บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบของวิธีประมาณพารามิเตอร์ในการวิเคราะห์ความแปรปรวนร่วมระหว่างวิธีกำลังสองน้อยที่สุดกับวิธีตัวประมาณเอ็มที่ใช้เกณฑ์ความแกร่งของ Ramsay ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนชนิดหางยาวกว่าปกติว่าวิธีการใดจะส่งผลให้การวิเคราะห์ความแปรปรวนร่วมมีประสิทธิภาพกว่ากัน ซึ่งในบทนี้จะกล่าวถึงวิธีการวิเคราะห์ความแปรปรวนร่วมของแผนการทดลองแบบสุ่มตลอด (analysis of covariance of completely randomized design) ตัวสถิติที่ใช้ และรายละเอียดของแต่ละวิธีการประมาณรวมทั้งนำเสนอผลงานวิจัยที่เกี่ยวข้อง

2.1 ตัวสถิติที่ใช้ในการวิจัย

การวิเคราะห์ความแปรปรวนร่วมเป็นวิธีการหนึ่งที่ถูกนำมาใช้ทดสอบสมมุติฐานเพื่อเปรียบเทียบความแตกต่างระหว่างวิธีปฏิบัติ โดยนำหลักการของการวิเคราะห์ความแปรปรวนกับการวิเคราะห์ความถดถอยผสมผสานเข้าด้วยกัน เพื่อลดความแปรปรวนให้กับหน่วยทดลองซึ่งมีตัวแบบดังต่อไปนี้

ตัวแบบของการวิเคราะห์ความแปรปรวนของแผนแบบการทดลองแบบสุ่มตลอด

(1) $$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \qquad , i = 1,2,...n_j \quad j = 1,2,...p$$

ตัวแบบของการวิเคราะห์ความถดถอยเมื่อตัวแปรร่วมมีความสัมพันธ์เชิงเส้นกับตัวแปรตามคือ

(2) $$y_{ij} = \mu + \sum_{k=1}^{q}\beta_k(x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij}$$

จาก (1)และ (2) จะได้ว่า

$$y_{ij} = \mu + \tau_j + \sum_{k=1}^{q}\beta_k(x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij} \quad ; \quad \begin{array}{l} i=1,2,\ldots,n_j \\ j=1,2,\ldots,p \end{array}$$

โดยที่ p หมายถึง จำนวนวิธีปฏิบัติ

q หมายถึง จำนวนตัวแปรร่วม

n_j หมายถึง ขนาดตัวอย่างในแต่ละวิธีปฏิบัติที่ j

μ หมายถึง ค่าเฉลี่ยทั้งหมดของ y

τ_j หมายถึง อิทธิพลของวิธีปฏิบัติที่ j

β_k หมาบถึง ค่าสัมประสิทธิ์การถดถอยของตัวแปรร่วมที่ k

$\bar{x}_{..k}$ หมายถึง ค่าเฉลี่ยของตัวแปรร่วมที่ k

ε_{ij} หมายถึง ค่าความคลาดเคลื่อนของค่าสังเกตจากค่าเฉลี่ยที่ปรับแล้ว

ในการวิจัยครั้งนี้ใช้ตัวแบบและตัวสถิติดังนี้คือ

ตัวแบบเต็มรูป (full model)

$$y_{ij} = \mu + \tau_j + \sum_{k=1}^{q}\beta_k(x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij}$$

ตัวแบบลดรูป (reduced model)

$$y_{ij} = \mu + \sum_{k=1}^{q}\beta_k(x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij}$$

ในการทดสอบ

สมมุติฐานว่าง H_o : $\mu_1 = \mu_2 = \ldots = \mu_p$

เทียบกับ H_a : มีวิธีปฏิบัติอย่างน้อยที่สุด 2 วิธีปฏิบัติที่ต่างกัน

เมื่อ $\mu_j = \mu + \tau_j$

ตัวสถิติที่ใช้ คือ

$$F = \frac{(SSF-SSR)/(p-1)}{(SST-SSF)/(n-p-q)}$$

เมื่อ SSF หมายถึง ผลบวกกำลังสองของการประมาณตัวแบบเต็มรูป

SSR หมายถึง ผลบวกกำลังสองของการประมาณตัวแบบลดรูป

SST หมายถึง ผลบวกกำลังสองรวม

n หมายถึง ขนาดตัวอย่างทั้งหมดซึ่งเท่ากับ $\sum_{j=1}^{p} n_j$

จะปฏิเสธสมมุติฐานเมื่อ ค่าสถิติ $F > F_{\alpha}(p-1,\ n-p-q)$

จากตัวแบบเต็มรูป

$$Y_{ij} = \mu + \tau_j + \sum_{k=1}^{q}\beta_k(X_{ijk} - \bar{X}_{..k}) + \varepsilon_{ij}$$

