บทที่ 2

สถิติที่ใช้ในการวิจัย

การวิจัยครั้งนี้สนใจศึกษา การประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลาเมื่อข้อมูลมีค่าผิดปกติ ในบทนี้จะกล่าวถึง ลักษณะทั่วไปของอนุกรมเวลา และจะกล่าวถึงรายละเอียดของวิธีการประมาณค่าพารามิเตอร์เป็นลำดับต่อไป

อนุกรมเวลาคงที่

ในการวิจัยครั้งนี้ข้อมูลอนุกรมเวลาสามารถเขียนในรูปแบบทั่วไปได้ดังนี้

สมมุติให้ $z_1, z_2, \ldots, z_t$ คืออนุกรมเวลา

$a_1, a_2, \ldots, a_t$ คือค่าผิดพลาดสุ่ม ที่มีการแจกแจงแบบปกติปลอมปน

รูปแบบทั่วไปของอนุกรมเวลา ARMA(p,q) โดยอนุกรมเวลานี้เป็นอนุกรมเวลาคงที่คือ

$$z_t = K + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

โดยที่ $\phi_1, \ldots, \phi_p$ คือสัมประสิทธิ์การถดถอย

$\theta_1, \ldots, \theta_q$ คือสัมประสิทธิ์ค่าเฉลี่ยเคลื่อนที่

K คือเทอมคงที่

เรียกสมการข้างต้นว่า ตัวแบบค่าเฉลี่ยเคลื่อนที่และการถดถอย อันดับที่ p,q

โดยที่ p คืออันดับของรูปแบบการถดถอย (Autoregressive Model)

q คืออันดับของรูปแบบค่าเฉลี่ยเคลื่อนที่ (Moving-Average Model)

ซึ่งเป็นรูปแบบคงที่ ที่ศึกษาในครั้งนี้มี 2 รูปแบบคือ

1. ตัวแบบอัตตสัมพันธ์อันดับที่หนึ่ง หรือ AR(1) มีสมการคือ

$$z_t = K + \phi_1 z_{t-1} + a_t \quad \text{โดยที่} \ |\phi_1| < 1$$

2. ตัวแบบค่าเฉลี่ยเคลื่อนที่อันดับที่หนึ่ง หรือ MA(1) มีสมการคือ

$$z_t = K + a_t - \theta_1 a_{t-1} \quad \text{โดยที่} \ |\theta_1| < 1$$

การวิจัยครั้งนี้ศึกษากรณีที่ $K = 0$

วิธีการประมาณค่าพารามิเตอร์

สำหรับการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของสมการการพยากรณ์ในครั้งนี้ได้ทำการศึกษา 3 วิธีดังนี้

1. วิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล

หลักการประมาณค่าโดยวิธีนี้คือการทำให้ผลรวมค่ากำลังสองของความคลาดเคลื่อนภายใต้พารามิเตอร์ของตัวแบบมีค่าต่ำที่สุด

จากรูปแบบ ARMA(p,q) สมการแสดงผลรวมค่ากำลังสองของความคลาดเคลื่อนของขบวนการนี้คือ

$$s(\phi,\theta) = \sum_{t=-\infty}^{n} [a_t|(\phi,\theta)]^2$$

1.1 เมื่อ z_t เป็นอนุกรมเวลาแบบ AR(1)

ตัวประมาณกำลังสองน้อยที่สุด (Ordinary Least Square Estimator: OLS) คือตัวประมาณที่ทำให้ฟังก์ชัน $s(\phi)$ มีค่าต่ำที่สุด

$$\begin{aligned} s(\phi) &= \sum_{t=-\infty}^{n} \left[a_t|\phi\right]^2 \\ &= \sum_{t=-\infty}^{n} \left[z_t - \phi_1 z_{t-1}|\phi\right] \\ \frac{d\,s(\phi)}{d\,\phi_1} &= \sum_{t=-\infty}^{n} \left[z_t z_{t-1} - \phi_1 z_{t-1}^2\right] \\ &= 0 \end{aligned}$$

