บทที่ 2

สถิติที่ใช้ในการวิจัย

การวิจัยครั้งนี้ ทำการศึกษาถึงวิธีการประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ โดยใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS Method) วิธีตัวประมาณ M (M-Estimator Method) วิธีตัวประมาณ Bounded-Influence (Bounded-Influence Estimator Method : BI Estimator Method) ซึ่งวิธีตัวประมาณ M และวิธีตัวประมาณ BI จะใช้เกณฑ์ความแกร่งของ Huber และของ Tukey และในการคำนวณจะอาศัยวิธีกำลังสองน้อยที่สุดแบบทั่วไป (Generalized Least Squares Method : GLS Method) และวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted Least Squares Method : WLS Method) ซึ่งรายละเอียดของทุกวิธีการมีดังต่อไปนี้

วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS Method)

การประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด เป็นวิธีที่มีรากฐานมาจากทฤษฎีการประมาณค่าเชิงเส้น (Theory of Linear Estimation) โดยมีหลักเกณฑ์คือ หาค่าประมาณสัมประสิทธิ์การถดถอยที่ทำให้ผลบวกกำลังสองของความแตกต่างระหว่างค่าจริงกับค่าประมาณ มีค่าน้อยที่สุด

ในกรณีต้องการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด จากสมการความสัมพันธ์ระหว่างตัวแปรตาม $\mathbf{y}$ และตัวแปรอิสระ $\mathbf{x}$ คือ

$$\mathbf{y} = \mathbf{x}\beta + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times(p+1)} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

เมื่อ $\mathbf{y}$ คือ เวกเตอร์ของตัวแปรตามขนาด nx1

$\mathbf{x}$ คือ เมตริกซ์ของตัวแปรอิสระขนาด nx(p+1) และแรงค์เต็ม (Full Rank)

$\boldsymbol{\beta}$ คือ เวกเตอร์ของพารามิเตอร์หรือสัมประสิทธิ์การถดถอยที่ไม่ทราบค่าขนาด (p+1)x1

$\boldsymbol{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนขนาด nx1

n คือ จำนวนค่าสังเกตทั้งหมด

p คือ จำนวนของตัวแปรอิสระ

โดยที่ ε_i เป็นตัวแปรสุ่มที่เป็นอิสระกัน มีการแจกแจงแบบสมมาตรที่มีค่าเฉลี่ยเท่ากับ 0 ($E(\boldsymbol{\varepsilon}) = \underline{0}$) และความแปรปรวนคงที่ ($V(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$)

จากหลักการของวิธี OLS จะหาตัวประมาณของ $\boldsymbol{\beta}$ ($\hat{\boldsymbol{\beta}}$) ที่ทำให้ผลบวกกำลังสองของความแตกต่างระหว่างค่าจริงกับค่าประมาณหรืออาจเรียกว่าผลบวกกำลังสองของความคลาดเคลื่อน (Sum of Square of Errors : SSE) มีค่าน้อยที่สุดซึ่งก็คือ

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \quad \text{หรือ} \quad \min_{\beta} \sum_{i=1}^{n} (\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}) \tag{1}$$

เมื่อ

$$\begin{aligned} \text{SSE} &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ &= (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{x}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{x}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{x}^T\mathbf{x}\boldsymbol{\beta} \\ &= \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{x}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{x}^T\mathbf{x}\boldsymbol{\beta} \end{aligned}$$

และการแก้สมการ (1) เพื่อหา $\hat{\boldsymbol{\beta}}$ ทำได้โดยการหาอนุพันธ์ (Differentiate) ของสมการ (1) เทียบกับ $\hat{\boldsymbol{\beta}}$ แล้วกำหนดให้เท่ากับศูนย์

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}} (\mathbf{y}^T\mathbf{y} - 2\hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{y} + \hat{\boldsymbol{\beta}}^T\mathbf{x}^T\mathbf{x}\hat{\boldsymbol{\beta}}) = \underline{0}$$

$$-2\mathbf{x}^T\mathbf{y} + 2\mathbf{x}^T\mathbf{x}\hat{\boldsymbol{\beta}} = \underline{0}$$

$$(\mathbf{x}^T\mathbf{x})\hat{\boldsymbol{\beta}} = \mathbf{x}^T\mathbf{y}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}$$

เมื่อ $\hat{\beta}$ คือ เวกเตอร์ของตัวประมาณค่าพารามิเตอร์หรือค่าสัมประสิทธิ์การถดถอย

โดยที่ $E(\hat{\beta}) = \beta$

และเมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ คือ

$$\begin{aligned} Cov(\hat{\beta}) &= Cov((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Cov(\mathbf{y})\,\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \end{aligned}$$

และตัวประมาณของวิธี OLS จะมีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimatior : BLUE)

วิธีกำลังสองน้อยที่สุดแบบทั่วไป (Generalized Least Squares Method : GLS Method) **และวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก** (Weighted Least Squares Method : WLS Method)

วิธีกำลังสองน้อยที่สุดแบบทั่วไป เป็นวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีที่ ε ไม่เป็นไปตามข้อตกลงเบื้องต้นของวิธีกำลังสองน้อยที่สุด คือ

เหตุการณ์ที่ ε_i มีความสัมพันธ์กัน ($E(\varepsilon_i\varepsilon_j) \neq 0$, $i \neq j$) และ/หรือ ε_i มีความแปรปรวนไม่เท่ากัน ($V(\varepsilon_i) \neq \sigma^2$)

นั่นคือ $V(\varepsilon) \neq \sigma^2 I_n$

เมื่อ I_n คือ เมตริกซ์เอกลักษณ์ (Identity Matrix)

หรือ $V(\varepsilon) = \sigma^2 \Omega$

เมื่อ Ω คือ เมตริกซ์แบบนอนซิงกูลาร์ (Nonsingular Matrix) และเป็นเมตริกซ์ที่เป็นบวกแน่นอน (Positive Definite Matrix)

ดังนั้นการแก้ไขจะใช้การแปลงข้อมูล (Transform Data) เพื่อทำให้ ε_i มีคุณสมบัติเป็นไปตามข้อตกลงของวิธีกำลังสองน้อยที่สุด คือ $E(\varepsilon) = \underline{0}$ และ $V(\varepsilon) = \sigma^2 I_n$ โดยใช้วิธีดังนี้