เราสามารถเขียนให้อยู่ในรูปสมการถดถอย โดยสร้างตัวแปรหุ่น (dummy) เพื่อแสดงอิทธิพลของวิธีปฏิบัติได้ดังนี้

$$D_{i1} = \begin{cases} 1 & \text{, ถ้าเป็นวิธีปฏิบัติที่ 1} \\ 0 & \text{, อื่นๆ} \end{cases}$$

$$D_{i2} = \begin{cases} 1 & \text{, ถ้าเป็นวิธีปฏิบัติที่ 2} \\ 0 & \text{, อื่นๆ} \end{cases}$$

$$D_{ip-1} = \begin{cases} 1 & \text{, ถ้าเป็นวิธีปฏิบัติที่ p-1} \\ 0 & \text{, อื่นๆ} \end{cases}$$

จะได้ตัวแบบเต็มรูปใหม่คือ

$$y_{ij} = \alpha_0 + \alpha_1 D_{i1} + \alpha_2 D_{i2} + \ldots + \alpha_{p-1} D_{ip-1} + \sum_{k=1}^{q} \beta_k (x_{ijk} - \bar{x}_{..k}) + \varepsilon_{ij}$$

เมื่อ $\alpha_0, \alpha_1, \alpha_2, \ldots \alpha_{p-1}$ คือ สัมประสิทธิ์การถดถอยของตัวแปรหุ่น dummy ซึ่งจะแสดงอิทธิพลของวิธีปฏิบัติและเขียนให้อยู่ในรูปเมตริกซ์จะได้

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

เมื่อ $\underset{\sim}{y}$ คือ เวกเตอร์ของค่าสังเกตหรือตัวแปรตามที่มีขนาด n x 1

X คือ เมตริกซ์ของตัวแปรอิสระที่มีขนาด n x (p+q)

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของสัมประสิทธิ์ความถดถอยที่มีขนาด (p+q) x 1

$\underset{\sim}{\varepsilon}$ คือ เวกเตอร์ของค่าความคลาดเคลื่อนที่มีขนาด n x 1

$$\underset{\sim}{y} = \begin{array}{cc} \begin{bmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \\ \\ \underset{\sim}{y}_p \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \\ \\ n_p \end{array} \end{array}$$

$$X = \begin{array}{cc} \begin{bmatrix} \underset{\sim}{1} & \underset{\sim}{1} & \underset{\sim}{0} & \underset{\sim}{0} \ldots (\underset{\sim}{x}_{11} - \bar{x}_{..1}\underset{\sim}{1}) & (\underset{\sim}{x}_{12} - \bar{x}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{x}_{1q} - \bar{x}_{..q}\underset{\sim}{1}) \\ \underset{\sim}{1} & \underset{\sim}{0} & \underset{\sim}{1} & \underset{\sim}{0} \ldots (\underset{\sim}{x}_{21} - \bar{x}_{..1}\underset{\sim}{1}) & (\underset{\sim}{x}_{22} - \bar{x}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{x}_{2q} - \bar{x}_{..q}\underset{\sim}{1}) \\ \underset{\sim}{1} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{1} \ldots (\underset{\sim}{x}_{31} - \bar{x}_{..1}\underset{\sim}{1}) & (\underset{\sim}{x}_{32} - \bar{x}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{x}_{3q} - \bar{x}_{..q}\underset{\sim}{1}) \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \underset{\sim}{1} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \ldots (\underset{\sim}{x}_{p1} - \bar{x}_{..1}\underset{\sim}{1}) & (\underset{\sim}{x}_{p2} - \bar{x}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{x}_{pq} - \bar{x}_{..q}\underset{\sim}{1}) \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \\ n_3 \\ \cdot \\ \cdot \\ \cdot \\ n_p \end{array} \end{array}$$

$$\underset{\sim}{\beta} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \cdot \\ \cdot \\ \alpha_{p-1} \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_q \end{bmatrix}$$

$$\underset{\sim}{\varepsilon} = \begin{array}{cc} \begin{bmatrix} \underset{\sim}{\varepsilon}_1 \\ \underset{\sim}{\varepsilon}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{\varepsilon}_p \end{bmatrix} & \begin{array}{c} n_1 \\ n_2 \\ \cdot \\ \cdot \\ \cdot \\ n_p \end{array} \end{array}$$

