บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้สิ่งที่สนใจศึกษา คือ ศึกษาเปรียบเทียบประสิทธิภาพของวิธีที่จะใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุของตัวแบบการถดถอยเชิงเส้น เมื่อค่าผิดพลาดมีการแจกแจงแบบต่าง ๆ โดยใช้ค่าเฉลี่ยความผิดพลาดกำลังสองเป็นเกณฑ์ในการพิจารณา ในกรณีที่ค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ การแจกแจงแบบเบ้ และการแจกแจงแบบดับเบิ้ลเอกซ์โปเนนเชียลโดยจะทำการศึกษาเปรียบเทียบระหว่างวิธีกำลังสองน้อยที่สุดวิธีบูตสแตรป และวิธีตัวประมาณชนิด M ที่ใช้เกณฑ์ความแกร่งของ Huber และสำหรับการแจกแจงแบบเบ้จะทำการศึกษา การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยใช้การแปลงในรูปการยกกำลังของ Box และ Cox (1964 :211-243) แปลงข้อมูลให้มีการแจกแจงเข้าใกล้ภาวะปกติเสียก่อนแล้วจึงนำไปประมาณค่าสัมประสิทธิ์การถดถอยพหุ 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)

2.2.4 วิธีตัวประมาณชนิด M (M-estimator method)

2.2.5 วิธีบูตสแตรป (Bootstrap method)

2.3 การแปลงข้อมูลด้วยการแปลงที่ใช้การยกกำลัง (Power Transformation)

2.4 ผลงานวิจัยที่เกี่ยวข้อง

2.1 ข้อสมมติทั่วไปของตัวแบบการถดถอยเชิงเส้น

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) $$\underset{\sim}{y} = X\underset{\sim}{B} + \underset{\sim}{\varepsilon}$$

เมื่อ

$$\underset{\sim}{Y} = \underset{n\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}, \quad \underset{\sim}{X} = \underset{n\times(p+1)}{\begin{bmatrix} 1 & X_{11}\ldots X_{1p} \\ 1 & X_{21}\ldots X_{2p} \\ \vdots & \vdots \\ 1 & X_{n1}\ldots X_{np} \end{bmatrix}}, \quad \underset{\sim}{\beta} = \underset{n\times 1}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}}, \quad \underset{\sim}{\varepsilon} = \underset{n\times 1}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

ซึ่งมีข้อสมมติ คือ

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

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

ข) $E(\varepsilon\varepsilon') = \sigma^2 I_n$

$$E(\varepsilon\varepsilon') = \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \ldots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_1\varepsilon_2) & E(\varepsilon_2^2) & \ldots & E(\varepsilon_2\varepsilon_n) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ E(\varepsilon_1\varepsilon_n) & E(\varepsilon_2\varepsilon_n) & \ldots & E(\varepsilon_n^2) \end{bmatrix}_{n\times n}$$

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

$$= \sigma^2 I_n$$

ค) X เป็นชุดของค่าตัวเลขที่คงที่ หรือ X เป็นเมตริกซ์ค่าคงที่ หมายถึงการสุ่มตัวอย่างที่ซ้ำกัน จะเกิดค่าหลายค่าในเวกเตอร์ $\underset{\sim}{y}$ ซึ่งแปรผันในเวกเตอร์ $\underset{\sim}{\varepsilon}$ และจำนวนค่าสังเกตของ X จะต้องมีจำนวนเกินจำนวนพารามิเตอร์ที่จะประมาณ

ง) X มี full rank กล่าวคือ rank (X) $= p < n$ ทั้งนี้จะต้องไม่มีความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ ซึ่งเราเรียกความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระนี้ว่าภาวะความไม่เป็นอิสระซึ่งกันและกัน

2.2 การประมาณค่าสัมประสิทธิ์การถดถอยพหุ

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

ให้ $\hat{\underset{\sim}{\beta}}$ เป็นค่าประมาณของ β โดยที่ค่าประมาณนี้กระทำโดยการอาศัยข้อมูลจากค่าสังเกต

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

โดยหลักเกณฑ์ดังกล่าวเราสามารถหาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุดังต่อไปนี้

2.2.1 วิธีกำลังสองน้อยที่สุดแบบธรรมดา (Ordinary Least Square Method (OLS))

วิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นวิธีนี้มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of Linear Estimation) โดย คาร์ล เฟดริก เกาส์ (Karl Friedrick Gauss) ในปี ค.ศ. 1777-1855 และ อังเดร แอนดรีวิช มาร์คอฟ (Andrei Andrewich Markov) ในปี ค.ศ. 1856-1922 โดยมีหลักในการประมาณค่าสัมประสิทธิ์การถดถอยคือทำให้ผลบวกกำลังสองของค่าผิดพลาดมีค่าน้อยที่สุด ซึ่งแสดงรายละเอียดได้ดังนี้