ถ้ากำหนดให้ Ω เป็นเมตริกซ์สมมาตรที่เป็นบวกแน่นอน (Symmetric Positive Definite Matrix) ขนาด nxn แล้วเราสามารถหาเมตริกซ์สมมาตรซึ่งมิใช่เอกฐาน P ที่ทำให้

$$P^T P = P P = \Omega$$

และทำการแปลงข้อมูลโดย

$$\mathbf{y}^* = P^{-1}\mathbf{y}$$
$$\mathbf{x}^* = P^{-1}\mathbf{x}$$
$$\varepsilon^* = P^{-1}\varepsilon$$

แล้วสามารถเขียนสมการถดถอยได้คือ

$$\mathbf{y}^* = \mathbf{x}^*\beta + \varepsilon^*$$

โดยที่ $E(\varepsilon^*) = P^{-1}E(\varepsilon) = \underline{0}$

และ $V(\varepsilon^*) = E(\varepsilon^*\varepsilon^{*T})$

$$\begin{aligned} E(\varepsilon^*\varepsilon^{*T}) &= E(P^{-1}\varepsilon\,\varepsilon^T P^{-1}) \\ &= P^{-1}E(\varepsilon\,\varepsilon^T)P^{-1} \\ &= \sigma^2 P^{-1}\Omega P^{-1} \\ &= \sigma^2 P^{-1}P P P^{-1} \\ &= \sigma^2 I_n \end{aligned}$$

นั่นคือ $E(\varepsilon^*) = \underline{0}$ และ $E(\varepsilon^*\varepsilon^{*T}) = \sigma^2 I_n$ ซึ่งสอดคล้องกับข้อตกลงเบื้องต้นของวิธี OLS ดังนั้นการประมาณค่า β จะหาจากการ

$$\min_{\beta}\sum_{i=1}^{n}(\varepsilon^{*T}\varepsilon^*) = \min_{\beta}\sum_{i=1}^{n}(\varepsilon^T\Omega^{-1}\varepsilon)$$

แล้วตัวประมาณ ($\hat{\beta}$) อยู่ในรูปของ

$$\begin{aligned}\hat{\beta} &= (\mathbf{x}^{*T}\mathbf{x}^{*})^{-1}\ \mathbf{x}^{*T}\mathbf{y}^{*} \\ &= \left\{(P^{-1}\mathbf{x})^T(P^{-1}\mathbf{x})\right\}^{-1}\left\{(P^{-1}\mathbf{x})^T(P^{-1}\mathbf{y})\right\} \\ &= \left\{\mathbf{x}^T(P^T)^{-1}P^{-1}\mathbf{x}\right\}^{-1}\left\{\mathbf{x}^T(P^T)^{-1}P^{-1}\mathbf{y}\right\} \\ &= (\mathbf{x}^T\Omega^{-1}\mathbf{x})^{-1}\,\mathbf{x}^T\Omega^{-1}\mathbf{y}\end{aligned}$$

ซึ่ง $\hat{\beta}$ จะเรียกว่า ตัวประมาณแบบ GLS

และเมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ คือ

$$\begin{aligned}\mathrm{Cov}(\hat{\beta}) &= \sigma^2(\mathbf{x}^{*T}\mathbf{x}^{*})^{-1} \\ &= \sigma^2\left\{(P^{-1}\mathbf{x})^T(P^{-1}\mathbf{x})\right\}^{-1} \\ &= \sigma^2\left\{\mathbf{x}^T(P^T)^{-1}P^{-1}\mathbf{x}\right\}^{-1} \\ &= \sigma^2(\mathbf{x}^T\Omega^{-1}\mathbf{x})^{-1}\end{aligned}$$

และในกรณีที่ ε ไม่มีความสัมพันธ์กันแต่มีความแปรปรวนไม่เท่ากันและเมตริกซ์ความแปรปรวนร่วมของ ε เขียนได้โดย

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

เมื่อ $\mathbf{W} = \Omega^{-1}$ และ Ω เป็นเมตริกซ์ทแยงมุม (Diagonal Matrix) จะได้ว่า $\mathbf{W}$ เป็นเมตริกซ์ทแยงมุมโดยมีสมาชิกแนวทแยงมุม คือ $w_1, w_2, \ldots, w_n$ แล้ว จะเรียกวิธีนี้ว่า วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weight Least Square Method : WLS Method) ซึ่งเป็นกรณีเฉพาะของวิธี GLS และสมการปกติของวิธี WLS อยู่ในรูปของ

$$(\mathbf{x}^T\mathbf{W}\mathbf{x})\hat{\beta} = \mathbf{x}^T\mathbf{W}\mathbf{y}$$

แล้วตัวประมาณ ($\hat{\beta}$) อยู่ในรูปของ

$$\hat{\beta} = (\mathbf{x}^T\mathbf{W}\mathbf{x})^{-1}\,\mathbf{x}^T\mathbf{W}\mathbf{y}$$

และเรียก w_i ว่า น้ำหนัก (Weight) สำหรับค่าสังเกตที่ i , $i = 1, 2, \ldots, n$
สำหรับเมตริกซ์ความแปรปรวนร่วม $\hat{\beta}$ คือ

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2(\mathbf{x}^T\mathbf{W}\mathbf{x})^{-1}$$

และพบว่า ค่าสังเกตที่มีความแปรปรวนมากจะถูกให้น้ำหนักที่น้อยกว่าค่าสังเกตที่มีความแปรปรวนน้อยซึ่งหลักการให้น้ำหนักสำหรับแต่ละค่าสังเกตแบบนี้จะถูกนำไปใช้ในการหาค่า $\hat{\beta}$ เมื่อข้อมูลมีค่าผิดปกติ โดยจะกล่าวถึงในวิธีตัวประมาณถัดไป

วิธีตัวประมาณ M (M-Estimator Method)

Huber (1964) ได้ศึกษาตัวประมาณที่มีความแกร่ง (Robust) และเรียกตัวประมาณนี้ว่า ตัวประมาณ M ซึ่งวิธีตัวประมาณ M จะถูกพัฒนามาจากหลักการพื้นฐานของการประมาณแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation)

