้อยที่สุดแบบธรรมดา (ordinary least squares) : 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 ถ้า E( ) = 0 และความแปรปรวนหาได้จาก

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

ตัวประมาณ GLS มีประสิทธิภาพมากกว่าตัวประมาณ OLS ในความหมายว่า $V(\underset{\sim}{\beta}_{OLS}) - V(\underset{\sim}{\beta}_{GLS})$ เป็นเมตริกซ์กึ่งบวกแน่นอน (positive semidefinite matrix) ในกรณีที่ $\Omega \neq I_n$

2.2.3 <u>วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก</u> (Weighted least square method)

เมื่อค่าผิดพลาด $\underset{\sim}{\varepsilon}$ ไม่มีความสัมพันธ์กัน โดยที่ความแปรปรวนไม่เท่ากัน เมตริกซ์ของความแปรปรวนร่วมของ $\underset{\sim}{\varepsilon}$ แทนด้วย

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} \frac{1}{w_1} & \cdot & \cdot & \cdot & 0 \\ \cdot & \frac{1}{w_2} & & & \cdot \\ \cdot & & & & \cdot \\ 0 & \cdot & \cdot & & \frac{1}{w_n} \end{bmatrix}$$

เมื่อ $W = \Omega^{-1}$ และ Ω เป็นเมตริกซ์ทแยงมุม (diagonal matrix) จะได้ว่า W เป็นเมตริกซ์ทแยงมุมด้วย โดยมีสมาชิกตามแนวทแยงมุมเป็น $w_1, w_2, \ldots, w_n$ ซึ่งสมการปกติของ GLS อยู่ในรูปของ

$$(X'\Omega^{-1} X)\,\hat{\underset{\sim}{\beta}} = X'\Omega^{-1}\underset{\sim}{y}$$

จะได้ว่าสมการของ Weighted least square หรือ WLS อยู่ในรูปของ

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

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

$\therefore$ $\hat{\beta}$ เป็นตัวประมาณของ Weighted least square ค่า w_i ถูกเรียกว่าเป็น weights โดยค่าสังเกตที่มีค่า w_i น้อยจะมีความแปรปรวนมากกว่าค่าสังเกตที่มีค่า w_i มาก

2.3 <u>วิธีที่ใช้ตัวประมาณชนิด M</u> (M-estimator method)

ปี ค.ศ.1964 P.J. Huber ได้ศึกษาถึงฟังก์ชันของความผิดพลาดซึ่งเรียกว่า ตัวประมาณชนิด M (M-estimator) ตัวประมาณชนิด M จะประมาณพารามิเตอร์ของค่าน้อยที่สุดของ $\sum_{i=1}^{n} \rho\,(\varepsilon_i/s)$ เมื่อ ε_i เป็นค่าผิดพลาดของค่าสังเกตตัวที่ i และ s เป็นตัวประมาณที่เหมาะสมของการกระจายของตัวอย่างของ

ในตัวแบบการถดถอยเชิงเส้น เราสามารถหาตัวประมาณชนิด M จาก

$$(2.3.1)\qquad \min_{\underset{\sim}{\beta}} \sum_{i=1}^{n} (\varepsilon_i/s) = \min_{\underset{\sim}{\beta}} \sum_{i=1}^{n} \rho\,|(y_i - X_i\underset{\sim}{\beta})/s|$$

เมื่อ s เป็นตัวประมาณที่แกร่งของสเกล ซึ่งนิยมใช้มัธยฐานของค่าสัมบูรณ์ของความเบี่ยงเบน (median absolute deviation) หรือ MAD (mosteller and Tukey : 1977) และทำการปรับด้วยค่าคงที่ 0.6745 ซึ่งจะทำให้ s เป็นตัวประมาณที่ไม่เอนเอียงโดยประมาณของเมื่อ n มีขนาดใหญ่และการแจกแจงของค่าผิดพลาดเป็นแบบปกติ ดังนั้น s เขียนได้เป็น

$$s = \text{median}\ |\ \varepsilon_i - \text{median}\ (\varepsilon_i)|\ /0.6745$$

ในการที่จะหาค่าน้อยที่สุดของสมการ (2.3.1) จะหาอนุพันธ์ (differentiate) ของ ρ เทียบกับ $\underset{\sim}{\beta}$ แล้วกำหนดให้เท่ากับ 0 โดย $\psi = \rho'$ ดังนี้

$$(2.3.2) \qquad \sum_{i=1}^{n} x_{ij}\psi\left[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s\right] = \underset{\sim}{0} \quad , j = 1,2,\ldots,m$$