นิยามที่ 2.1

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

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

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

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

จาก (2.2.1.1) ผลบวกกำลังสองของค่าผิดพลาด คือ

$$
\begin{aligned}
(2.2.1.2)\qquad SSE &= \underset{\sim}{\hat{\varepsilon}}'\underset{\sim}{\hat{\varepsilon}} \\
&= (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}) \\
&= (\underset{\sim}{y}' - \hat{\underset{\sim}{\beta}}'X')(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}) \\
&= \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'X\hat{\underset{\sim}{\beta}} - \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}} \\
(2.2.1.3)\qquad &= \underset{\sim}{y}'\underset{\sim}{y} - 2\hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y} + \hat{\underset{\sim}{\beta}}'X'X\hat{\underset{\sim}{\beta}}
\end{aligned}
$$

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

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

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

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

เมื่อ $(X'X)^{-1}X'$ เป็นเมตริกซ์ของค่าคงที่ และสมาชิกใน $\hat{\underset{\sim}{\beta}}$ เป็นฟังก์ชั่นเชิงเส้นของ $\underset{\sim}{Y}$ ดังนั้น

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

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

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

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

$$
\begin{aligned}
Cov(\hat{\underset{\sim}{\beta}}) &= Cov(L\underset{\sim}{y}) \\
&= L\, Cov(\hat{\underset{\sim}{y}})\, L' \\
&= L\, \sigma^2 p L' \\
&= \sigma^2\, LpL' \\
(2.2.1.6)\qquad &= \sigma^2\, (X'X)^{-1}
\end{aligned}
$$

018558

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 คือเมตริกซ์ใด ๆ ซึ่งมีขนาด k x n

กำหนด $M = (X'X)^{-1}X'$

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

กล่าวคือ $\hat{\underset{\sim}{\beta}}_j = \sum_{i=1}^{n} m_{ji} y_i \; ; \; j = 1, 2, \ldots, k$

เมื่อ $\sum_{i=1}^{n} m_{ji} = M$

ซึ่งแสดงว่า $\hat{\underset{\sim}{\beta}}$ เป็นฟังก์ชั่นเชิงเส้นของ $\underset{\sim}{y}$

ในทำนองเดียวกัน

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

$$\hat{\underset{\sim}{\beta}}_j^* = \sum_{i=1}^{n} (m_{ji} + C_{ij})Y_i$$

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

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

จะได้ว่า

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

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

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

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

จาก (2.2.1.6) จะได้ว่า

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

ซึ่งเป็นเมตริกซ์กึ่งบวกแน่นอน (positive semi-definite 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 วิธีกำลังสองน้อยที่สุดแบบทั่วไป (General Least Square Method (GLS))

ในทางปฏิบัติจะได้ว่า $E(\varepsilon_i\varepsilon_j) \neq 0$ ทุกค่าของ $i \neq j$ และ $E(\varepsilon_i\varepsilon_j) = \sigma^2$ สำหรับทุกค่าของ $i = j$ เช่นในตัวแบบถดถอยแบบอัตโนมัติหรือในบางกรณีเกิดปัญหาเรื่อง heterocedasticity ในสถาณการณ์เช่นนี้ $E(\underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}') = \sigma^2\Omega$ โดยที่ Ω เป็น เมตริกซ์สมมาตรบวกแน่นอนขนาด n X n และ $\underset{\sim}{\varepsilon} \sim N_n(0, \sigma^2\Omega)$

เนื่องจาก Ω เป็น เมตริกซ์สมมาตรบวกแน่นอนเราจึงหาเมตริกซ์ P ซึ่งทำให้ $P'P = \Omega$ ได้

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

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

ดังนั้น เราสามารถปรับสมการ (2.2.2.1) ให้อยู่ในรูปดังนี้

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

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

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

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

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

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

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

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

และความแปรปรวนของตัวประมาณจะมีค่าเป็น

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

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

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

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

2.2.3 วิธีกำลังสองแบบถ่วงน้ำหนัก (Weighted Least Square Method (WLS))

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

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

$$= \sigma^2 \begin{bmatrix} 1/w_1 & 0 \ldots & 0 \\ 0 & 1/w_2 \ldots & 0 \\ . & . & . \\ . & . & . \\ 0 & 0 \ldots & 1/w_n \end{bmatrix}_{n \times n}$$

เมื่อ $W = \Omega^{-1}$ และ Ω เป็นเมตริกซ์ทแยงมุม (diagonal matrix) จะได้ว่า W เป็นเมตริกซ์ทแยงมุมด้วย โดยมีเส้นทแยงมุมเป็น W_1, W_2, W_3,...W_n และสมการของ weighted least square หรือ WLS อยู่ในรูปของ

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

ซึ่งเราเรียก w_i ว่าเป็น weight โดยที่ค่าสังเกตที่มี w_i น้อยจะมีความแปรปรวนมากกว่าค่าสังเกตที่มี w_i มาก