แก้สมการได้

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n} z_t z_{t-1}}{\sum_{t=3}^{n} z_{t-1}^2}$$

1.2 เมื่อ z_t เป็นอนุกรมเวลาแบบ MA(1)

ตัวประมาณกำลังสองน้อยที่สุด คือตัวประมาณที่ทำให้ฟังก์ชั่น $s(\theta)$ มีค่าต่ำที่สุด

$$s(\theta) = \sum_{t=-\infty}^{n} \left[a_t|\theta\right]^2$$

$$= \sum_{t=-\infty}^{n} \left[z_t + \theta_1 a_{t-1}|\theta\right]^2$$

$$\frac{ds(\theta)}{d\theta_1} = \sum_{t=-\infty}^{n} \left[z_t a_{t-1} + \theta_1 a_{t-1}^2\right]$$

$$= 0$$

แก้สมการได้

$$\hat{\theta}_1 = -\frac{\sum_{t=2}^{n} z_t a_{t-1}}{\sum_{t=3}^{n} a_{t-1}^2}$$

การหาค่าประมาณพารามิเตอร์ของอนุกรมนี้ใช้ค่า a_t ซึ่งไม่ทราบค่า ดังนั้นจึงต้องทำการประมาณค่า a_t และวิธีหนึ่งในการประมาณค่าคือวิธีการพยากรณ์ย้อนหลัง (Backforecasting) รายละเอียดแสดงไว้ใน ภาคผนวก ก.

2. วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว

หลักการประมาณค่าโดยวิธีนี้คือการ ทำให้ค่าของสมการภาวะน่าจะเป็นภายใต้พารามิเตอร์ของตัวแบบมีค่าสูงที่สุด

จากรูปแบบARMA(p,q)สมการภาวะน่าจะเป็นของขบวนการนี้คือ

$$L(\phi,\theta,\sigma_a|z) = (2\Pi\sigma_a^2)^{-n/2}|M_n^{(p,q)}|^{1/2} \exp\{-[s(\phi,\theta)]/[2\sigma_a^2]\}$$

เมื่อ

$$s(\phi,\theta) = \sum_{t=-\infty}^{n}[a_t|z,\phi,\theta]^2$$

2.1 เมื่อ z_t เป็นอนุกรมเวลาแบบ AR(1)

ตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator:MLE) คือตัวประมาณที่ทำให้สมการภาวะน่าจะเป็น $L(\phi,\sigma_a|z)$ มีค่าสูงที่สุด

$$\underset{\phi,\sigma_a}{\text{maximize}}\ L(\phi,\sigma_a|z) = (2\Pi\sigma_a^2)^{-n/2}|M_n^1|^{1/2} \exp\{-[s(\phi)]/[2\sigma_a^2]\}$$

เมื่อ

$$|M_n^1| = 1 - \phi_1^2$$

$$s(\phi) = (1-\phi_1^2)z_1^2 + \sum_{t=2}^{n}(z_t - \phi_1 z_{t-1})^2$$

2.2 เมื่อ z_t เป็นอนุกรมเวลาแบบ MA(1)

ตัวประมาณภาวะน่าจะเป็นสูงสุด คือตัวประมาณที่ทำให้สมการภาวะ-น่าจะเป็น $L(\theta,\sigma_a|z)$ มีค่าสูงที่สุด

$$\underset{\theta,\sigma_a}{\text{maximize}}\ L(\theta,\sigma_a|z) = (2\Pi\sigma_a^2)^{-n/2}|M_n^1|^{1/2} \exp\{-[s(\theta)]/[2\sigma_a^2]\}$$

เมื่อ

$$|M_n^1| = (1-\theta_1^2)/(1-\theta_1^{2(n+1)})$$

$$s(\theta) = \sum_{t=0}^{n}E^2(a_t|z)$$

หลังจากได้ค่าประมาณของพารามิเตอร์ต่างๆแล้ว จึงนำค่าที่ได้ไปเป็นค่าประมาณเริ่มต้นสำหรับการปรับปรุงข้อมูลในขบวนการทำซ้ำจนกระทั่งได้ค่าประมาณของพารามิเตอร์คงที่