จะได้ว่า ฟังก์ชัน ψ ไม่เป็นฟังก์ชันเชิงเส้น วิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยส่วนใหญ่จะใช้วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักซ้ำหลายรอบ (iteratively reweighted least square) ของ Beaton and Tukey (1964) โดยจะต้องประมาณ และ s เราสามารถเขียนสมการ m สมการของ (2.3.2) ได้ดังนี้

$$\sum_{i=1}^{n} x_{ij}\psi\left[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s\right] = \sum_{i=1}^{n} x_{ij}\frac{\psi\left[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s\right]}{(y_i - \underset{\sim}{X}\underset{\sim}{\beta})/s}(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s = \underset{\sim}{0}$$

$$(2.3.3) \qquad \sum_{i=1}^{n} x_{ij}\, w_{io}\,(y_i - \underset{\sim}{X}_i\underset{\sim}{\beta}) = 0 \quad , j = 1,2,\ldots, m$$

$$w_{io} = \begin{cases} \dfrac{\psi\left|(y_i - x_i\,{}_o)/s\right|}{(y_i - \underset{\sim}{X}_i\underset{\sim}{\beta}_o)/s} & ; \ y_i \neq \underset{\sim}{X}_i\hat{\underset{\sim}{\beta}}_o \\ 1 & ; \ y_i = \underset{\sim}{X}_i\hat{\underset{\sim}{\beta}}_o \end{cases}$$

สมการเหล่านี้อาจเขียนในรูปเมตริกซ์ดังนี้

$$(2.3.4) \qquad X'W_oX\underset{\sim}{\beta} = X'W_o\underset{\sim}{Y}$$

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

ดังนั้น ตัวประมาณ $\hat{\underset{\sim}{\beta}}$ ที่หาได้ในครั้งแรกจะมีค่าเป็น

$$(2.3.5) \qquad \hat{\underset{\sim}{\beta}} = (X'W_oX)^{-1}X'W_o\underset{\sim}{Y}$$

จะทำครั้งถัดไปโดยคำนวณน้ำหนัก ซึ่งใช้ β แทน β และทำซ้ำ ๆ กันจนกระทั่งได้ค่า $\hat{\underset{\sim}{\beta}}$ ที่ค่อนข้างคงที่ Holland และ Welsch (1977) กล่าวว่ากระบวนการกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักทำซ้ำ ๆ กันนี้ต้องการเพียงโปรแกรมคอมพิวเตอร์ของวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักวิธีมาตรฐานเท่านั้น

พังก์ชันซึ่งใช้เกณฑ์ของความแกร่งได้มีผู้ศึกษาไว้หลายท่านด้วยกัน ซึ่งจะนำเสนอฟังก์ชันตามเกณฑ์ความแกร่งชนิดต่าง ๆ ในตารางที่ 2.1 และรูปที่ 2.1, 2.2 และ 2.3 โดย

เปรียบเทียบกับวิธีกำลังสองน้อยที่สุดซึ่งเป็นวิธีที่ไม่แกร่งด้วย Huber (ค.ศ.1964) ได้สร้างฟังก์ชัน Huber't ขึ้น โดยที่ t เป็นค่าคงที่สำหรับกำหนดขอบเขตค่าผิดพลาด ในที่นี้ใช้ $t = 2$ Ramsay (ค.ศ.1977) ได้สร้างฟังก์ชัน Ramsay Ea ขึ้น โดยมีลักษณะเป็นฟังก์ชันเอ็กซ์โปเนนเชียล ในที่นี้ใช้ $a = 0.3$ เป็นค่าคงที่สำหรับกำหนดขอบเขตของฟังก์ชัน โดยมีขอบเขตที่ $|\varepsilon_i/s| = 1/a$ ลักษณะของ ψ ของ Ramsay จะค่อย ๆ คล้อยลาดลงเรื่อย ๆ นั่นคือ ฟังก์ชัน ψ เข้าหาศูนย์เมื่อ $|\varepsilon_i|$ มีขนาดใหญ่ขึ้นเรื่อย ๆ Andrews et al (ค.ศ. 1972, 1974) ได้สร้างฟังก์ชันคลื่นของ Andrew Hampel ได้สร้างฟังก์ชัน Hampel (ค.ศ.1971, 1974) โดยฟังก์ชัน ψ ของทั้งสองวิธีเป็นแบบคล้อยลาดลงอย่างรวดเร็ว ลักษณะฟังก์ชันคลื่นของ Andrew เป็นคลื่นแบบ sine และฟังก์ชัน Hampel มีจุดเปลี่ยนแปลงขอบเขตของฟังก์ชัน 3 จุด ได้แก่ a, b, c เมื่อ $a = 1.7$, $b = 3.4$ และ $c = 8.5$

เกณฑ์ของฟังก์ชันที่แกร่งที่สนใจศึกษาวิจัยครั้งนี้คือ เกณฑ์ของ Ramsay โดยมีฟังก์ชันของ ρ และ ψ สำหรับชุดของตัวประมาณคือ