ในกรณีต้องการประมาณค่าสัมประสิทธิ์การถดถอย จากหลักการของวิธีภาวะ-น่าจะเป็นสูงสุดจะต้องทราบลักษณะการแจกแจงหรือรูปแบบของฟังก์ชันความหนาแน่น (Density Function) ของความคลาดเคลื่อน คือ $f(\varepsilon_i)$

เมื่อ $\varepsilon_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$

และ

$$\mathbf{x}_i^T = (1 \;\; x_{1i} \;\; x_{2i} \;\; \ldots \;\; x_{pi})$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)x1}$$

ε_i คือ ความคลาดเคลื่อนของค่าสังเกตที่ i

y_i คือ ตัวแปรตามของค่าสังเกตที่ i

$\mathbf{x}_i^T$ คือ เวกเตอร์ของตัวแปรอิสระ p ตัว ของค่าสังเกตที่ i

$\boldsymbol{\beta}$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอย

n คือ จำนวนตัวอย่าง

แล้วจะหาตัวประมาณของ $\boldsymbol{\beta}$ ($\hat{\boldsymbol{\beta}}$) ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) ของความคลาดเคลื่อนมีค่าสูงสุด คือ

$$\max \sum_{i=1}^{n} \ln f(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \quad = \quad \min -\sum_{i=1}^{n} \ln f(y_i - \mathbf{x}_i^T \boldsymbol{\beta})$$

เมื่อ $\sum_{i=1}^{n} \ln f(y_i - \mathbf{x}_i^T \boldsymbol{\beta})$ เป็นฟังก์ชันภาวะน่าจะเป็น

และในการหา $\hat{\boldsymbol{\beta}}$ ทำได้โดยการหาอนุพันธ์ของ $\sum_{i=1}^{n} \ln f(y_i - \mathbf{x}_i^T \boldsymbol{\beta})$ เทียบกับ $\hat{\boldsymbol{\beta}}$ แล้วกำหนดให้ฟังก์ชันมีค่าเท่ากับศูนย์ดังนี้

$$\sum_{i=1}^{n} \mathbf{x}_i \frac{f'(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{f(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})} \quad = \quad \underline{0}$$

เมื่อ $f'(z) = \dfrac{df(z)}{dz}$

และถ้า ε_i มีการแจกแจงแบบปกติแล้ว $\hat{\beta}$ ซึ่งเป็นตัวประมาณแบบภาวะน่าจะเป็นสูงสุด จะมีรูปแบบเดียวกับตัวประมาณจากวิธี OLS แต่ถ้าไม่ทราบรูปแบบฟังก์ชันการแจกแจงของความคลาดเคลื่อน ($f(\varepsilon_i)$) แล้วทำให้ไม่สามารถคำนวณ f'/f ได้และไม่สามารถหา $\hat{\beta}$ ได้ ดังนั้น Huber ได้เสนอวิธีตัวประมาณ M ซึ่งวิธีนี้จะมีการกำหนดฟังก์ชัน ψ แทน f'/f และการกำหนดฟังก์ชัน ψ ที่เหมาะสม (คือฟังก์ชัน ψ เป็นตัวแทนที่ดีในการอธิบายลักษณะการแจกแจงของε_i) จะมีผลต่อความแกร่งและการมีประสิทธิภาพของตัวประมาณ M

ดังนั้นหลักการของตัวประมาณ M ก็คือ จะหาตัวประมาณของ β ที่ทำให้ผลรวมของฟังก์ชันความผิดพลาด (ρ) มีค่าน้อยที่สุดดังนี้

$$\min_{\beta}\sum_{i=1}^{n}\rho\left(\frac{y_i-\mathbf{x}_i^T\beta}{\sigma}\right) \tag{2}$$

โดยที่ ρ เป็น ฟังก์ชันของความผิดพลาดที่ถูกเลือกให้เหมาะสม และถ้า $\rho(u)=\frac{1}{2}u^2$ แล้ว $\hat{\beta}$ ก็คือ ตัวประมาณจากวิธี OLS และถ้า $\rho(u)=-\ln f(u)$ แล้ว $\hat{\beta}$ ก็คือ ตัวประมาณภาวะน่าจะเป็นสูงสุด

และ σ เป็น ค่าพารามิเตอร์ของสเกล (Scale Parameter) จากการแจกแจงของความคลาดเคลื่อน

ในการแก้สมการ (2) เพื่อหา $\hat{\beta}$ ที่เหมาะสมทำได้โดยการหาอนุพันธ์ของ

$$\sum_{i=1}^{n}\rho\left(\frac{y_i-\mathbf{x}_i^T\beta}{\sigma}\right)$$

เทียบกับ $\hat{\beta}$ แล้วกำหนดให้เท่ากับศูนย์คือ

$$\sum_{i=1}^{n}\psi\left(\frac{e_i}{\hat{\sigma}}\right)\mathbf{x}_i = \underline{0} \tag{3}$$

เมื่อ $\psi(u) = \frac{\partial}{\partial u}\rho(u)$ และถ้า $\psi(u) = u$ แล้ววิธีตัวประมาณ M ก็คือวิธี OLS

$$e_i = y_i - \mathbf{X}_i^T\hat{\beta}$$

และ $\hat{\sigma}$ เป็นตัวประมาณของสเกล โดยในที่นี้ใช้ตัวประมาณแบบมัธยฐานของค่าสัมบูรณ์ของความเบี่ยงเบน (Median Absolute Deviation : MAD) ซึ่งถูกเสนอโดย Mosteller และ Tukey (1977) และจะถูกปรับด้วยค่าคงที่ 1.4826 ซึ่งจะทำให้ $\hat{\sigma}$ เป็นตัวประมาณที่ไม่เอนเอียงของ σ เมื่อ n มีขนาดใหญ่ และ $\hat{\sigma}$ มีรูปแบบดังนี้

$$\hat{\sigma} = 14826 \ \underset{i}{\mathrm{med}}\left(\left|e_i - \underset{j}{\mathrm{med}}(e_j)\right|\right) \ , i = 1, 2, \ldots, n$$