การปรับปรุงข้อมูล

ข้อมูลผิดปกติที่เกิดขึ้นในอนุกรมเวลา จะมีผลทำให้การประมาณค่าพารามิเตอร์ของอนุกรมมีความเอนเอียงสูง โดยเฉพาะค่าประมาณของ σ_a^2 จะมีค่ามากกว่าจริง(Over Estimate) การแก้ไขข้อมูลผิดปกติที่เกิดในอนุกรม ต้องอาศัยขบวนการทำซ้ำซึ่งเสนอโดย Chang และ Tioa(1983) , Chang , Tiao และ Chen(1988) , Chen และ Liu(1993)

วิธีการปรับปรุงข้อมูลมีหลักการคือ

กำหนดให้ $\{Y_t\}$เป็นอนุกรมเวลา ARMA(p,q) ซึ่งมีรูปแบบ

$$Y_t = \frac{\theta(B)}{\phi(B)} a_t \qquad t = 1....,n$$

และกำหนดรูปแบบของอนุกรมเวลาที่มีอิทธิพลจากปัจจัยภายนอกมากระทบ(อนุกรมมีค่าผิดปกติ) m ครั้ง ณ. เวลา $t_1, t_2, ..., t_m$ คือ

$$Y_t' = Y_t + \sum_{j=1}^{m} \omega_j I_t(t_j)$$

เมื่อ

$$I_t(t_j) = \begin{cases} 1 & , t = t_j \\ 0 & , t \neq t_j \end{cases}$$

จากรูปแบบข้างต้นจะเห็นได้ว่านอกจากการประมาณค่าพารามิเตอร์ของอนุกรมแล้วจะต้องทำการประมาณค่าของข้อมูลผิดปกติด้วยซึ่งค่าประมาณของข้อมูลผิดปกติจะแตกต่างกันไปสำหรับแต่ละรูปแบบของอนุกรม นอกจากการประมาณค่าผิดปกติแล้ว Chen และ Liu ยังได้เสนอขั้นตอนการตรวจสอบข้อมูลผิดปกติโดยใช้ตัวสถิติ $\hat{\tau}(t)$ [1] ในการตรวจสอบความผิดปกติของข้อมูล

---

[1] Chen,C.,and Lui,L. Journal of the American Statistical Association , March 1993,Vol 88 , No. 421 ,pp.284-297

ขั้นตอนในการปรับปรุงแบ่งเป็นขั้นตอนใหญ่ 2 ขั้นตอนคือ

1. ประมาณค่าพารามิเตอร์เริ่มต้นและตรวจสอบค่าผิดปกติ

1.1 หลังจากการประมาณค่าพารามิเตอร์ของอนุกรมโดยวิธีตัวประมาณภาวะน่าจะเป็นสูงสุดแล้ว ให้คำนวณค่าความคลาดเคลื่อนจาก

$$\hat{e}_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} z_t$$

เมื่อ

$$\hat{\phi}(B) = (1 - \hat{\phi}_1 B - \dots - \hat{\phi}_p B^p)$$

$$\hat{\theta}(B) = (1 - \hat{\theta}_1 B - \dots - \hat{\theta}_q B^q)$$

และในกรณีนี้ $p = 1$

$q = 1$

ซึ่งจะได้ค่าประมาณความคลาดเคลื่อนสำหรับรูปแบบ AR(1) และ MA(1) คือ

$$\text{AR(1)}: \hat{e}_t = (1 - \hat{\phi}_1 B) z_t$$

$$\text{MA(1)}: \hat{e}_t = \frac{1}{(1 - \hat{\theta}_1)} z_t$$

ค่าประมาณของความแปรปรวนคือ

$$\hat{\sigma}_a^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{e}_t^2$$

1.2 สำหรับ $t = 1, \dots, n$ คำนวณค่า $\hat{\tau}(t)$ โดยที่

$$\hat{\tau}(t) = \frac{\hat{\omega}_j}{\left[\mathrm{Var}(\hat{\omega})\right]^{(1/2)}}$$