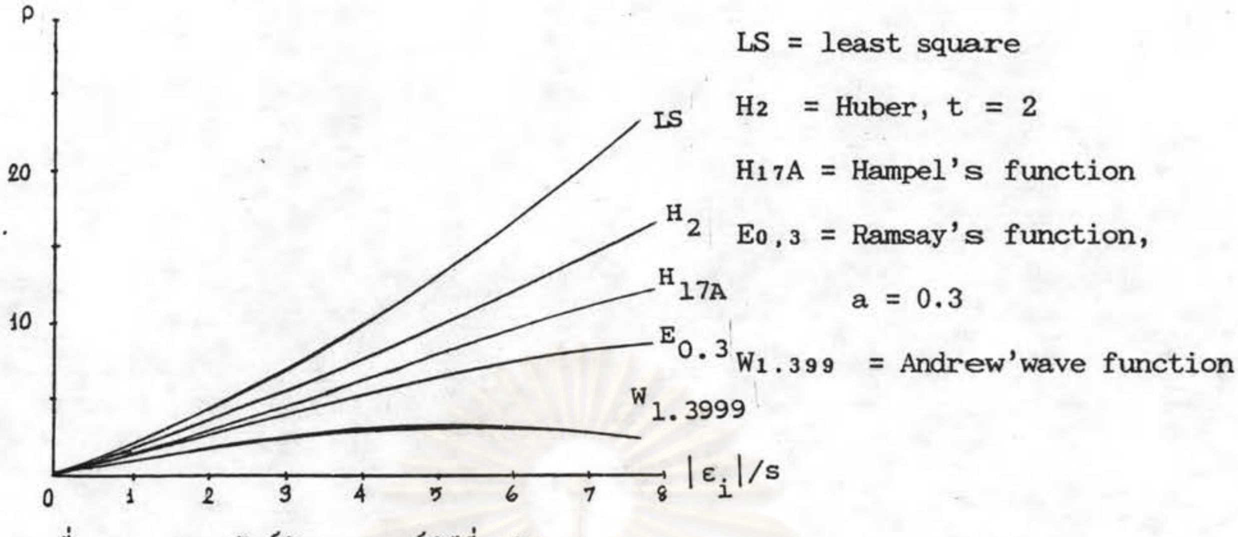
$$\rho(\varepsilon_i/s) = a^{-2} \quad 1-\exp(-a|\varepsilon_i|/s) \cdot (1+a|\varepsilon_i|/s)$$

$$\text{และ} \quad \psi(\varepsilon_i/s) = (\varepsilon_i/s) \exp(-a|\varepsilon_i|/s)$$

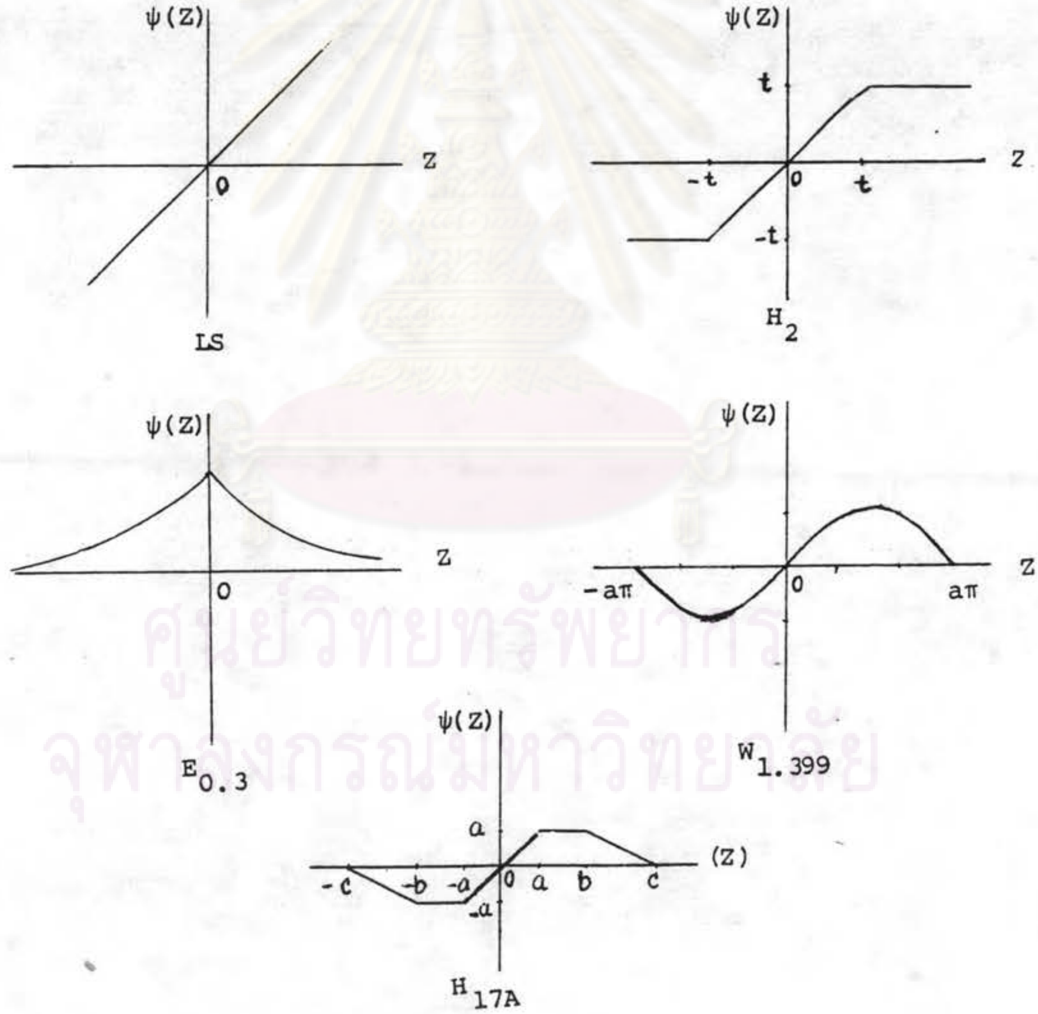
ซึ่งอ้างถึงตัวประมาณนี้ด้วยตัวประมาณ Ea โดยที่ $a = 0.3$ Ea จะมีขอบเขตและมีค่าเข้าสู่ขอบเขตเมื่อ $|\varepsilon_i|/s = 1/a$ ค่าผิดพลาดที่มีค่ามาก (extreme value) จะมีอิทธิพลลดลงเรื่อย ๆ และอิทธิพลเหล่านี้จะอยู่ในสัดส่วนที่ถูกตัดออกไป นั่นคือ ค่าสังเกตที่ผิดปกติมาก ๆ จะอยู่ในสัดส่วนที่ถูกตัดออกจากตัวอย่าง