2.2.3 วิธีตัวประมาณชนิด M (M-Estimator Method)

เมื่อปี ค.ศ. 1964 P.J. Huber ได้ศึกษาฟังก์ชั่นความผิดพลาดซึ่งเรียกว่าตัวประมาณชนิด M (M-estimator) ตัวประมาณชนิด M จะประมาณพารามิเตอร์ของค่าน้อยที่สุดของ $\sum_i^n \rho(\varepsilon_i/s)$ เมื่อ ε_i เป็นค่าผิดพลาดของค่าสังเกตตัวที่ i และ s เป็นตัวประมาณของการกระจายของตัวอย่างของ ε_i

ในตัวแปรการถดถอยเชิงเส้นเราสามารถหาตัวประมาณชนิด M จาก

$$\text{(2.3.1)} \qquad \min_{\underset{\sim}{\beta}} \sum_i^n \rho(\varepsilon_i/s) = \min_{\underset{\sim}{\beta}} \sum_i^n \rho\{(\underset{\sim}{y_i} - X_i'\underset{\sim}{\beta})/s\}$$

เมื่อ s เป็นตัวประมาณที่แกร่งของสเกล

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

$$\text{(2.3.2)} \qquad \sum x_{ij}\,\psi\{(\underset{\sim}{Y_i} - X_i'\underset{\sim}{\beta})/s\} = 0$$

$$\text{(2.3.3)} \qquad \sum \chi\,\{(Y_i - X_i'\underset{\sim}{\beta})/s\} = 0$$

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

ในปี ค.ศ.1971 P.J. Bickel ได้ศึกษาต่อจากงานของ Huber's Proposal 2 และทำการเสนอวิธีที่เรียกว่า One Step Huber โดยการกำหนดค่าเริ่มต้นได้แก่ ตัวประมาณตำแหน่งลำดับที่ 1 ซึ่งอาจใช้ค่าเฉลี่ย หรือค่ามัธยฐาน (mean or median) และตัวประมาณที่แกร่งของสเกล โดยสามารถหาค่าประมาณพารามิเตอร์ตำแหน่งลำดับถัดไปได้โดยที่สามารถเขียนอยู่ในรูปของ

$$T^{(m+1)} = T^{(m)} + \frac{(1/n)s^{(0)} \Sigma\psi\{(y_i - T^{(m)})/s^{(0)}\}}{(1/n)\Sigma\psi'\{(y_i - T^{(m)})/s^{(0)}\}} \quad (2.3.4)$$

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

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

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

ต่อมาในปี ค.ศ. 1977 Huber ได้เสนอรูปแบบของค่าน้อยที่สุดของฟังก์ชั่นความผิดพลาดสำหรับ β และสเกล ซึ่งอยู่ในรูปของ

$$Q(\beta, \sigma) = (1/n)\Sigma\rho\{(\underset{\sim}{y}_i - X_i'\underset{\sim}{\beta})/\sigma\}\sigma \quad (2.3.5)$$

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

เมื่อเราหาอนุพันธ์บางส่วนเมื่อเทียบกับ $\underset{\sim}{\beta}$ และ σ จากสมการที่ (2.3.2) และ (2.3.3) จะสามารถหาค่าประมาณของ $\underset{\sim}{\beta}$ และสเกล(scale) โดยการ iteration สมการ (2.3.2) และ (2.3.3) พร้อม ๆ กัน เรียกว่า simultaneous estimation of location and scale ซึ่งในทางปฏิบัติจะกระทำได้ยาก

โดยทั่วไปค่า s ที่ประมาณขึ้นมาจะถูกเลือกไม่ให้มีอิทธิพลกับความผิดพลาดที่มีขนาดใหญ่ ดังนั้นฟังก์ชันความผิดพลาดต่าง ๆ จะมีค่าความผิดพลาดน้อยที่สุดเมื่อกำหนดค่า s คงที่

และวิธีการประมาณค่าสัมประสิทธิ์การถดถอยส่วนใหญ่จะใช้วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักซ้ำหลาย ๆ รอบ (iteratively reweighted least square) ของ

Beaton และ Tukey (1964) โดยจะต้องประมาณ β และ s เราสามารถเขียนสมการ m สมการของ (2.3.5) ได้ดังนี้

$$X'W_oX\underset{\sim}{\beta} = X'W_o\underset{\sim}{y}$$

2.2.5 การประมาณค่าพารามิเตอร์โดยวิธีบูตสแตรป (Bootstrap Method)

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้ เป็นวิธีที่เสนอขึ้นโดย Bradley Efron ในปี ค.ศ. 1979 โดยมีหลักเกณฑ์ดังนี้คือ ทำการสุ่มตัวอย่างจากข้อมูลที่เก็บรวบรวมมาแบบใส่คืน (with replacement) ขนาดเท่ากับจำนวนตัวอย่างหรือข้อมูลที่มีอยู่แล้วนั้น เพื่อสร้างข้อมูลชุดใหม่แล้วนำมาใช้ในการประมาณค่าพารามิเตอร์ที่สนใจ