จากสมการ (3) พบว่า เป็นสมการแบบไม่เป็นเชิงเส้น (Nonlinear) ดังนั้นการแก้สมการจะอาศัยเทคนิคการทำซ้ำเพื่อหาค่าที่เหมาะสม ซึ่งในที่นี้ใช้วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักซ้ำหลายรอบ (The Iteratively Reweighted Least Squares Procedure : IRLS Procedure) ของ Beaton และTukey (1964) ในการคำนวณของวิธี IRLS จะต้องมีการกำหนดค่าเริ่มต้นของ $\hat{\beta}$ และคำนวณ $\hat{\sigma}$ แล้วหาน้ำหนักถ่วงที่เหมาะสมของแต่ละค่าสังเกตเพื่อหาตัวประมาณค่า β จากวิธี WLS ดังนี้

จากสมการ (3)

$$\sum_{i=1}^{n} \psi\left(\frac{e_i}{\hat{\sigma}}\right)\mathbf{X}_i = \underline{0}$$

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{e_i}{\hat{\sigma}}\right) = \frac{\sum_{i=1}^{n} x_{ij}\psi\left(\frac{e_i}{\hat{\sigma}}\right)}{\left(\frac{e_i}{\hat{\sigma}}\right)} \cdot \left(\frac{e_i}{\hat{\sigma}}\right) = 0$$

$$\sum_{i=1}^{n} x_{ij}w_i\left(\frac{e_i}{\hat{\sigma}}\right) = 0$$

เมื่อ $w_i\left(\frac{e_i}{\hat{\sigma}}\right) = \psi\left(\frac{e_i}{\hat{\sigma}}\right) \bigg/ \left(\frac{e_i}{\hat{\sigma}}\right)$, $j = 0, 1, 2, \ldots, p$

และ p คือ จำนวนตัวแปรอิสระ

ดังนั้น ตัวประมาณของ β จากวิธี WLS อยู่ในรูปของ

$$\hat{\beta} = (\mathbf{x}^T\mathbf{W}\mathbf{x})^{-1}\mathbf{x}^T\mathbf{W}\mathbf{y}$$

เมื่อ $\mathbf{W}$ เป็น เมตริกซ์ทแยงมุมขนาด nxn ที่มี w_i เป็นสมาชิกแนวทแยงมุม

โดยขั้นตอนของการคำนวณเมื่อใช้วิธี IRLS มีรายละเอียดดังนี้

ขั้นที่ 1 มีการกำหนดค่าเริ่มต้นของ $\hat{\beta}$ โดยใช้วิธี OLS

$$\hat{\beta}^{(0)} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{y}$$

ขั้นที่ 2 เข้าสู่การทำซ้ำรอบที่ 1 เพื่อหาค่า $e_i^{(1)}$ และ $\hat{\sigma}^{(1)}$ โดยนำ $\hat{\beta}^{(0)}$ มาคำนวณ e_i และ $\hat{\sigma}$ ดังนี้

ขั้นที่ 2.1

$$e_i^{(1)} = y_i - \mathbf{x}_i^T\hat{\beta}^{(0)} \quad , i = 1, 2, \ldots , n$$

ขั้นที่ 2.2

$$\hat{\sigma}^{(1)} = 1.4826 \ \underset{i}{\text{med}} \left(\left| e_i^{(1)} - \underset{j}{\text{med}}(e_j^{(1)}) \right| \right)$$

ขั้นที่ 3 การทำซ้ำรอบที่ 1 เพื่อหาน้ำหนักของแต่ละค่าสังเกต ตามเงื่อนไขของเกณฑ์ความแกร่งหรือรูปแบบฟังก์ชัน ψ ที่ถูกเลือก

$$w_i^{(1)} = \begin{cases} \dfrac{\psi(e_i^{(1)}/\hat{\sigma}^{(1)})}{e_i^{(1)}/\hat{\sigma}^{(1)}} & , e_i^{(1)} \neq 0 \\ 1 & , e_i^{(1)} = 0 \end{cases}$$

เมื่อ $w_i^{(1)}$ คือ น้ำหนักของค่าสังเกตที่ i ในการทำซ้ำรอบที่ 1

ขั้นที่ 4 การทำซ้ำรอบที่ 1 เพื่อหา $\hat{\beta}^{(1)}$

$$\hat{\beta}^{(1)} = (\mathbf{x}^{T}\mathbf{W}^{(1)}\mathbf{x})^{-1}\mathbf{x}^{T}\mathbf{W}^{(1)}\mathbf{y}$$

เมื่อ $\mathbf{W}^{(1)}$ คือ เมตริกซ์ทแยงมุมของน้ำหนักในรอบที่ 1 ขนาด nxn

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

ขั้นที่ 5 หาค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณของสัมประสิทธิ์การถดถอยเริ่มต้นกับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ 1 ของสัมประสิทธิ์การถดถอยทุกค่า

ขั้นที่ 6 ถ้าค่าสัมบูรณ์ของขั้นที่ 5 อย่างน้อยหนึ่งค่า มากกว่า 0.001 ให้ทำขั้นต่อไป แต่ถ้าค่าสัมบูรณ์จากขั้นที่ 5 ของสัมประสิทธิ์การถดถอยทุกค่าไม่มากกว่า 0.001 ก็จะได้ค่าประมาณสัมประสิทธิ์การถดถอยจากรอบที่ 1

ขั้นที่ 7 เข้าสู่ขั้นที่ 2 โดยเป็นการทำซ้ำรอบที่ k , k = 2, 3, ...

ขั้นที่ 7.1

$$e_i^{(k)} = y_i - \mathbf{x}_i^{T}\hat{\beta}^{(k-1)} \quad , i = 1, 2, \ldots , n$$

ขั้นที่ 7.2

$$\hat{\sigma}^{(k)} = 1.4826 \ \underset{i}{\text{med}} \left(\left| e_i^{(k)} - \underset{j}{\text{med}}(e_j^{(k)}) \right| \right)$$

ขั้นที่ 8 เข้าสู่ขั้นที่ 3 ในการทำซ้ำรอบที่ k

$$w_i^{(k)} = \frac{\psi(e_i^{(k)}/\hat{\sigma}^{(k)})}{e_i^{(k)}/\hat{\sigma}^{(k)}}$$

ขั้นที่ 9 เข้าสู่ขั้นการประมาณค่าสัมประสิทธิ์การถดถอย ในการทำซ้ำรอบที่ k

$$\hat{\beta}^{(k)} = (\mathbf{X}^T\mathbf{W}^{(k)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(k)}\mathbf{y}$$