ตัวแบบลดรูปคือ

$$y_{ij} = \alpha + \sum_{k=1}^{q}\beta_k(x_{ijk} - \overline{x}_{..k}) + \varepsilon_{ij}$$

เมื่อเขียนให้อยู่ในรูปเมตริกซ์จะได้

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

เมื่อ $\underset{\sim}{y}$ คือ เวกเตอร์ของค่าสังเกตหรือตัวแปรตาม ที่มีขนาด $n \times 1$

X คือ เมตริกซ์ของตัวแปรอิสระที่มีขนาด $n \times (q+1)$

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของสัมประสิทธิ์ความถดถอยที่มีขนาด $(q+1) \times 1$

$\underset{\sim}{\varepsilon}$. คือ เวกเตอร์ของค่าความคลาดเคลื่อนที่มีขนาด n x 1

$$X = \begin{array}{c} \left[\begin{array}{l} \underset{\sim}{1} \;\; (\underset{\sim}{X}_{11}-\bar{X}_{..1}\underset{\sim}{1}) \;\; (\underset{\sim}{X}_{12}-\bar{X}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{X}_{1q}-\bar{X}_{..q}\underset{\sim}{1})_1 \\ \underset{\sim}{1} \;\; (\underset{\sim}{X}_{21}-\bar{X}_{..1}\underset{\sim}{1}) \;\; (\underset{\sim}{X}_{22}-\bar{X}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{X}_{2q}-\bar{X}_{..q}\underset{\sim}{1})_2 \\ \underset{\sim}{1} \;\; (\underset{\sim}{X}_{31}-\bar{X}_{..1}\underset{\sim}{1}) \;\; (\underset{\sim}{X}_{32}-\bar{X}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{X}_{3q}-\bar{X}_{..q}\underset{\sim}{1})_3 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{1} \;\; (\underset{\sim}{X}_{p1}-\bar{X}_{..1}\underset{\sim}{1}) \;\; (\underset{\sim}{X}_{p2}-\bar{X}_{..2}\underset{\sim}{1}) \ldots (\underset{\sim}{X}_{pq}-\bar{X}_{..q}\underset{\sim}{1})_p \end{array}\right] \end{array} \begin{array}{l} n_1 \\ n_2 \\ n_3 \\ \cdot \\ \cdot \\ \cdot \\ n_p \end{array}$$

$$\underset{\sim}{\beta} = \left[\begin{array}{c} \alpha \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_q \end{array}\right]$$

$$\underset{\sim}{\varepsilon} = \left[\begin{array}{c} \underset{\sim}{\varepsilon}_1 \\ \underset{\sim}{\varepsilon}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{\varepsilon}_p \end{array}\right] \begin{array}{l} n_1 \\ n_2 \\ \\ \\ \\ n_p \end{array}$$

ตารางที่ 2.1 แสดงข้อมูลของการวิเคราะห์ความแปรปรวนร่วมของแผนแบบการทดลองแบบสุ่มตลอด

วิธีปฏิบัติที่ 1			วิธีปฏิบัติที่ 2			วิธีปฏิบัติที่ p		
y	x_1	$x_2 \ldots x_q$	y	x_1	$x_2 \ldots x_q$	y	x_1	$x_2 \ldots x_q$
y_{11}	x_{111}	$x_{112} \ldots x_{11q}$	y_{12}	x_{121}	$x_{122} \ldots x_{12q}$	y_{1p}	x_{1p1}	$x_{1p2} \ldots x_{1pq}$
y_{21}	x_{211}	$x_{212} \ldots x_{21q}$	y_{22}	x_{221}	$x_{222} \ldots x_{22q}$	y_{2p}	x_{2p1}	$x_{2p2} \ldots x_{2pq}$
y_{31}	x_{311}	$x_{312} \ldots x_{31q}$	y_{31}	x_{321}	$x_{322} \ldots x_{32q}$	y_{3p}	x_{3p1}	$x_{3p2} \ldots x_{3pq}$
.	.	. .	.	.	. .	.	.	. .
.	.	. .	.	.	. .	.	.	. .
.	.	. .	.	.	. .	.	.	. .
y_{n1}	x_{n11}	$x_{n12} \ldots x_{n1q}$	y_{n1}	x_{n21}	$x_{n22} \ldots x_{n2q}$	y_{np}	x_{np1}	$x_{np2} \ldots x_{npq}$