ตารางที่ 2.1 ฟังก์ชันเกณฑ์ของความแกร่งชนิดต่าง ๆ

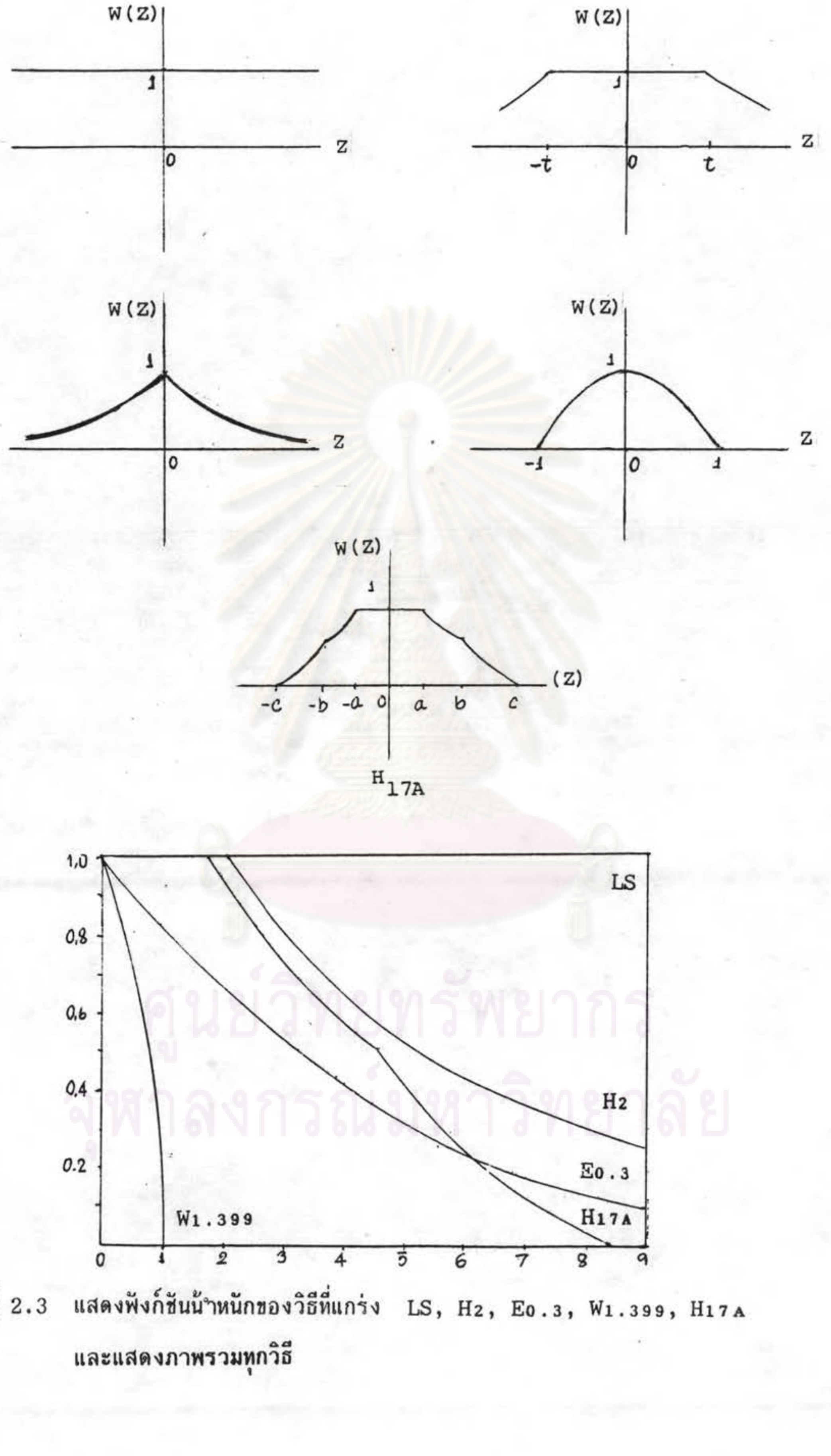
เกณฑ์วิธีการต่าง ๆ	$\rho(Z)$	$\psi(Z)$	$W(Z)$	ขอบเขต
วิธีกำลังสองน้อยที่สุด	$\frac{1}{2}Z^2$	Z	1.0	$\|Z\| < \infty$
Huber't ฟังก์ชัน $t = 2$	$\frac{1}{2}Z^2$ $\|Z\|\, t-\frac{1}{2}t^2$	Z $t\,\mathrm{sign}(Z)$	1.0 $\frac{t}{\|Z\|}$	$\|Z\| \leqslant t$ $\|Z\| > t$
Ramsay's Ea ฟังก์ชัน $a = 0.3$	$a^{-2}\left[1-\exp(-a\|Z\|)\right]\cdot(1+a\|Z\|)$	$Z\exp(-a\|Z\|)$	$\exp(-a\|Z\|)$	$\|Z\| < \infty$
Andrew's wave ฟังก์ชัน $a = 1.399$	$a\|1-\mathrm{COS}(Z/a)$ 0	$\mathrm{SIN}(Z/a)$ 0	$\frac{\mathrm{SIN}\ (Z/a)}{Z/a}$ 0	$\|Z\| \leqslant a\pi$ $\|Z\| > a\pi$
Hampel's 17 A ฟังก์ชัน $a = 1.7$ $b = 3.4$ $c = 8.5$	$\frac{1}{2}Z^2$ $a\|Z\| - \frac{1}{2}a^2$ $\left[\frac{a(c\|Z\|-\frac{1}{2}Z^2)}{c-b} - \frac{(Z)}{6}a^2\right]$ $a(b-c-a)$	Z $a.\mathrm{SIGN}(Z)$ $\left[a\mathrm{SING}(Z)\cdot(c-\|Z\|)\right]/\ c-b$ 0	1.0 $a\|Z\|$ $\frac{a(c-\|Z\|)}{Z\ (c-b)}$ 0	$Z \leqslant a$ $a < \|Z\| < b$ $b < \|Z\| \leqslant c$ $\|Z\| > c$