ขั้นที่ 10 หาค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณของสัมประสิทธิ์การถดถอยรอบที่ k-1 กับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k ของสัมประสิทธิ์การถดถอยทุกค่า

ขั้นที่ 11 ถ้าค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณของสัมประสิทธิ์การถดถอยรอบที่ k-1 กับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k ของสัมประสิทธิ์การถดถอยทุกค่า มีอย่างน้อยหนึ่งค่า มากกว่า 0.001 ให้กลับไปทำซ้ำขั้นที่ 7 จนถึงขั้นที่ 10 เรื่อย ๆ จนกระทั่งค่าสัมบูรณ์ของผลต่างระหว่างค่าประมาณของสัมประสิทธิ์การถดถอยรอบที่ k-1 กับค่าประมาณสัมประสิทธิ์การถดถอยรอบที่ k ทุกค่า ไม่มากกว่า 0.001 แล้ว จะได้ค่าประมาณสัมประสิทธิ์การถดถอยจากรอบที่ k

สำหรับเมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ Huber (1973) ได้ศึกษาพบว่า $\hat{\beta}$ มีคุณสมบัติ ลู่เข้าสู่การแจกแจงแบบปกติเมื่อใกล้อนันต์ (Asyemtotic Normal) โดย

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 \frac{E[\psi^2(\varepsilon/\sigma)]}{\{E[\psi'(\varepsilon/\sigma)]\}^2}(\mathbf{X}^T\mathbf{X})^{-1}$$

และเมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ โดยประมาณ คือ

$$\hat{C}ov(\hat{\beta}) = \frac{(n\hat{\sigma})^2}{n-(p+1)} \frac{\sum_{i=1}^{n}\psi^2\left[(y_i - \mathbf{x}_i^T\hat{\beta})/\hat{\sigma}\right]}{\left\{\sum_{i=1}^{n}\psi'\left[(y_i - \mathbf{x}_i^T\hat{\beta})/\hat{\sigma}\right]\right\}^2}(\mathbf{X}^T\mathbf{X})^{-1}$$

และเมื่อใช้หลักการคำนวณของวิธี WLS ได้

$$\hat{C}ov(\hat{\beta}) = \frac{\sum_{i=1}^{n} w_i \cdot (y_i - \mathbf{x}_i^T\hat{\beta})^2}{n-(p+1)}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$$

จากสมการ (2) และ (3) จะต้องมีการเลือกฟังก์ชัน ψ หรืออาจเรียกว่าเกณฑ์ความแกร่งที่เหมาะสม เพื่อทำให้ได้สัมประสิทธิ์การถดถอยที่มีความแกร่งและมีประสิทธิภาพ ซึ่งในที่นี้ใช้เกณฑ์ความแกร่ง 2 เกณฑ์ คือ

1. เกณฑ์ความแกร่งของ Huber (1964) ซึ่งมีรูปแบบของฟังก์ชัน ρ ฟังก์ชัน ψ และฟังก์ชัน w ดังนี้

$$\rho(u) = \begin{cases} \frac{1}{2}u^2 & , |u| \le c_H \\ c_H|u| - \frac{1}{2}c_H^2 & , |u| > c_H \end{cases}$$

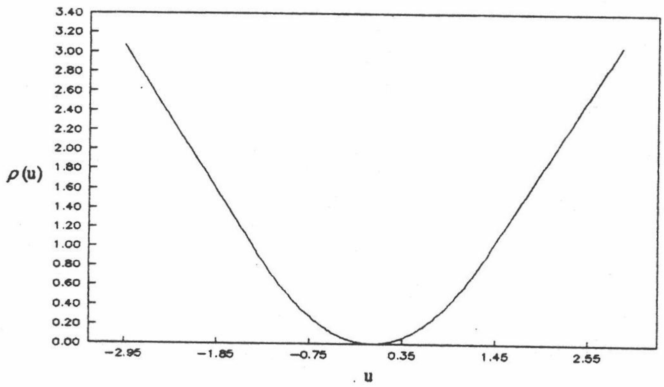
เมื่อ $\psi(u) = \frac{\partial}{\partial u}\rho(u)$ แล้ว

$$\psi(u) = \begin{cases} -c_H & , u < -c_H \\ u & , -c_H \le u \le c_H \\ c_H & , u > c_H \end{cases}$$

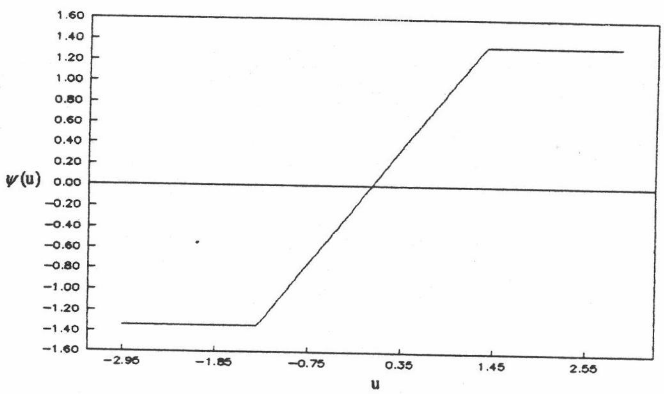
และ $\mathbf{w(u)} = \dfrac{\psi(\mathbf{u})}{\mathbf{u}}$ แล้ว

$$\mathbf{w(u)} = \begin{cases} \dfrac{-c_H}{u} & , \quad u < -c_H \\ 1 & , \quad -c_H \le u \le c_H \\ \dfrac{c_H}{u} & , \quad u > c_H \end{cases}$$