2.2 วิธีการประมาณพารามิเตอร์

2.2.1 วิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method)

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 $\hat{\underset{\sim}{\beta}}$ ซึ่งทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (sum square error) หรือ SSE มีค่าน้อยที่สุด

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

และการหาค่าน้อยที่สุดของผลบวกกำลังสองของความคลาดเคลื่อนทำได้โดยการหาอนุพันธ์ (differentiate) เทียบกับ $\hat{\beta}_i$ แล้วกำหนดให้เท่ากับ 0

$$
\frac{\partial\, SSE}{\partial\, \hat{\underset{\sim}{\beta}}} = \begin{bmatrix} \dfrac{\partial\, SSE}{\partial \hat{\beta}_1} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial\, SSE}{\partial \hat{\beta}_p} \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} = \underset{\sim}{0}
$$

$$
\begin{aligned}
-2X'\underset{\sim}{y} + 2X'X\hat{\underset{\sim}{\beta}} &= 0 \\
(X'X)\hat{\underset{\sim}{\beta}} &= X'\underset{\sim}{y}
\end{aligned}
$$

ดังนั้น

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

2.2.2 <u>การประมาณพารามิเตอร์ด้วยวิธีตัวประมาณเอ็ม</u>

ปี ค.ศ. 1964 P.J. Huber ได้ศึกษาฟังก์ชันความคลาดเคลื่อนซึ่งเรียกว่าตัวประมาณชนิดเอ็ม(M-estimator) ซึ่งตัวประมาณนี้จะประมาณพารามิเตอร์ของค่าน้อยที่สุดของ $\sum_{i=1}^{n}\rho(\varepsilon_i/s)$ เมื่อ ε_i เป็นค่าคลาดเคลื่อนของค่าสังเกตที่ i และ s เป็นตัวประมาณของการกระจายของตัวอย่าง

(1) $$\min_{\underset{\sim}{\beta}}\sum_{i=1}^{n}\rho(\varepsilon_i/s) = \min_{\underset{\sim}{\beta}}\sum_{i=1}^{n}\rho\left|(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s\right| \quad , \; X_i = \underset{\sim}{X}_i'$$

เมื่อ S เป็นตัวประมาณที่แกร่งของสเกล ซึ่งนิยมใช้มัธยฐานของค่าสัมบูรณ์ของความเบี่ยงเบน(Median Absolute Deviation) และทำการปรับด้วยค่าคงที่ 0.6745 ซึ่งจะทำให้ S เป็นตัวประมาณที่ไม่เอนเอียงของ เมื่อ n มีขนาดใหญ่ และการแจกแจงของค่าคลาดเคลื่อนเป็นแบบปกติ ดังนั้น S เขียนได้เป็น

$$S = \frac{\text{median}\left|\varepsilon_i - \text{median}\ (\varepsilon_i)\right|}{0.6745}$$

จากสมการ (1) หาอนุพันธ์ของ ρ เทียบกับ β_j แล้วกำหนดให้ = 0 โดย $\Psi = \rho'$

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

โดยที่ $m = p+q-1$ ในกรณีของตัวแบบเต็มรูป

$m = q$ ในกรณีของตัวแบบลดรูป

วิธีการที่ใช้ประมาณค่าสัมประสิทธิ์การถดถอยส่วนใหญ่จะใช้วิธีกำลังสองน้อยที่สุดซึ่งถ่วงน้ำหนักซ้ำหลายรอบ (iteratively reweighted least square) จากสมการ (2) สามารถเขียนสมการ m สมการได้ดังนี้

$$\sum_{i=1}^{n} X_{ij}\Psi[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s] = \frac{\sum_{i=1}^{n} X_{ij}\Psi[(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s] \ .\ (y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s}{(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta})/s} = 0$$

$$\sum_{i=1}^{n} X_{ij}W_{i0}(y_i - \underset{\sim}{X}_i'\underset{\sim}{\beta}) = 0 \quad , \quad j=1,2,\ldots,m$$

