บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้สิ่งที่สนใจศึกษาคือการเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์ที่ได้จากการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่ายด้วยวิธีกำลังสองน้อยที่สุด และวิธีค่าสัมบูรณ์ต่ำสุดเมื่อค่าคลาดเคลื่อนสุ่มในสมการถดถอยมีการแจกแจงแบบปกติ สม่ำเสมอ และแบบหางยาว ในที่นี้จะศึกษาเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการพยากรณ์ของวิธีกำลังสองน้อยที่สุด และของวิธีค่าสัมบูรณ์ต่ำสุด โดยจะกล่าวถึงในรายละเอียดในหัวข้อต่อไปนี้

2.1 การประมาณค่าพารามิเตอร์

ในการวิจัยครั้งนี้ได้ใช้วิธีการประมาณค่าพารามิเตอร์ 2 วิธีการคือ วิธีกำลังสองน้อยที่สุด และวิธีการค่าสัมบูรณ์ต่ำสุด ซึ่งรายละเอียดของแต่ละวิธีเป็นดังนี้

2.1.1 วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method)

วิธีการหาตัวประมาณค่าพารามิเตอร์วิธีนี้ เป็นวิธีที่มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of Linear Estimation) โดยมีหลักเกณฑ์ดังนี้ คือ ทำให้ผลบวกกำลังสองของความคลาดเคลื่อน $\Sigma \varepsilon_t^2$ มีค่าน้อยที่สุด ภายใต้ข้อสมมติเบื้องต้นของสมการถดถอยดังนี้

1. ค่าคลาดเคลื่อน (ε_t) เป็นตัวแปรสุ่ม
2. $E(\varepsilon_t) = 0$ (Zero mean)
3. $E(\varepsilon_s \varepsilon_t) = \sigma^2$ เมื่อ $s = t$ (Homoscedasticity)
4. $E(\varepsilon_s \varepsilon_t) = 0$ เมื่อ $s \neq t$ (Nonautocorrelation)
5. ตัวแปรอิสระ x เป็นค่าคงที่ (Nonstochastic)

จากหลักเกณฑ์ดังกล่าวข้างต้น จะทำการหาตัวประมาณกำลังสองต่ำสุด $\hat{\beta}_1$ และ $\hat{\beta}_2$ ของ β_1 และ β_2 ตามลำดับ จากสมการต่อไปนี้

$$\frac{\partial}{\partial \beta_1} \Sigma(y_t-\beta_1-\beta_2 x_t)^2 = \Sigma y_t - n\beta_1 - \beta_2 \Sigma x_t = 0$$

$$\frac{\partial}{\partial \beta_2} \Sigma(y_t-\beta_1-\beta_2 x_t)^2 = \Sigma x_t y_t - \beta_1 \Sigma x_t - \beta_2 \Sigma x_t^2 = 0$$

ได้

$$\beta_1 = \hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

$$\beta_2 = \hat{\beta}_2 = \frac{\Sigma x_t y_t - (\Sigma x_t \; \Sigma y_t)/n}{\Sigma x_t^2 - (\Sigma x_t)^2/n}$$

2.1.2 วิธีค่าสัมบูรณ์ต่ำสุด

วิธีค่าสัมบูรณ์ต่ำสุด สำหรับงานวิจัยนี้ได้ศึกษาโดยใช้เทคนิคโปรแกรมเชิงเส้น (Linear Programming Techniques) คำนวณหาค่าพารามิเตอร์ ($\hat{\beta}$) โดยมีหลักการทำให้ผลรวมของค่าสัมบูรณ์ของค่าคลาดเคลื่อนมีค่าต่ำสุด

$$\text{Minimize } \Sigma |\varepsilon_t| \qquad \text{------(1)}$$

ด้วยเงื่อนไข $\varepsilon_t = y_t-\beta_1-\beta_2 x_t \quad t = 1,2,\ldots,n$

จาก (1) แปลงให้เป็นรูปโปรแกรมเชิงเส้น ดังนี้

กำหนดให้ $z = \Sigma |\varepsilon_t|$

เนื่องจาก ε_t ไม่มีข้อจำกัดด้านเครื่องหมาย ได้ว่า

$$\varepsilon_t = \varepsilon_t^+ - \varepsilon_t^- , \quad \varepsilon_t^+ , \varepsilon_t^- \geq 0 \qquad \text{-------(2)}$$

นิยาม

$$\varepsilon_t^+ = \begin{cases} \varepsilon_t & , \ \varepsilon_t \geq 0 \\ 0 & , \ \varepsilon_t < 0 \end{cases}$$

--------(3)

$$\varepsilon_t^- = \begin{cases} 0 & , \ \varepsilon_t \geq 0 \\ -\varepsilon_t & , \ \varepsilon_t < 0 \end{cases}$$

เพราะฉะนั้น ได้ว่า $\varepsilon_t^+ . \varepsilon_t^- = 0$ --------(4)