รูปที่ 2.1 แสดงพังก์ชันของเกณฑ์วิธีที่แกร่ง



รูปที่ 2.2 แสดงพังก์ชันอิทธิพล (ψ) ของเกณฑ์วิธีที่แกร่ง



รูปที่ 2.3 แสดงฟังก์ชันน้ำหนักของวิธีที่แกร่ง LS, H_2, $E_{0.3}$, $W_{1.399}$, H_{17A} และแสดงภาพรวมทุกวิธี

ผู้ศึกษาหลายท่าน เช่น Andrew (ค.ศ.1964), Hogg (ค.ศ.1979) และ Hocking (ค.ศ.1978) บันทึกไว้ว่า การเริ่มต้นด้วยค่า β_0 ในการประมาณที่แกร่งต้องเลือกอย่างระมัดระวัง และการใช้ผลจากวิธีกำลังสองน้อยที่สุดจะได้ผลดี นอกจากนี้ L1-norm อาจจะเป็นค่าเริ่มต้นที่ดีด้วย Andrew (ค.ศ.1974) และ Dutter (ค.ศ.1977) ได้แนะนำกระบวนการสำหรับหาค่าเริ่มต้นด้วย

ในปัจจุบันเป็นการยากที่จะให้ความสนับสนุนอย่างแข็งขันเกี่ยวกับโครงสร้างของความผิดพลาดของตัวประมาณ $\hat{\beta}$ จากการถดถอยที่แกร่งในขั้นสุดท้าย การพิจารณาถึงเมตริกซ์ของความแปรปรวนร่วมของ $\hat{\beta}$ มีความสำคัญถ้าเราทำโครงสร้างของช่วงที่เชื่อถือได้และทำการอ้างอิงตัวแบบอื่น Huber (1973) ได้แสดงว่าถ้าค่าลู่เข้าสู่ $\hat{\beta}$ มีการแจกแจงแบบปกติโดยประมาณด้วยเมตริกซ์ความแปรปรวนร่วม

$$\frac{\sigma^2 E\left[\psi^2(\varepsilon/\sigma)\right]}{\{E\left[\psi'(\varepsilon/\sigma)\right]\}^2}(X'X)^{-1}$$

เมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ โดยประมาณคือ

$$\frac{(ns)^2}{n-p}\cdot\frac{\sum_{i=1}^{n}\psi^2\left[(y_i - X_i'\hat{\beta})/s\right]}{\left\{\sum_{i=1}^{n}\psi'\left[(y_i - X_i'\hat{\beta})/s\right]\right\}^2}$$