สำหรับการหาตัวประมาณโดยวิธีนี้ สามารถทำได้ 3 แบบ คือ โดยการคำนวณหาจากสูตร โดยใช้เทคนิคมอนติคาร์โล และโดยหาจากการกระจายของอนุกรมเทย์เลอร์ (Taylor series) ซึ่งส่วนใหญ่จะใช้เทคนิคมอนติคาร์โล โดยนำคอมพิวเตอร์มาใช้เป็นเครื่องมือช่วยโดยมีขั้นตอนการทำงานดังนี้

1. สร้างตัวเลขสุ่ม เพื่อนำไปใช้ในการสุ่มตัวอย่างแบบใส่คืน

2. จากตัวอย่างที่ได้แต่ละชุดนำมาหาค่าสถิติหรือค่าประมาณของพารามิเตอร์ที่สนใจ

ตัวอย่าง การประมาณค่าความผิดพลาดมาตรฐานของสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient ($\hat{\rho}$) หรือ SD ($\hat{\rho}$) โดยวิธีบูตสแตรป

ให้ $x_1, x_2, \ldots, x_n$ iid F ซึ่ง F เป็นการแจกแจงที่ไม่ทราบ

$x_i = (y_i, z_i)$; $i = 1, 2, \ldots, n$

จาก
$$\hat{\rho} = \frac{\sum_{i=1}^{n} Y_i Z_i - \sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Z_i / n}{[\{\sum_{i=1}^{n} Y_i - (\sum_{i=1}^{n} Y_i)^2/n\}\{\sum_{i=1}^{n} Z_i - (\sum_{i=1}^{n} Z_i)^2/n\}]^{1/2}}$$

จะเห็นว่าไม่สามารถหาค่าประมาณของ SD ( ) ได้จากข้อมูลชุดนี้โดยใช้สูตรทั่วไป ซึ่ง $SD(x) = \sum_{i=1}^{n} (x_i - \bar{x})^2/n-1$ คือ เมื่อ n คือขนาดของตัวอย่าง จึงนำเอาวิธีบูตสแตรปมาใช้ในการประมาณค่า โดยใช้เทคนิคมอนติคาร์โล

ขั้นตอน การประมาณค่า SD $(\hat{\rho})$ โดยวิธีบูตสแตรป

1) ให้ F เป็นการแจกแจง x_i ; $i = 1,2,\ldots,n$

2) สุ่ม x_i แบบใส่คืน (with replacment) ขนาด n

ดังนั้น x_i แต่ละตัวจะมีโอกาสถูกเลือกเป็นตัวอย่าง = 1/n

ได้ $x_1^*, x_2^*, \ldots, x_n^*$ iid F

เรียก x_i^* นี้ว่า ตัวอย่างของบูตสแตรป (bootstrap sample) คำนวณค่า P

3) กระทำในข้อ 2) B ครั้ง จะได้ $\hat{\rho}^{*1}$, $\hat{\rho}^{*2}$, ... , $\hat{\rho}^{*B}$

4) คำนวณหาค่าประมาณของ SD $(\hat{\rho})$ โดยวิธีบูตสแตรป $(\sigma_B(\hat{\rho}))$

ได้ $\sigma_B(\hat{\rho}) = \mathrm{var}(\hat{\rho})^{1/2}$

ซึ่ง $\mathrm{var}(\hat{\rho})$ คือ ค่าประมาณความแปรปรวน (variance) ของ $\hat{\rho}$ ภายใต้ F

$$\sigma_B(\hat{\rho}) = \left[\frac{\sum_{i=1}^{B} (\hat{\rho}^{*}_{i} - \bar{\hat{\rho}}^{*})}{B-1}\right]^{1/2}$$

เมื่อ

$$\bar{\hat{\rho}}^{*} = \frac{\sum_{i=1}^{B} \hat{\rho}^{*i}}{B}$$

และได้ว่า $\bar{\hat{\rho}}^{*}$ เป็นค่าประมาณของค่าคาดหมายของ $\hat{\rho}$ $(E(\hat{\rho}))$ ที่ได้จากวิธีบูตสแตรป

ในการศึกษาครั้งนี้ได้นำเอาวิธีบูตสแตรปมาใช้ในการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยเชิงเส้น และหาค่าความแปรปรวน ในกรณีที่ไม่ทราบลักษณะการแจกแจงของความคลาดเคลื่อน ซึ่งมีรายละเอียดในการหาแสดงในรูปของเวคเตอร์และเมตริกซ์ได้ดังนี้