นั่นคือ อย่างน้อย 1 ตัวแปรใน ε_t^+ และ ε_t^- เท่ากับ 0 เสมอ ดังนั้น

$$|\varepsilon_t| = \varepsilon_t^+ + \varepsilon_t^-$$

และให้ $u_t = \varepsilon_t^+$ และ $v_t = \varepsilon_t^-$

เนื่องจากพารามิเตอร์ β_1 และ β_2 ไม่มีข้อจำกัดด้านเครื่องหมาย

จึงเขียนใหม่เป็น $\beta_1 = \beta_1^+ - \beta_1^-$

$$\beta_2 = \beta_2^+ - \beta_2^-$$

โดยที่ $\beta_1^+ , \ \beta_1^- , \ \beta_2^+ , \ \beta_2^- \geq 0$

สามารถแปลงปัญหา (1) ได้เป็น

$$\text{Minimize } \Sigma(u_t + v_t)$$

เงื่อนไข : $(\beta_1^+ - \beta_1^-) + (\beta_2^+ - \beta_2^-) + u_t - v_t = y_t \ , \quad t = 1,\ldots,n$

$$u_t \ , v_t \geq 0$$

$$\beta_1^+ , \ \beta_1^- , \ \beta_2^+ , \ \beta_2^- \geq 0$$

สามารถเขียนสมการซิมเพล็กซ์ ได้ดังนี้

ยกตัวอย่าง เมื่อ n = 5

$$\text{Minimize } Z = u_1+u_2+u_3+u_4+u_5+v_1+v_2+v_3+v_4+v_5+0(\beta_1^+-\beta_1^-)$$
$$+0(\beta_2^+-\beta_2^-)$$

เงื่อนไข :
$$(\beta_1^+ - \beta_1^-) + (\beta_2^+ - \beta_2^-)x_1 + u_1 - v_1 = y_1$$
$$(\beta_1^+ - \beta_1^-) + (\beta_2^+ - \beta_2^-)x_2 + u_2 - v_2 = y_2$$
$$(\beta_1^+ - \beta_1^-) + (\beta_2^+ - \beta_2^-)x_3 + u_3 - v_3 = y_3$$
$$(\beta_1^+ - \beta_1^-) + (\beta_2^+ - \beta_2^-)x_4 + u_4 - v_4 = y_4$$
$$(\beta_1^+ - \beta_1^-) + (\beta_2^+ - \beta_2^-)x_5 + u_5 - v_5 = y_5$$
$$u_t , v_t \geq 0$$
$$\beta_1^+ , \beta_1^- , \beta_2^+ , \beta_2^- \geq 0$$

2.2 ผลงานวิจัยที่เกี่ยวข้อง

ผู้ศึกษาเกี่ยวกับตัวประมาณพารามิเตอร์ที่แกร่งในสมการถดถอยเชิงเส้นนั้น ได้ทำการศึกษาแตกต่างกันออกไป ซึ่งวิธีการที่นำมาใช้มีทั้งเป็นเชิงเส้น (linear) และไม่เป็นเชิงเส้น (nonlinear) ในหัวข้อนี้จะเสนอวิธีที่น่าสนใจพร้อมทั้งข้อสรุปต่างๆ ดังนี้คือ

การศึกษาตัวประมาณที่แกร่งของสัมประสิทธิ์การถดถอยเชิงเส้นแบบพหุ โดยวิธีหาค่าน้อยที่สุดของส่วนเบี่ยงเบนยกกำลังที่ p (p^{th} Power Deviation หรือ Lp-norm regression) ศึกษาโดย Forsythe (ค.ศ. 1972 :159-166) โดยพารามิเตอร์ของตัวแบบถูกเลือกเพื่อหาค่าน้อยที่สุดของ $\Sigma|\varepsilon_i|^p$; $(1<p<2)$ สำหรับคำนวณหาสัมประสิทธิ์การถดถอยพหุ เมื่อ $1<p<2$ จะทำเป็นโปรแกรมไม่เป็นเชิงเส้น Forsythe ได้ศึกษากระบวนการนี้สำหรับตัวแบบถดถอยเชิงเส้นตรงด้วยการจำลองแบบมอนติคาร์โล โดยใช้การแจกแจงปกติปลอมปนหลายแบบ เขาสังเกตว่า เมื่อค่าผิดพลาดมีการแจกแจงแบบหางยาววิธี Lp-norm ที่ $p = 1.5$ จะใช้ประมาณค่าสัมประสิทธิ์การถดถอยพหุได้ดีกว่าวิธีกำลังสองน้อยที่สุด และเมื่อการแจกแจงของความคลาดเคลื่อนเป็นแบบปกติวิธี Lp-norm ที่ $p = 1.5$ จะให้ผลใกล้เคียงกับวิธีกำลังสองน้อยที่สุด