และเมตริกซ์ความแปรปรวนร่วมที่ประมาณโดยโปรแกรมคอมพิวเตอร์วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก คือ

$$\frac{\sum_{i=1}^{n} W_i(y_i - X_i'\hat{\beta})^2}{n-p}(X'WX)^{-1}$$

คำแนะนำอื่น ๆ ของ Welsch (ค.ศ.1975) และ Hill (ค.ศ.1979) กล่าวว่า ไม่มีข้อตกลงทั่วไปเกี่ยวกับการประมาณสำหรับเมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ ที่ดีที่สุด ท่านทั้งสองกล่าวว่า เมตริกซ์ความแปรปรวนร่วมสามารถแสดงผลออกมาน้อยสำหรับเมตริกซ์ของ X ที่มีค่าผิดปกติ และเป็นเงื่อนไขที่ไม่ดี ได้แก่ เงื่อนไขเกี่ยวกับ multicollinearity สามารถจะมีผลกระทบต่อการประมาณการความถดถอยที่แกร่งด้วย ในหลายกรณีด้วยกันเราสามารถอ้างอิงเกี่ยวกับ $\hat{\beta}$ โดยใช้กระบวนการคล้ายคลึงกับทฤษฎีปกติได้

2.4 การแปลงข้อมูลด้วยการแปลงที่ใช้การยกกำลัง (Power Transformation)

การแปลงข้อมูลด้วยการแปลงที่ใช้การยกกำลัง (Power Transformation) ได้รวมการแปลงซึ่งใช้ลอกการิทึมในวงศ์ (family) ของการแปลง การแปลงข้อมูลอาจจะทำให้ข้อสมมติของความเป็นปกติใช้ได้สำหรับแปลงข้อมูลที่ใช้ศึกษาการแจกแจงของค่าผิดพลาดเป็นแบบเบ้

Tukey (ค.ศ.1957 : 602-632) เป็นท่านแรกที่ศึกษารายละเอียดเกี่ยวกับการแปลง ท่านได้แนะนำว่า การแปลงข้อมูลจะช่วยให้ตัวแบบเข้าใกล้เชิงเส้น โดยที่ข้อมูลจะมีการกระจายเท่ากัน และเป็นแบบปกติมากขึ้น

Box และ Cox (ค.ศ.1962 : 211-243) ได้พิจารณาการแปลงในตัวแปรตามและสามารถทำในตัวแปรอิสระด้วย การแปลงของเขาอยู่ในรูปของ

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

ซึ่งเป็นข้อดีของการแปลง เมื่อเปรียบเทียบกับการแปลงอย่างง่าย y_i^{λ} ซึ่งจะต่อเนื่องใน $\lambda = 0$ ดังนั้น

$$L_{t_{\lambda \to 0}} \frac{y_i^{\lambda} - 1}{\lambda} = \log_e y_i$$

Box และ Cox พิจารณาตัวแบบของการถดถอย

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

เมื่อ $\varepsilon_i \sim I_n(0, \sigma^2)$ ดังนั้น ฟังก์ชันความหนาแน่นร่วมของค่าสังเกต $y_1, y_2, \ldots, y_n$ เป็น

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

โดยที่ J เป็น จาโคเบียนของการแปลงจากตัวแปร y_i เป็น $Y_i^{(\lambda)}$ ดังนั้น

$$J = \prod_{i=1}^{n} \left| \frac{dy_i^{(\lambda)}}{dy_i} \right| = \prod_{i=1}^{n} y_i^{\lambda - 1}$$

จะได้ว่า ฟังก์ชัน log-likelihood อยู่ในรูปของ

$$\mathrm{Log}_e L = - \frac{n}{2} \log_e 2 - \frac{n}{2} \log_e - \frac{1}{2\sigma^2} \sum_{i=1}^{n_1} |y_i^{(\lambda)} - \underset{\sim}{x}_i \underset{\sim}{\beta}|^2 + (\lambda-1) \sum_{i=1}^{n} \log_e y_i$$

ถ้าเราหารแต่ละ y_i ด้วย geometric mean ของ y แล้ว เราจะได้ $\sum_{i=1}^{n_1} \log_e y_i = 0$ ดังนั้น เทอมสุดท้ายจะหายไป และค่ามากที่สุดของ L เท่ากับ $\frac{n}{2} \log_e \sigma^2_{\varepsilon|\lambda}$ เมื่อ $\hat{\sigma}^2_{\varepsilon|\lambda}$ เป็นผลรวมกำลังสองของค่าผิดพลาดจากการถดถอยของ $Y^{(\lambda)}$ บน x_i ดังนัน ขั้นตอนต่าง ๆ มีดังนี้

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

ค่าผิดพลาดมาตรฐานของสัมประสิทธิ์การถดถอยสำหรับค่า λ ใด ๆ สามารถหาได้จากโปรแกรมการถดถอย ซึ่งเป็นค่าผิดพลาดมาตรฐานที่มีเงื่อนไข บนค่าสมมติ λ ค่าผิดพลาดมาตรฐาน หรือช่วงความเชื่อมั่นสำหรับ $\hat{\lambda}$ สามารถหาได้จากการหาส่วนกลับของ likelihood-ratio test statistics

เราต้องการที่จะทดสอบสมมติฐาน $\lambda = \lambda_o$ จะได้ว่า

$$\Theta = \frac{\max L(\lambda_o, \beta, \sigma^2)}{\max L(\lambda, \beta, \sigma^2)} = \left[\frac{\hat{\sigma}^2_{\varepsilon|\hat{\lambda}}}{\hat{\sigma}_{\varepsilon|\lambda_0}} \right]^{\frac{n}{2}}$$