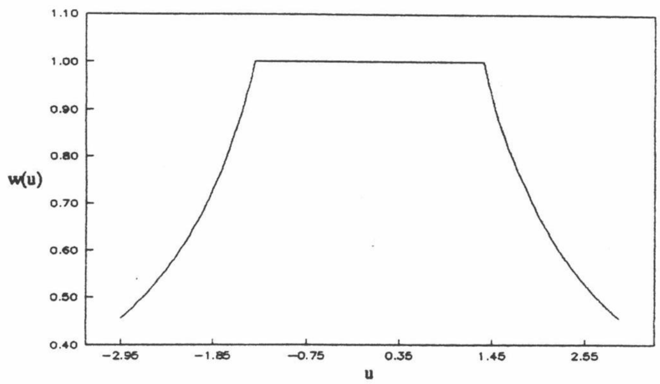
ฟังก์ชัน ρ มีลักษณะเป็นแบบ convex



ฟังก์ชัน ψ มีลักษณะเป็นแบบ monotone



ฟังก์ชัน w



เมื่อ u คือ ความคลาดเคลื่อนมาตรฐาน $(e_i/\hat{\sigma})$

และ c_H คือ ค่าคงที่ (Tuning Constant) ซึ่งกำหนดมาเพื่อให้ตัวประมาณที่ได้มีประสิทธิภาพตามที่ต้องการ เมื่อเทียบกับตัวประมาณจากวิธี OLS

ในที่นี้ c_H = 1.345 มีความหมายว่า เมื่อความคลาดเคลื่อนมาตรฐานมีการแจกแจงแบบปกติแล้ว ตัวประมาณจากวิธีตัวประมาณ M ที่ใช้เกณฑ์ความแกร่งของ Huber จะมีประสิทธิภาพ 95% เมื่อเทียบกับตัวประมาณจากวิธี OLS

และจากลักษณะของฟังก์ชัน ρ ฟังก์ชัน ψ และฟังก์ชัน w พบว่า มีการกำหนดขอบเขตของความคลาดเคลื่อนมาตรฐาน โดยถือว่าความคลาดเคลื่อนมาตรฐานที่อยู่ภายในขอบเขต จะมีการแจกแจงแบบปกติมาตรฐานจึงมีการให้น้ำหนักเท่ากับ 1 แก่ค่าสังเกตเหล่านี้ ส่วนความคลาดเคลื่อนมาตรฐานที่อยู่นอกขอบเขตจะถูกลดอิทธิพลลงโดยมีการให้น้ำหนักที่น้อยกว่า 1 ดังนั้นเกณฑ์ความแกร่งของ Huber เหมาะสำหรับข้อมูลที่ความคลาดเคลื่อนมาตรฐานมีค่าผิดปกติที่ไม่รุนแรงมากนัก

2. เกณฑ์ความแกร่งของ Tukey (1974) ซึ่งมีรูปแบบของฟังก์ชัน ρ ฟังก์ชัน ψ และฟังก์ชัน w ดังนี้

$$\rho(u) = \begin{cases} \dfrac{u^2}{2} - \dfrac{u^4}{2c_B^2} + \dfrac{u^6}{6c_B^4} & , |u| \le c_B \\ \dfrac{c_B^2}{6} & , |u| > c_B \end{cases}$$

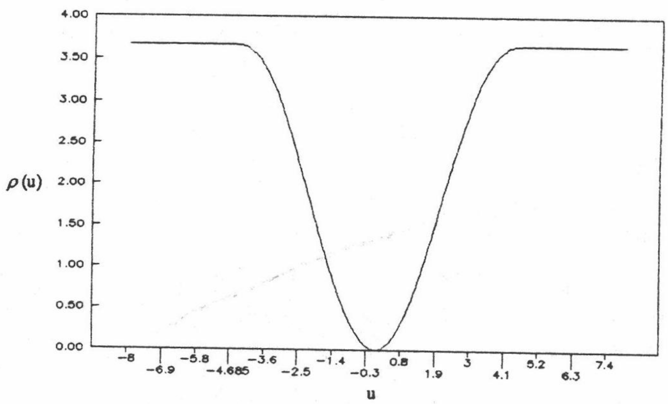
เมื่อ $\psi(u) = \frac{\partial}{\partial u}\rho(u)$ แล้ว

$$
\psi(u) = \begin{cases} 0 & , \quad u < -c_B \\ u\left[1 - \left(\dfrac{u}{c_B}\right)^2\right]^2 & , \quad -c_B \le u \le c_B \\ 0 & , \quad u > c_B \end{cases}
$$

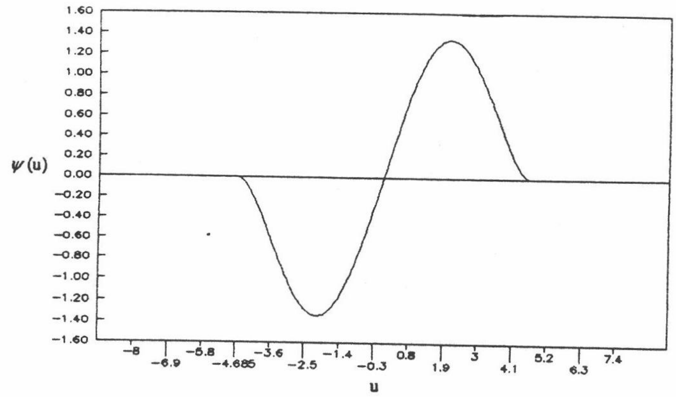
และ $w(u) = \frac{\psi(u)}{u}$ แล้ว

$$
w(u) = \begin{cases} 0 & , \quad u < -c_B \\ \left[1 - \left(\dfrac{u}{c_B}\right)^2\right]^2 & , \quad -c_B \le u \le c_B \\ 0 & , \quad u > c_B \end{cases}
$$

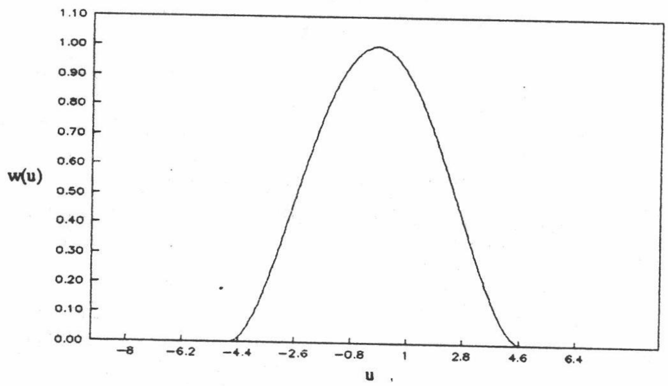
ฟังก์ชัน ρ มีลักษณะเป็นแบบ nonconvex



ฟังก์ชัน ψ มีลักษณะเป็นแบบ redesoending



ฟังก์ชัน w



เมื่อ u คือ ความคลาดเคลื่อนมาตรฐาน $(e_i/\hat{\sigma})$

และ c_B คือ ค่าคงที่ (Tuning Constant) ซึ่งกำหนดมาเพื่อให้ตัวประมาณที่ได้มีประสิทธิภาพตามที่ต้องการเมื่อเทียบกับตัวประมาณจากวิธี OLS