เมื่อ $\hat{\omega}_j$ คือค่าประมาณของค่าผิดปกติ

$\mathrm{Var}(\hat{\omega})$ คือความแปรปรวนของการประมาณค่าผิดปกติ

ค่าประมาณ $\hat{\omega}_j$ และ $\mathrm{Var}(\hat{\omega})$ สำหรับรูปแบบ AR(1) และ MA(1) คือ

รูปแบบ AR(1)

$$\hat{\omega}_t = z_t - \frac{\hat{\phi}_1}{1+\hat{\phi}_1^2}(z_{t-1} + z_{t+1})$$ [2]

$$Var(\hat{\omega}) = (1+\hat{\phi}^2)^{-1}\sigma_a^2$$

รูปแบบ MA(1)

$$\hat{\omega}_t = \frac{1}{1+\hat{\theta}_1^2}\left[a_t + \hat{\theta}_1 a_{t+1} + \hat{\theta}_1^2 a_{t+2} + \ldots\right]$$

$$Var(\hat{\omega}) = (1 - \hat{\theta}^2)\sigma_a^2$$ [3]

ถ้า $\max|\hat{\tau}(t)| > C_c$ อนุกรมมีค่าผิดปกติ

ค่า C_c คือค่าวิกฤติ ในที่นี้ให้ $C_c = 3.5$ (C_c มีค่าเป็น 3.0,3.5 และ 4.0 สำหรับการตรวจสอบค่าผิดปกติที่มีความไวต่อการตรวจสอบ สูง กลาง และต่ำ ตามลำดับ [4])

1.3 ถ้าไม่พบค่าผิดปกติให้ไปทำขั้นตอนที่ 1.4

ถ้ามีค่าผิดปกติให้ขจัดค่าผิดปกติออกจากความคลาดเคลื่อนและค่าสังเกต

$$\tilde{Z}_t = Z_t - \hat{\omega}_t$$

$$\tilde{e}_t = \hat{e}_t - \hat{\omega}_t$$

หลังจากขจัดค่าผิดปกติแล้ว ทำการประมาณค่าความแปรปรวนของอนุกรมใหม่จากสูตร $\hat{\sigma}_a^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t^2$ โดยใช้ค่าสังเกตชุดที่เพิ่งขจัดค่าผิดปกติออก จากนั้นกลับไปทำขั้นตอนที่ 1.2 เพื่อตรวจหาค่าผิดปกติเพิ่มเติม ทำซ้ำขั้นตอนที่1.2 และ 1.3 จนกระทั่งไม่พบค่าผิดปกติเพิ่ม (โดยใช้ค่าประมาณพารามิเตอร์ของอนุกรมชุดเดิมที่คำนวณได้ในขั้นตอนที่ 1.1)

---

[2] A. J. Fox, "Outlier In Time Series", Journal of the Royal Statistical Society,Ser. B,34(1972),pp.352

[3] Cheng I, Tiao G. C. and Chen C. "Estimation of Time Series Parameters in the Presence of Outliers" Technometrics. ,30(1988),pp.194

[4] Ih Chang ,George C. Tiao and Chung Chen. Technometric ,May 1988,Vol 30 No. 2 , pp.196

1.4 ถ้าไม่พบค่าผิดปกติในการทำซ้ำรอบแรก แสดงว่าอนุกรมไม่มีค่าผิดปกติให้หยุดขบวนการและใช้ค่าประมาณพารามิเตอร์ของอนุกรมเวลาที่ได้ในการพยากรณ์

ถ้าพบค่าผิดปกติให้ประมาณค่าพารามิเตอร์โดยวิธีตัวประมาณภาวะน่าจะเป็นสูงสุดของอนุกรมใหม่ แล้วไปทำขั้นตอนที่ 2.1