2.3.1 การหาตัวประมาณโดยวิธีบูตสแตรป

จาก $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

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

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

$\underset{\sim}{\varepsilon}_i \sim$ iid F ; ไม่ทราบการแจกแจง F

การหาตัวประมาณกระทำตามขั้นตอนต่อไปนี้

1. คำนวณหาค่า $\hat{\beta}$ โดยวิธีกำลังสองต่ำสุด

ได้ $\hat{\beta} = (X'X)^{-1} X'Y$

และ $\hat{Y} = X\hat{\underset{\sim}{\beta}}$

$\underset{\sim}{\varepsilon} = Y-\hat{Y}$ ซึ่ง $\sum_{i=1}^{n} \varepsilon_i = 0$

2. สุ่ม ε_i แบบใส่คืน (with replacement) ขนาด n

ได้ $\varepsilon^*_1, \varepsilon^*_2, \ldots, \varepsilon^*_n$

$\varepsilon^*_i \sim$ iid F

3. นำค่า ε^*_i มาพิจารณาโดยรวมไว้ในสมการ

จะได้ $Y^* = X\hat{\beta} + \varepsilon^* = Y + \varepsilon^*$

คำนวณหาค่า $\hat{\beta}^*$ โดยวิธีกำลังสองต่ำสุด โดยยึดหลักเกณฑ์เดียวกัน คือทำการหา $\hat{\beta}^*$ ซึ่งเป็นตัวประมาณกำลังสองต่ำสุดของ β^* ที่ทำให้

SEE* = $\sum_{i=1}^{n} \varepsilon^{*2}_i = \hat{\varepsilon}^{*\prime} \hat{\varepsilon}^*$ มีค่าต่ำสุด

โดยการหาอนุพันธ์ของ SEE เทียบกับ β แล้วกำหนดให้เท่ากับ 0 นั่นคือ

$$
\begin{aligned}
SEE^* = \hat{\varepsilon}^{*\prime} \hat{\varepsilon}^* &= (Y^* - X\hat{\beta}^*)'(Y^* - X\hat{\beta}^*) \\
&= (Y^{*\prime} - \hat{\beta}^{*\prime}X')(Y^* - X\hat{\beta}^*) \\
&= Y^{*\prime} Y^* - Y^{*\prime}X\hat{\beta}^* - \hat{\beta}^{*\prime} X' Y^* + \hat{\beta}^{*\prime} X' X\hat{\beta}^* \\
&= Y^{*\prime} Y^* - 2\hat{\beta}^{*\prime} X' Y^* + \hat{\beta}^{*\prime} X' X\hat{\beta}^*
\end{aligned}
$$

$$\frac{\partial SSE}{\partial \beta} = -2X'Y^* + 2X'X\hat{\beta} = 0$$

$$X'X\hat{\beta}^* = X'Y^*$$

$$\hat{\beta}^* = (X'X)^{-1} X'Y^*$$

4. กระทำตามขั้นตอนในข้อ 2-3 ซ้ำ B ครั้ง จะได้ $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \ldots\ldots \hat{\beta}^{*B}$

5. คำนวณหาค่า $\bar{\hat{\beta}}^* = \sum_{i=1}^{B} \hat{\beta}^{*i}/B$

จากหลักการของวิธีบูตสแตรป จะได้ว่า $\bar{\hat{\beta}}^*$ เป็นค่าประมาณของ $E(\hat{\beta}) = B$ ซึ่งการศึกษาครั้งนี้จึงสนใจและใช้ $\bar{\hat{\beta}}^*$ เป็นตัวประมาณของ β ที่ได้จากวิธีบูตสแตรป

ในที่นี้จะเห็นว่า ε_i^* เป็นตัวอย่างที่สุ่มได้โดยวิธีบูตสแตรป (bootstrap sample) ซึ่งมีคุณสมบัติดังนี้คือ

1. $E(\varepsilon_i^*) = 0 \quad ; \quad i = 1, 2 \ldots.n$
2. $cov(\varepsilon_i^*, \varepsilon_j^*) = 0$
3. $V(\varepsilon_i^*) = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$

พิสูจน์ $\quad E(\varepsilon_i^*) = 0$

เมื่อ ε_i^* เป็นการสุ่มตัวอย่างแบบใส่คืนจาก $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots., \hat{\varepsilon}_n$

ให้ b_i = จำนวนครั้งที่หน่วย i ตกอยู่ในตัวอย่าง

$$b_i \quad \text{Binomial} \quad (n, p = \frac{1}{n})$$

$$E(b_i) = \frac{n}{n} = 1$$

$$V(b_i) = n\frac{1}{n}(1 - \frac{1}{n}) = \frac{n-1}{n}$$

$$cov(b_i, b_j) = -n.p_i.p_j = -n\frac{1}{n}\frac{1}{n} = -\frac{1}{n}$$

$$\begin{aligned}
E(\varepsilon_i^*) &= E\{E(\varepsilon_i^* / \hat{\varepsilon}_i, \ldots., \hat{\varepsilon}_n)\} \\
&= E\{\sum_{i=1}^{n} p_i \varepsilon_i^*) \\
&= E\{\sum_{i=1}^{n} \frac{1}{n} \varepsilon_i^*\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i E(b_i) \\
E(\varepsilon_i^*) &= 0 \quad \because \sum_{i=1}^{n} \hat{\varepsilon}_i = 0
\end{aligned}$$