ในที่นี้ c_B = 4.685 มีความหมายว่า เมื่อความคลาดเคลื่อนมาตรฐานมีการแจกแจงแบบปกติแล้ว ตัวประมาณจากวิธีตัวประมาณ M ที่ใช้เกณฑ์ความแกร่งของ Tukey จะมีประสิทธิภาพ 95% เมื่อเทียบกับตัวประมาณจากวิธี OLS

และจากลักษณะของฟังก์ชัน ρ , ฟังก์ชัน ψ และฟังก์ชัน w พบว่าความคลาดเคลื่อนมาตรฐานที่มีค่าสูงมากหรือต่ำมากกว่าขอบเขตที่กำหนด ค่าเหล่านี้จะไม่มี

อิทธิพลคือ มีการให้น้ำหนักเท่ากับ 0 ส่วนความคลาดเคลื่อนมาตรฐานที่อยู่ภายในขอบเขต จะถูกให้น้ำหนักที่มีค่าตั้งแต่ 0 จนถึง 1 ดังนั้นเกณฑ์ความแกร่งของ Tukey เหมาะสำหรับข้อมูลที่มีค่าผิดปกติที่รุนแรงมาก ๆ

วิธีตัวประมาณ Bounded-Influence (Bounded-Influence Estimator Method : BI Estimator Method)

วิธีตัวประมาณ BI เป็น วิธีการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับข้อมูลที่เกิดค่าผิดปกติในตัวแปรตามและตัวแปรอิสระ ณ ตำแหน่งเดียวกัน (Outliers occurring at high leverage points) โดยวิธีนี้จะพัฒนามาจากวิธีตัวประมาณ M คือ วิธีตัวประมาณ BI มีการใช้ฟังก์ชันความผิดพลาดที่พิจารณาความผิดปกติที่เกิดจากความคลาดเคลื่อนและเกิดจากตัวแปรอิสระไปพร้อม ๆ กัน แทนฟังก์ชัน ρ ของวิธีตัวประมาณ M ซึ่งเป็นฟังก์ชันความผิดพลาดที่พิจารณาแต่ความผิดปกติที่ความคลาดเคลื่อนเท่านั้น ดังนั้นฟังก์ชันความผิดพลาดของวิธีตัวประมาณ BI มีรูปแบบ คือ $\tau\,(\mathbf{X}_i, (y_i - \mathbf{X}_i^T\beta)/\sigma)$

และการหาตัวประมาณของสัมประสิทธิ์การถดถอยทำได้โดยการหาตัวประมาณของ β ที่ทำให้ผลรวมของฟังก์ชันความผิดพลาดมีค่าน้อยที่สุด คือ

$$\min_{\beta}\sum_{i=1}^{n}\tau\,(\mathbf{X}_i\,,\,(y_i-\mathbf{X}_i^T\beta\,/\sigma)) \qquad (4)$$

เมื่อ y_i คือ ตัวแปรตามของค่าสังเกตที่ i

$\mathbf{X}_i^T$ คือ เวกเตอร์ ของตัวแปรอิสระ p ตัว ของค่าสังเกตที่ i

$\mathbf{X}_i^T = (1 \quad x_{1i} \quad x_{2i} \quad \ldots \quad x_{pi})$

β คือ เวกเตอร์ของสัมประสิทธิ์การถดถอย

σ คือ ค่าพารามิเตอร์สเกลของการแจกแจงของความคลาดเคลื่อน

จากสมการ (4) เพื่อหา $\hat{\beta}$ ที่เหมาะสมทำได้โดยการหาอนุพันธ์ของ

$$\sum_{i=1}^{n} \tau \left(\mathbf{x}_i \, , \, (y_i - \mathbf{x}_i^T \beta / \sigma) \right)$$

เทียบกับ $\hat{\beta}$ แล้วกำหนดให้เท่ากับศูนย์ คือ

$$\sum_{i=1}^{n} \eta \left(\mathbf{x}_i \, , \, (y_i - \mathbf{x}_i^T \hat{\beta} / \hat{\sigma}) \right) \mathbf{x}_i \quad = \quad \underline{0} \tag{5}$$

และมีนักสถิติหลายท่านทำการศึกษา ความสัมพันธ์ระหว่างฟังก์ชัน η กับฟังก์ชัน ψ โดยเกี่ยวข้องกับฟังก์ชันน้ำหนักของตัวแปรอิสระ และในที่นี้ใช้ฟังก์ชัน η ที่มีรูปแบบความสัมพันธ์กับฟังก์ชัน ψ จากการเสนอของ Sohweppe (1975) ดังนี้

$$\eta \left(\mathbf{X}_i \, , e_i / \hat{\sigma} \right) \quad = \quad v(\mathbf{X}_i) \cdot \psi \left(\frac{e_i}{v(\mathbf{X}_i) \cdot \hat{\sigma}} \right) \tag{6}$$

เมื่อ $e_i \quad = \quad y_i - \mathbf{x}_i^T \hat{\beta}$

$\hat{\sigma}$ คือ ตัวประมาณสเกลของการแจกแจงของความคลาดเคลื่อน ในที่นี้ใช้ตัวประมาณแบบมัธยฐานของค่าสัมบูรณ์ของความเบี่ยงเบน (Median Absolute Deviation : MAD) เช่นเดียวกับวิธีตัวประมาณ M

และ $v(\mathbf{X}_i)$ คือ ฟังก์ชันน้ำหนักของ $\mathbf{X}_i$ ซึ่งจะถูกนำมาให้น้ำหนักแก่ค่า $\mathbf{X}_i$ ที่ผิดปกติ และได้มีผู้คิดรูปแบบ $v(\mathbf{X}_i)$ ที่เหมาะสมหลายแบบและแต่ละรูปแบบจะเป็นฟังก์ชันของสมาชิกแนวทแยงมุมของเมตริกซ์ Hat (Hat Matrix : $\mathbf{H}$) ที่มีขนาด nxn และ สมการของ $\mathbf{H}$ คือ

$$\begin{aligned} \mathbf{H} &= \mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T \\ &= \begin{bmatrix} h_{11} & h_{21} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix}_{n \times n} \end{aligned}$$