2. ประมาณค่าผิดปกติและพารามิเตอร์ของอนุกรม

2.1 สมมติว่ามีค่าผิดปกติ m ค่าที่เวลา $t_1, t_2, \ldots, t_m$ ผลกระทบของค่าผิดปกติประมาณได้จาก

$$\hat{e}_t = \sum_{j=1}^{m} \omega_j \Pi(B) I_t(t_j) + a_t$$

โดยถือว่า $\{\hat{e}_t\}$ เป็นตัวแปรตาม

$\{I_t(t_j)\}$ เป็นตัวแปรอิสระ

และ

$$I_t(t_j) = \begin{cases} 1, & t = t_j \\ 0, & t \neq t_j \end{cases}$$

2.2 คำนวณค่า $\hat{\tau}_j$ ของค่า $\hat{\omega}_j$ ที่ประมาณได้เมื่อ

$$\hat{\tau}_j = \frac{\hat{\omega}_j}{[\mathrm{Var}(\hat{\omega})]^{(1/2)}} \qquad ,j = 1,2,\ldots,m$$

ถ้า $\min_j |\hat{\tau}_j| = \hat{\tau}_\gamma \leq C_c$ เมื่อ C_c เป็นค่าเดียวกับที่ใช้ในขั้นตอนที่ 1.2 ให้ตัดค่าที่ t_γ ออกจากชุดของข้อมูลผิดปกติ แล้วกลับไปทำขั้นตอนที่ 2.1 โดยมีค่าผิดปกติเหลืออยู่ m - 1 ตัว

ถ้า $\min_j |\hat{\tau}_j| > C_c$ ให้ไปทำขั้นตอนที่ 2.3

2.3 ขจัดอิทธิพลของค่าผิดปกติที่ตรวจพบในขั้นที่ 2.2 ออกจากอนุกรม

2.4 ประมาณค่าพารามิเตอร์ของอนุกรมเวลาที่ขจัดค่าผิดปกติแล้ว ถ้าค่าประมาณของส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน(Residual Standard Error) ต่างจากรอบก่อนหน้ามากกว่า ε ให้กลับไปทำขั้นตอนที่ 2.1 ใหม่ จนกระทั่งได้ความแตกต่างน้อยกว่า ε ในที่นี้ $\varepsilon = 0.001$

3. วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว

หลักการประมาณค่าโดยวิธีนี้คือการทำให้ผลรวมค่ากำลังสองของความคลาดเคลื่อนภายใต้พารามิเตอร์ของตัวแบบมีค่าต่ำที่สุด

จากรูปแบบ ARMA(p,q) สมการแสดงผลรวมค่ากำลังสองของความคลาดเคลื่อนของขบวนการนี้คือ

$$s(\phi,\theta) = \sum_{t=-\infty}^{n}[a_t|(\phi,\theta)]^2$$

3.1 เมื่อ z_t เป็นอนุกรมเวลาแบบ AR(1)

ตัวประมาณกำลังสองน้อยที่สุดคือ

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n} z_t z_{t-1}}{\sum_{t=3}^{n} z_{t-1}^2}$$

3.2 เมื่อ z_t เป็นอนุกรมเวลาแบบ MA(1)

ตัวประมาณกำลังสองน้อยที่สุดคือ

$$\hat{\theta}_1 = -\frac{\sum_{t=2}^{n} z_t a_{t-1}}{\sum_{t=3}^{n} a_{t-1}^2}$$

หลังจากได้ค่าประมาณของพารามิเตอร์แล้ว ให้นำค่าที่ได้ไปเป็นค่าประมาณเริ่มต้นสำหรับการปรับปรุงข้อมูลในขบวนการทำซ้ำเช่นเดียวกับขบวนการทำซ้ำของวิธีตัวประมาณภาวะน่าจะเป็นสูงสุด แต่ในการประมาณค่าพารามิเตอร์ของอนุกรมเวลาให้ประมาณด้วยวิธีกำลังสองน้อยที่สุด