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

และเขียนให้อยู่ในรูปเมตริกซ์ได้เป็น

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

เมื่อ W เป็นเมตริกซ์ขนาด nxn และมีสมาชิกตามเส้นทแยงมุมเป็น $W_{1o}, W_{2o}, \ldots, W_{no}$

$$W = \begin{bmatrix} W_{1o} & 0 \ldots & 0 \\ 0 & W_{2o} & . \\ . & . & . \\ . & . & . \\ 0 & \ldots & W_{no} \end{bmatrix}$$

ในการประมาณที่แกร่งจำเป็นอย่างยิ่งต้องเลือกค่าเริ่มต้นของ β_j อย่างระมัดระวัง ซึ่งการใช้ $\hat{\beta}_j$ จากวิธีกำลังสองน้อยที่สุดจะให้ผลดี ในที่นี้จึงใช้ $\underset{\sim}{\hat{\beta}}_{OLS}$ มาคำนวณค่า W_i เมื่อได้ค่า W_i แล้วก็นำไปหาค่า $\underset{\sim}{\beta}_j$ ใหม่โดยจะกระทำซ้ำๆจนกระทั่งได้ค่า $\hat{\underset{\sim}{\beta}}$ ที่ค่อนข้างคงที่ ดังนั้นตัวประมาณ $\hat{\underset{\sim}{\beta}}_m$ จึงหาได้จากสมการ

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

จากการศึกษาของ ปราณี รัตนัง (พ.ศ. 2530) ในเรื่องการประมาณสัมประสิทธิ์การถดถอยพหุเมื่อความคลาดเคลื่อนมีการแจกแจงแบบหางยาวกว่าปกติ พบว่าสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปนจะมีอิทธิพลจากมากไปหาน้อยที่ทำให้วิธีของตัวประมาณเอ็มโดย

เกณฑ์ความแกร่งของ Ramsay ดีกว่าวิธีกำลังสองน้อยที่สุด ดังนั้นจึงเป็นที่สงสัยว่าถ้าเป็นเรื่องของการวิเคราะห์ความแปรปรวนร่วมแล้ว วิธีของตัวประมาณเอ็มโดยใช้เกณฑ์ความแกร่งของ Ramsay และตัวประมาณสเกลความคลาดเคลื่อนแบบ MAD จะยังคงให้ประสิทธิภาพการวิเคราะห์ดีอยู่หรือไม่ และจากการศึกษาของ Jeffrey B. and Raymond H.Myere ใน Robust Analysis of Covariance พบว่าเมื่อความคลาดเคลื่อนมีการแจกแจงแบบหางยาวกว่าปกติ วิธีการของตัวประมาณเอ็มโดยใช้เกณฑ์ความแกร่งของ Huber และ Beaton จะทำให้เกิดความน่าจะเป็นของความความผิดพลาดประเภทที่ 1 ต่ำกว่าวิธีกำลังสองน้อยที่สุด เมื่อเปอร์เซนต์การปลอมปนเพิ่มขึ้น ด้วยเหตุนี้เองการวิจัยครั้งนี้จึงใช้เกณฑ์ความแกร่งของ Ramsay ซึ่งจะมีฟังก์ชันของ ρ และ Ψ_i คือ

$$\rho \quad = \quad a^{-2}\left[1-\exp(-a|\varepsilon_i|/s).(1+a|\varepsilon_i|)/s\right]$$

$$\Psi i \quad = \left(\varepsilon_i/s\right)\exp(-a|\varepsilon_i|/s\ )$$

โดยฟังก์ชัน Ψi จะมีลักษณะเป็นฟังก์ชันเอ็กซ์โปเนนเชียลที่ลาดลงเรื่อยๆ และจะลู่เข้าหาค่าศูนย์เมื่อ ε_i มีขนาดใหญ่ และจะใช้ a = 0.3 เป็นค่าคงที่ในการกำหนดขอบเขตฟังก์ชันโดยที่ขอบเขตคือ $|\varepsilon_i|/s \ = 1/a$ เมื่อพิจารณา Ψi พบว่า ถ้าค่าผิดพลาดมีค่ามากหรือมีค่าสังเกตที่ผิดปกติมาก ๆ ค่าเหล่านี้จะมีอิทธิพลลดลงเรื่อย ๆ