เมื่อ h_{ii} คือ สมาชิกแนวทแยงมุมของ $\mathbf{H}$ และจะเรียก h_{ii} ว่า ค่า Leverage (Leverage Value) หรือ ค่า hat (Hat Value) ซึ่งค่า h_{ii} เป็น ค่าที่ใช้วัดความผิดปกติของค่าตัวแปรอิสระของตำแหน่งที่ i (Measure of Leverage) โดยใช้หลักการพิจารณาจากระยะห่างระหว่างค่าของตัวแปรอิสระของตำแหน่งที่ i กับจุดศูนย์กลางของตัวแปรอิสระทั้งหมด (Center of The Regressor Space) ซึ่ง $1/n \le h_{ii} < 1$ และ $\sum_{i=1}^{n} h_{ii} = p+1$ และค่าเฉลี่ยของ h_{ii} เท่ากับ $(p+1)/n$ เมื่อ p คือ จำนวนตัวแปรอิสระ และ n คือ ขนาดตัวอย่าง

และ Hoaglin และ Welsh (1978) ได้เสนอเกณฑ์ที่ใช้ในการตัดสินว่าตัวแปรอิสระของตำแหน่งที่ i มีค่าผิดปกติมาก (High Leverage Point) โดยพิจารณาจากการที่ตำแหน่งที่ i มีค่า $h_{ii} > 2(p+1)/n$

ส่วนรูปแบบของ $v(\mathbf{X}_i)$ ในที่นี้จะใช้รูปแบบที่ถูกเสนอโดย Welsh (1980) และ Krasker และ Welsh (1982) คือ

$$v(\mathbf{X}_i) = \frac{(1-h_{ii})}{\sqrt{h_{ii}}}$$

แล้วเมื่อนำ $v(\mathbf{X}_i)$ ไปแทนในสมการ (6) จะได้

$$\eta\left(\mathbf{X}_i, \frac{e_i}{\hat{\sigma}}\right) = \left(\frac{1-h_{ii}}{\sqrt{h_{ii}}}\right)\psi\left(\frac{e_i\sqrt{h_{ii}}}{\hat{\sigma}(1-h_{ii})}\right)$$

และจากสมการ (5) สามารถเขียนเป็น

$$\sum_{i=1}^{n} v(\mathbf{X}_i)\,\psi\left(\frac{e_i}{v(\mathbf{X}_i)\cdot\hat{\sigma}}\right)\mathbf{X}_i = \underline{0} \qquad (7)$$

และ Myers (1990) ได้ทำการศึกษาพบว่า ฟังก์ชัน ψ เป็นฟังก์ชันของ DFFITS คือ

$$\psi\left(\frac{e_i}{\hat{\sigma}(1-h_{ii})^{1/2}}\left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2}\right) = \psi(\mathrm{DFFITS}_i)$$

ซึ่งค่า $DFFITS_i$ เป็น ค่าที่ใช้วัดอิทธิพล (Influence Measure) ของค่าสังเกตที่ i ที่มีต่อการพยากรณ์ของค่าที่ i ($\hat{y}_i$)

ดังนั้นในการหา $\hat{\beta}$ ที่เหมาะสมจะพิจารณาสมการ (7) แทนสมการ (5) และเนื่องจากสมการ (7) เป็นสมการแบบไม่เป็นเชิงเส้น (Nonlinear) จึงใช้วิธีกำลังสองน้อย-ที่สุดแบบถ่วงน้ำหนักซ้ำหลายรอบ (IRLS) เช่นเดียวกับวิธีตัวประมาณ M โดย

$$\hat{\beta} = (\mathbf{x}^T\mathbf{W}\mathbf{x})^{-1}\mathbf{x}^T\mathbf{W}\mathbf{y}$$

เมื่อ $\mathbf{W}$ คือ เมตริกซ์แนวทแยงมุมขนาด nxn และมี w_i เป็นสมาชิกแนวทแยงมุมโดยที่

$$w_i(DFFITS_i) = \frac{\psi(DFFITS_i)}{DFFITS_i}$$

สำหรับเมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ คือ

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 \frac{E[\psi^2(DFFITS_i)]}{\{E[\psi'(DFFITS_i)]\}^2}(\mathbf{x}^T\mathbf{x})^{-1}$$

และเมตริกซ์ความแปรปรวนร่วมของ $\hat{\beta}$ โดยประมาณ คือ

$$\hat{\mathrm{C}}\mathrm{ov}(\hat{\beta}) = \frac{(n\hat{\sigma})^2}{n-(p+1)} \frac{\sum_{i=1}^{n}\psi^2(DFFITS_i)}{\left\{\sum_{i=1}^{n}\psi'(DFFITS_i)\right\}^2}(\mathbf{x}^T\mathbf{x})^{-1}$$

และเมื่อใช้หลักการคำนวณของวิธี WLS ได้

$$\hat{\mathrm{C}}\mathrm{ov}(\hat{\beta}) = \frac{\sum_{i=1}^{n} w_i \cdot (y_i - \mathbf{x}_i^T\hat{\beta})^2}{n-(p+1)}(\mathbf{x}^T\mathbf{W}\mathbf{x})^{-1}$$

ส่วนเกณฑ์ความแกร่งหรือรูปแบบของฟังก์ชัน ψ จะใช้เกณฑ์ความแกร่งของ Huber และของ Tukey เช่นเดียวกับวิธีตัวประมาณ M โดย Walker (1984) ได้เสนอค่าคงที่ (Tuning Constants) ในการพิจารณาค่าผิดปกติของ $DFFITS_i$ ของทั้งเกณฑ์ 2 ดังนี้

เกณฑ์ความแกร่งของ Huber

$$c_H^* = 1.345 \frac{(2n(p+1))^{1/2}}{n-2(p+1)}$$

และ เกณฑ์ความแกร่งของ Tukey

$$c_B^* = 4.685 \frac{(2n(p+1))^{1/2}}{n-2(p+1)}$$

เมื่อ p คือ จำนวนตัวแปรอิสระ และ n คือ ขนาดตัวอย่าง