พิสูจน์ $\mathrm{cov}\ (\varepsilon_i^*, \varepsilon_j^*) = 0$

$\mathrm{cov}\ (\varepsilon_i^*, \varepsilon_j^*) = 0$

เพราะว่า ε_i^* และ ε_j^* เป็นอิสระแก่กัน เนื่องจากการสุ่มแบบใส่คืนจาก $\hat{\varepsilon}_1, \hat{\varepsilon}_2, .., \hat{\varepsilon}_n$

พิสูจน์ $V(\varepsilon_i^*) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2$

จาก
$$
\begin{aligned}
V(\varepsilon_i^*) &= E(\varepsilon_i^{*2}) - \{E(\varepsilon_i^*)\}^2 \\
&= E(\varepsilon_i^{*2}) \\
&= E\{\sum_{i=1}^{n} pi\ \ \varepsilon_i^*\} \\
&= E\{\sum_{i=1}^{n} \frac{1}{n} \varepsilon_i^{*2}\} \\
&= E\{\sum_{i=1}^{n} \frac{1}{n} Bi\ \hat{\varepsilon}_i^{*2}\} \\
&= \frac{1}{n}\{\sum_{i=1}^{n} \hat{\varepsilon}_i^2\ E(bi)\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 . 1 \\
&= \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2
\end{aligned}
$$

2.3 การแปลงข้อมูลที่อยู่ในรูปยกกำลัง (Power Transformation)

การแปลงข้อมูลที่อยู่ในรูปยกกำลัง (Power Transformation) ได้รวมการแปลงซึ่งใช้ลอกการิทึมในวงศ์ (family) ของการแปลง การแปลงข้อมูลอาจทำให้ข้อสมมติของความเป็นไปได้สำหรับแปลงข้อมูลที่ใช้ศึกษาแจกแจงของความผิดพลาดเป็นแบบเบ้

Tukey (ค.ศ. 1960, 222-232) เป็นท่านแรกที่ศึกษารายละเอียดเกี่ยวกับการแปลงข้อมูลได้แนะนำว่า การแปลงข้อมูลจะช่วยให้ตัวแบบมีการแจกแจงเข้าใกล้เชิงเส้นโดยที่ข้อมูลจะมีการกระจายเท่ากันและเป็นปกติมากขึ้น

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

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

จึงเป็นข้อดีของการแปลง และเมื่อเปรียบเทียบกับการแปลงอย่างง่าย y_i ซึ่งต่อเนื่องเมื่อ $\lambda = 0$ ดังนั้น

$$\lim_{\lambda \to 0} \frac{y_i^{\lambda} - 1}{\lambda} = \ln y_i$$

Box และ Cox จึงพิจารณาตัวแบบของการถดถอยในรูปสมการเมตริกซ์ดังรูป

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

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

$$\left(\frac{1}{2\pi\sigma^2}\right)\frac{n}{2} \exp \left| -\frac{1}{2\sigma^2} \sum_{i}^{n} (y_i^{\lambda} - \underset{\sim}{X_i'}\underset{\sim}{\beta})^2 \right| . J$$

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

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

จะได้ว่าฟังก์ชัน log-likelihood อยู่ในรูปของ

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

และถ้าหาร y_i แต่ละตัวด้วย ค่าเฉลี่ยเรขาคณิต (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 จะได้ว่าขั้นตอนต่าง ๆ มีดังนี้

1. หาร Y แต่ละค่าด้วยค่าเฉลี่ยเรขาคณิตของ Y
2. สำหรับค่าของ λ ทำการถดถอย $y^{(\lambda)}$ บน x_i และคำนวนผลรวมกำลังสองของค่าผิดพลาด
3. เลือกค่า λ สำหรับ $\sigma^2_{\varepsilon|\lambda}$ ที่น้อยที่สุด จะได้ว่า λ' เป็น ภาวะน่าจะเป็นสูงสุด (maximum likelihood) ของ λ

ค่าผิดพลาดมาตรฐานของสัมประสิทธิ์การถดถอยสำหรับค่า λ ใด ๆ สามารถหาได้จากโปรแกรมการถดถอย ซึ่งเป็นค่าผิดพลาดมาตรฐานที่มีเงื่อนไข บนค่าสมมติ λ ค่าผิดพลาดมาตรฐาน หรือช่วงความเชื่อมั่นสำหรับ λ สามารถหาได้จากการหาส่วนกลับของอัตราส่วนภาวะน่าจะเป็นของการทดสอบทางสถิติ (likelihood ratio test statistics) และเมื่อต้องการที่จะทดสอบสมมติฐาน $\lambda = \lambda_0$ จะได้ว่า