และ $-2\log_e\Theta \sim \chi^2_{(1)}$ ดังนั้น ถ้าเราต้องการ 95 เปอร์เซนต์ช่วงความเชื่อมั่นสำหรับ λ เราจะพิจารณาทุกค่าของ λ ซึ่ง $-2\log_e\Theta$ มีค่าน้อยกว่า 3.84 เพราะว่า $Pr\ (-2\log_e\Theta < 3.84) = 0.95$ จะได้ว่า

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

เมื่อเราประเมินค่า $\hat{\sigma}^2_{\varepsilon|\lambda}$ สำหรับ λ ที่แตกต่างกันในการหา ML ของตัวประมาณ $\hat{\lambda}$ แล้วเราจะสามารถหาพิสัย (Range) ของ λ สำหรับช่วงความเชื่อมั่น 95 เปอร์เซนต์นี้ได้

ในกรณีพิเศษเราสามารถแปลงค่าตัวแปรอิสระ ดังนี้

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

และในบางกรณีเราสามารถแปลงค่าตัวแปรตาม y และตัวแปรอิสระ X โดยใช้ตัวเดียวกันบน y และตัวแปรอิสระทุกตัว

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

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

ถ้าเรากำหนด $\lambda = 1$ และ $\lambda = 0$ เราจะได้รูปแบบของสมการเชิงเส้นและสมการลอกการิทึมดังนี้

$$y_i = \beta_o + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i$$

$$\therefore \quad \text{Log}_e y_i = \alpha_o + \alpha_1 \log_e X_{i1} + \alpha_2 \log_e X_{i2} + \dots$$

$$+\alpha_k \log_e X_{ik} + \varepsilon_i$$

Raymond J. Carroll (ค.ศ.1978) ได้ศึกษาถึงวิธีการที่แกร่งสำหรับการทดสอบการแปลงข้อมูลที่ลู่เข้าสู่ภาวะปกติโดยประมาณ ดังนั้น ค่าผิดพลาดอาจจะเข้าใกล้ภาวะปกติเท่านั้น Hinkley (ค.ศ.1975) ได้กล่าวว่า มันเป็นข้อบกพร่องที่ไม่พิจารณาความเป็นไปได้ถึงเรื่องนี้ Carroll จึงพิจารณาวิธีการใหม่โดยอาศัยความรู้พื้นฐานการแปลงข้อมูลของ Box และ Cox กับวิธีการประมาณที่แกร่ง กล่าวคือ จะแทนทฤษฎีในภาวะน่าจะเป็นเดิมเมื่อการแจกแจงเป็นแบบปกติด้วยความหนาแน่นเป็นแบบปกติที่มีศูนย์กลางและแบบเอ็กซ์โปเนนเชียลส่วนหาง (Normal central and Exponential tail)

$$L(\underset{\sim}{\beta},\sigma,\lambda) = \sigma^{n} \sum_{i=1}^{n} \exp\left\{-\rho \frac{(y_i^{(\lambda)} - \underset{\sim}{X}_i'\underset{\sim}{\beta})}{\sigma} + (\lambda-1)\log_e y_i\right\}$$

สำหรับ λ ที่คงที่ เราจะกระทำเช่นเดียวกับปัญหาเกี่ยวกับตำแหน่ง โดยการหาค่าเริ่มต้นของค่า σ ซึ่งทำสมการข้างล่างมีค่ามากที่สุด

$$\sum_{i=1}^{n} \psi\{(y_i^{(\lambda)} - \underset{\sim}{X}_i\underset{\sim}{\beta})/\sigma\}\, x_i = 0$$

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

2.5 ผลงานวิจัยที่เกี่ยวข้อง

ผู้ศึกษาเกี่ยวกับเรื่องวิธีการที่มีความแกร่งสำหรับการถดถอยเชิงเส้นพหุนั้น มีไม่มากนัก ดังนั้น ผลงานวิจัยที่เกี่ยวข้องจึงมีอยู่น้อย แต่ก็ยังมีนักวิจัยบางท่านที่ได้ศึกษาเกี่ยวกับวิธีต่าง ๆ ที่ใช้ในการประมาณสัมประสิทธิ์การถดถอยเชิงเส้น ในหัวข้อนี้จะเสนอวิธีที่น่าสนใจ พร้อมทั้งข้อสรุปต่างๆ ดังนี้คือ

ก) การศึกษาตัวประมาณที่แกร่งของสัมประสิทธิ์การถดถอยเชิงเส้นตรง โดยวิธีหาค่าน้อยที่สุดของส่วนเบี่ยงเบนยกกำลังที่ P (P^{th} Power Deviation หรือ Lp-norm regression) ศึกษาโดย Forsythe (ค.ศ.1972 : 159-166) โดยพารามิเตอร์ของตัวแบบ