2.3 <u>ผลงานวิจัยที่เกี่ยวข้อง</u>

ผู้ที่ศึกษาวิจัยประสิทธิภาพการวิเคราะห์ความแปรปรวนร่วมเมื่อข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นนั้นมีน้อย แต่ในลักษณะของการประมาณพารามิเตอร์ในรูปแบบของการวิเคราะห์การถดถอยเชิงเส้นนั้น ได้มีผู้ทำการศึกษาไว้หลายท่านดังนี้

1. ปี ค.ศ. 1964 พี.เจ. ฮิวเบอร์ (P.J. Huber) ได้ศึกษากับฟังก์ชันของความคลาดเคลื่อน ซึ่งเรียกว่า ตัวประมาณเอ็ม (M-estimator) โดยที่ตัวประมาณเอ็มจะประมาณพารามิเตอร์ของ $\sum_{i=1}^{n}\rho(\varepsilon_i/s)$ เมื่อ ε_i เป็นค่าคลาดเคลื่อนของค่าสังเกตตัวที่ i

และ s เป็นตัวประมาณที่เหมาะสมของการกระจายของตัวอย่าง ต่อมาปี ค.ศ.1977 แรมเซย์ (Ramsay) ได้สร้างฟังก์ชันแรมเซย์ (Ramsay Ea) ขึ้นมีลักษณะเป็นฟังก์ชันเอ็กซ์โปเนนเชียล โดยมีขอบเขตที่ $|\varepsilon_i/s|=1/a$ ซึ่งฟังก์ชันของ แรมเซย์ จะค่อย ๆ ลาดลงนั่นคือ ลู่เข้าหาศูนย์เมื่อ $|\varepsilon_i|$ มีขนาดใหญ่ขึ้นเรื่อย ๆ ในปี ค.ศ.1971 แฮมเพล (Hampel)ได้สร้างฟังก์ชันแฮมเพลและปี ค.ศ. 1972 แอนดรูว์ (Andrews) ได้สร้างฟังก์ชันคลื่นของแอนดรูว์โดยฟังก์ชันทั้งสองจะเป็นแบบคล้อยลาดลงอย่างรวดเร็ว ลักษณะฟังก์ชันแอนดรูว์ เป็นคลื่นแบบ Sine และฟังก์ชันแฮมเพล Hampel' s function) มีจุดเปลี่ยนแปลงขอบเขตของฟังก์ชัน 3 จุด คือ a, b, c เมื่อ a = 1.7, b = 3.4 และ c = 8.5

2. ปี พ.ศ. 2530 ปราณี รัตนัง ได้ศึกษาเปรียบเทียบวิธีการประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงหางยาวกว่าปกติและแบบเบ้ ระหว่างวิธีกำลังสองน้อยที่สุดและวิธีตัวประมาณเอ็มที่ใช้เกณฑ์ความแกร่งของ Ramsay พบว่าสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปน จะมีอิทธิพลจากมากไปหาน้อยที่ทำให้วิธีตัวประมาณเอ็มดีกว่าวิธีกำลังสองน้อยที่สุด

3. ปี พ.ศ. 2532 ทรงพันธ์ ชุณหสวัสดิกุล ได้ศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อค่าประมาณสเกลเปลี่ยนไป โดยเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธีตัวประมาณเอ็ม ซึ่งใช้เกณฑ์ความแกร่งของ Huber และตัวประมาณสเกล วิธี STD (The Standard - Deviation of Location), วิธี MAD (The Median Absolute Deviation)และวิธี MBA (The Modified Biewight A estimator) พบว่าเมื่อการแจกแจงของความคลาดเคลื่อนเป็นแบบหางยาว สเกลแฟคเตอร์ และเปอร์เซ็นต์การปลอมปน เรียงลำดับจากมากไปน้อยที่ทำให้วิธี MAD เป็นตัวประมาณสเกลซึ่งใช้เกณฑ์ความแกร่งของ Huber เป็นวิธีการประมาณค่าที่ดีที่สุด