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

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

และ $-2 \ln \theta \; X^2(1)$ ดังนั้น ถ้าต้องการ 95 เปอร์เซนต์ของช่วงความเชื่อมั่นสำหรับ $\hat{\lambda}$ จะต้องพิจารณาทุก ๆ ค่าของ λ ซึ่ง $-2 \ln \theta$ มีค่าน้อยกว่า 3.84 เพราะว่า $Pr(-2 \ln \theta < 3.84) = 0.95$ จะได้ว่า

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

เมื่อประเมินค่า $\hat{\sigma}_{\varepsilon|\lambda}$ สำหรับ λ ที่แตกต่างกันในการหา ML ของตัวประมาณ λ 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

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

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

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) ได้ศึกษาถึงวิธีการที่แกร่งสำหรับการทดสอบการแปลงข้อมูลที่สู่เข้าสู่ภาวะปกติโดยประมาณ ดังนั้นค่าผิดพลาดอาจเข้าใกล้ภาวะปกติเท่านั้น Hinkley(ค.ศ.1975) ได้กล่าวว่ามันเป็นข้อพกพร่องที่ไม่พิจารณาความเป็นไปได้เรื่องนี้ Carroll (ค.ศ.1980) จึงพิจารณาวิธีการใหม่โดยอาศัยความรู้พื้นฐานการแปลงข้อมูลของ Box และ Cox กับวิธีการที่แกร่ง กล่าวคือ จะแทนทฤษฎีในภาวะน่าจะเป็นเดิมเมื่อการแจกแจงเป็นแบบปกติด้วยความหนาแน่นเป็นแบบปกติที่มีศูนย์กลาง การแจกแจงแบบปกติด้วยความหนาแน่นแบบเอกซ์โปเนนเชียล

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

สำหรับ λ ที่คงที่จะกระทำเช่นเดียวกับปัญหาที่เกี่ยวกับตำแหน่ง โดยการหาค่าเริ่มต้นของ β ซึ่งทำให้สมการข้างล่างมีค่ามากที่สุด

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

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ จะใช้วิธีตัวประมาณชนิด M เพื่อจะประมาณค่า $\hat{\underset{\sim}{\beta}}$ และ $\hat{\lambda}$ ซึ่งจะให้ค่า $\sigma^2_{\varepsilon/\lambda}$ ที่มีค่าน้อยที่สุด

2.4 ผลงานการวิจัยที่เกี่ยวข้อง

เนื่องจาก σ^2 เป็นค่าที่ไม่ทราบ จึงไม่สามารถหาค่า $V(\hat{\beta})$ ซึ่งเท่ากับ $\sigma^2(X'X)^{-1}$ ได้ในปี ค.ศ. 1979 Bradley Efron ได้ศึกษาและนำเอาวิธีบูตสแตรปมาใช้ในการประมาณค่าความผิดพลาดมาตรฐานของสัมประสิทธิ์ความถดถอย ซึ่งก็คือ $[V(\hat{\beta})]^{1/2}$ ในสมการถดถอยที่มีรูปแบบ (model) ที่ยุ่งยากซับซ้อนและไม่ทราบลักษณะการแจกแจงของความผิดพลาด ทั้งนี้ตั้งอยู่บนสมมติฐานที่ว่า รูปแบบของสมการถูกต้องและใช้ตัวประมาณของสัมประสิทธิ์ความถดถอยที่ได้จากวิธีกำลังสองต่ำสุดเป็นตัวประมาณของพารามิเตอร์ โดยนำคอมพิวเตอร์มาใช้ในการสุ่มตัวอย่างโดยใช้เทคนิควิธีมอนติคาร์โล ในกรณีที่สมการมีรูปแบบทั่วไปคือ $Y = X\beta + \varepsilon$ จะมีขั้นตอนการหาค่าประมาณของความผิดพลาดมาตรฐานของสัมประสิทธิ์ความถดถอย เช่นเดียวกับ