ถูกเลือกเพื่อหาค่าน้อยที่สุดของ $\sum_{i=1}^{n} |\varepsilon_i|^p$; $(1 < p < 2)$ สำหรับการคำนวณหาสัมประสิทธิ์การถดถอยพหุเมื่อ $1<p<2$ จะทำเป็นโปรแกรมนอนลิเนียร์ (nonlinear) Forsythe ได้ศึกษากระบวนการนี้สำหรับตัวแบบการถดถอยเชิงเส้นตรงด้วยการจำลองแบบมอนติคาร์โล โดยใช้การแจกแจงแบบปกติปลอมปนหลายแบบ เขาสังเกตว่า $p = 1.5$ เป็นข้อเลือกที่ผสมผสานดีในการที่จะนำมาใช้แทนวิธีกำลังสองน้อยที่สุดเมื่อค่าผิดพลาดไม่เป็นแบบปกติ และเมื่อการแจกแจงของค่าผิดพลาดเป็นแบบปกติ การใช้ $p = 1.5$ ทำให้ผลลัพธ์ในการประมาณจะมีประสิทธิภาพ 90 เปอร์เซนต์เทียบกับวิธีกำลังสองน้อยที่สุด

ข) การศึกษาวิธีการที่แกร่งสำหรับการถดถอยเชิงเส้น ของ Andrew (ค.ศ. 1974:523-615) ซึ่งมีฟังก์ชัน ψ เป็นแบบ sine โดยพิจารณาตัวแบบการถดถอยเชิงเส้น

$$y_i = \underset{\sim}{X}_i \underset{\sim}{\beta} + \sigma\varepsilon_i \quad , \quad i = 1,2,\ldots, n$$

เมื่อ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยเชิงเส้นที่ไม่ทราบค่า X เป็นเวกเตอร์ของแถวตัวแปรอิสระ σ เป็นสเกลพารามิเตอร์ที่ไม่ทราบค่า และ ε_i เป็นค่าผิดพลาด ซึ่งการประมาณสัมประสิทธิ์การถดถอยเชิงเส้น เป็นการหาค่ามากที่สุดของ

$$\sum_{i=1}^{n} \psi \{r_i(\underset{\sim}{\beta})/s(\underset{\sim}{\beta})\}$$

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

โดยที่ฟังก์ชัน ψ อยู่ในรูปของ

$$\psi(Z) = \begin{cases} C|1 + \cos(\frac{Z}{C})| & ; \quad |Z| < C\pi \\ 0 & ; \quad |Z| > C\pi \end{cases}$$

Daniel และ Wood (ค.ศ.1971) ได้ทำการทดลองโดยนำข้อมูลของ Brownee (ค.ศ.1985.sec.13.12) เมื่อพิจารณาจากตัวอย่าง 21 ค่า และตัวแปรอิสระ 3 ตัว ซึ่งมีค่าผิดพลาดที่ใหญ่ผิดปกติ ในค่าสังเกตที่ 1, 3, 4 และ 21 หาสมการการประมาณได้ 4 สมการ โดยที่สมการที่ 1 ใช้วิธีกำลังสองน้อยที่สุด สมการที่ 2 ใช้วิธีกำลังสองน้อยที่สุด แต่ตัดค่าผิดปกติออกไป 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) การศึกษาเปรียบเทียบตัวประมาณที่แกร่งหลายวิธีของความชัน, ค่าจุดตัดแกน y และสเกลของการถดถอยเชิงเส้น ของ Ramsay (ค.ศ.1977: 608-615) Ramsay ได้ศึกษาเปรียบเทียบวิธีการที่แกร่งของการถดถอยเชิงเส้นอย่างง่าย โดยศึกษาเปรียบเทียบเกณฑ์ของ Ramsay (ค.ศ.1977), Andrew (ค.ศ.1972), Forsythe (ค.ศ.1972) และ Hampel (ค.ศ.1971, 1974) และวิธีกำลังสองน้อยที่สุด Lp = 2 Ramsay ทำการทดลองโดยใช้การแจกแจงของค่าผิดพลาดเป็นแบบปกติปลอมปน ซึ่งมี $F = (1-p)N(0,1) + pN(0,c^2)$ เมื่อ p เป็นสัดส่วนของการปลอมปนโดยใช้ 0.0, 0.01, 0.05, 0.10 และ 0.25 และสเกลแฟคเตอร์ c = 3 และ 10 ขนาดตัวอย่าง n = 5, 20 และ 50 โดยตัวแปรอิสระมีการแจกแจงโดยใช้ Quatiles ที่เหมาะสมของแบบปกติและแบบสม่ำเสมอ ผลการทดลองปรากฏว่าเมื่อ c = 3 และ 10 เกณฑ์ของ Ramsay, Andrew และ Hampel จะดีเสมอกัน และยังมีค่าความแปรปรวนน้อยกว่าเกณฑ์ของ Forsythe และ Lp = 2 และเมื่อเปรียบเทียบความสามารถในการประมาณความชันโดยใช้ relative deficiency (1-minimum variance/variance) ปรากฏว่าที่ c = 3 เกณฑ์ของ Ramsay ดีที่สุด และที่ c = 10 เกณฑ์ของ Andrew ดีที่สุด แต่ใกล้เคียงกันกับของ Ramsay