ตัวอย่าง ในหัวข้อ 2.3 ได้ว่า $V(\hat{\beta}^*) = \sum_{i=1}^{B} (\hat{\beta}^{*i} - \bar{\hat{\beta}}^*)(\hat{\beta}^{*i} - \bar{\hat{\beta}}^*)' / B-1$ และสามารถหา $V(\hat{\beta}^*)$ โดยไม่ใช้เทคนิคมอนติคาร์โลได้ซึ่งมีค่าเท่ากับ $\sigma^2(X'X)^{-1}$ ซึ่ง $\hat{\sigma}^2 = \sum_{i=1}^{n} \varepsilon_i$ เป็น biased downward ของ σ^2 ทำให้ค่าประมาณของ $V(\hat{\beta})$ น้อยกว่าความเป็นจริงจึงใช้ $V(\hat{\beta}^*)$ ที่ได้จากวิธีบูตสแตรปเป็นค่าประมาณของ $V(\hat{\beta})$ และสรุปได้ว่าเมื่อ $B \rightarrow \infty$ ค่า $V(\hat{\beta}^*)$ ที่หาได้จะมีค่าเข้าใกล้ $V(\hat{\beta})$ และค่า B ซึ่งเป็นจำนวนครั้งในการทำบูตสแตรปที่เหมาะสมควรอยู่ในช่วง 50-200 ครั้ง ต่อมาในปี ค.ศ. 1983 David A. Freedman และ Stephen C. Peters ได้นำวิธีการนี้ไปใช้ในการประมาณค่า Standard error และพยากรณ์ในการวิเคราะห์ความถดถอยเชิงเส้น เมื่อขนาดของตัวอย่างมีจำนวนจำกัด (finite sample size) และไม่เป็นไปตามข้อตกลงเบื้องต้น ซึ่งพบมากในทางเศรษฐศาสตร์ เช่น ในกรณีที่เกิด Lag Structure Autoregressive, Structure Heteroscedasticity เป็นต้น สรุปผลได้ว่าวิธีบูตสแตรปสามารถให้ค่าประมาณของความผิดพลาดมาตรฐานของสัมประสิทธิ์ความถดถอยที่ดีและเข้าใกล้ค่าจริงมากกว่าค่าประมาณที่หาได้จากสูตรทั่วไป

ปราณี รัตนัง (2530) ได้ศึกษาการประมาณสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาดมีการแจกแจงแบบเบ้และมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ โดยการเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธีตัวประมาณชนิด M ซึ่งใช้เกณฑ์ความแกร่งของ (Ramsayและเกณฑ์การเปรียบเทียบที่ใช้คือ ค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองและค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง

ผลจากการศึกษาคือ

1. กรณีค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ เช่นการแจกแจงแบบปกติปลอมปน และการแจงแจงแบบที จะได้ว่าสเกลแฟกเตอร์ เปอร์เซนต์ของการปลอมปนจะมีอิทธิพลจากมากไปน้อยที่ทำให้วิธีตัวประมาณชนิด M ซึ่งใช้เกณฑ์ของ Ramsay ดีกว่าวิธีกำลังสองน้อยที่สุด ระดับความเป็นอิสระก็มีอิทธิพลที่ทำให้ได้ข้อสรุปเช่นเดียวกัน

2. กรณีค่าผิดพลาดมีการแจกแจงเบบแบ้ ได้แก่การแจกแจงแบบลอกนอร์มอลแกมมาและไวบูลล์ จะได้ว่าวิธีกำลังสองน้อยที่สุดและวิธีตัวประมาณชนิด M สามารถใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุใกล้เคียงกัน เมื่อใช้เทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox ในการแปลงข้อมูลให้เข้าสู่ภาวะปกติ

มาลี ตระการศิรินนท์ (2531) ได้ศึกษาเปรียบเทียบการประมาณค่าพารามิเตอร์ของรูปแบบสมการถดถอยเชิงเส้น ด้วยวิธีกำลังสองน้อยที่สุด และวิธีบุตสแตรป โดยใช้ค่าเฉลี่ยความผิดพลาดกำลังสองเป็นตัวเปรียบเทียบ

ผลการศึกษาพบว่า วิธีบูตสแตรป สามารถนำมาใช้ในการประมาณค่าพารามิเตอร์ใด ๆ ได้ ในกรณีที่ไม่ทราบลักษณะการแจกแจงของประชากรหรือค่าประมาณนั้นไม่สามารถหาได้โดยตรงจากสูตรทั่วไป และพบว่าเมื่อขนาดตัวอย่างจำนวนตัวแปรอิสระมีค่าน้อย วิธีบูตสแตรปมีประสิทธิภาพดีกว่าวิธีกำลังสองน้อยที่สุด

ทรงพันธ์ ชุณหสวัสดิกุล (2532) ได้ศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยที่ค่าประมาณสเกลเปลี่ยนไป โดยศึกษาเปรียบเทียบระหว่างวิธีกำลังสองน้อยที่สุด กับวิธีตัวประมาณชนิด M เมื่อใช้ตัวประมาณสเกล 3 วิธี คือ

1. The Standard Deviation of Location (STD)
2. The Median Absolute Deviation (MAD)
3. The Modified Biweight A-estimator (MBA)

ผลการศึกษาพบว่า เมื่อค่าความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ วิธีตัวประมาณชนิด M ที่ใช้ตัวประมาณสเกล MAD มีประสิทธิภาพดีกว่าวิธีกำลังสองน้อยที่สุด และวิธีตัวประมาณชนิด M ที่ใช้ตัวประมาณสเกลแบบอื่น แต่เมื่อค่าความผิดพลาดมีการแจกแจงแบบเบ้จะให้ผลการประมาณค่าสัมประสิทธิการถดถอยใกล้เคียงกันทุกวิธี