บทที่ 2

สถิติที่ใช้ในการวิจัย

ในกรณีที่ผู้วิจัยต้องการประมาณค่าสูญหายในข้อมูลอนุกรมเวลา วิธีการประมาณค่าสูญหายที่ใช้มีอยู่หลายวิธี สำหรับการวิจัยครั้งนี้จะศึกษาวิธีการประมาณค่าสูญหาย 2 วิธี คือ วิธี Between-Forecast Estimation และวิธี Fixed-Point Smoothing ดังนั้นในบทนี้จะกล่าวถึงรายละเอียดของวิธีการประมาณค่าสูญหายในข้อมูลอนุกรมเวลาแต่ละวิธี ซึ่งมีรายละเอียดดังนี้

2.1 <u>วิธีการประมาณค่าสูญหายที่ใช้ในการศึกษา</u>

ในการวิจัยครั้งนี้ข้อมูลอนุกรมเวลาสามารถเขียนในรูปแบบทั่วไปได้ดังนี้

สมมติให้ $Z_1, Z_2, ---, Z_T$ คืออนุกรมเวลา

$a_1, a_2, ---, a_T$ คือ random shock ที่มีการแจกแจงปกติ ค่าเฉลี่ย 0

ความแปรปรวน σ^2_a

รูปแบบทั่วไปของอนุกรมเวลาบ็อกซ์-เจนกินส์ โดยอนุกรมเวลานี้เป็นอนุกรมเวลาคงที่คือ

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + --- + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - --- - \theta_q a_{t-q}$$

โดยที่ $\phi_1, ---, \phi_p$ คือสัมประสิทธิ์ความถดถอย

$\theta_1, ---, \theta_q$ คือสัมประสิทธิ์ค่าเฉลี่ยเคลื่อนที่

เรียกสมการข้างต้นว่า Autoregressive Moving-Average Model อันดับที่ p,q

โดยที่ p คืออันดับของรูปแบบถดถอย (Autoregressive model)

q คืออันดับของรูปแบบค่าเฉลี่ยเคลื่อนที่ (Moving-Average model)

ซึ่งเป็นรูปแบบคงที่ (Stationary models) ที่ศึกษาในครั้งนี้มี 5 รูปแบบด้วยกันคือ

1. The First Order Autoregressive Model หรือ AR(1) มีสมการคือ

$$Z_t = \phi_1 Z_{t-1} + a_t \qquad \text{โดยที่} \quad |\phi_1| < 1$$

2. The Second Order Autoregressive Model หรือ AR(2) มีสมการคือ

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t \qquad \text{โดยที่} \quad |\phi_2| < 1$$

$$\phi_2 \pm \phi_1 < 1$$

3. The First Order Moving-Average Model หรือ MA(1) มีสมการคือ

$$Z_t = a_t - \theta_1 a_{t-1} \qquad \text{โดยที่} \quad |\theta_1| < 1$$

4. The Second Order Moving-Average Model หรือ MA(2) มีสมการคือ

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \qquad \text{โดยที่} \quad |\theta_2| < 1$$

$$\theta_2 \pm \theta_1 < 1$$

5. The First Order Autoregressive-First Order Moving-Average Model หรือ ARMA(1,1) มีสมการคือ

$$Z_t = \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1} \qquad \text{โดยที่} \quad |\phi_1| < 1$$

$$|\theta_1| < 1$$

สำหรับการศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในครั้งนี้ ได้ทำการศึกษา 2 วิธีดังนี้

2.1.1 <u>วิธี Between-Forecast Estimation</u>

วิธีการนี้อาศัยเทคนิคการพยากรณ์ล่วงหน้าและย้อนหลังของบ๊อกซ์-เจนกินส์ ดังสมการที่ (1.1) และ (1.3) โดยนำค่าพยากรณ์ที่ได้มาเฉลี่ยแบบถ่วงน้ำหนัก และตัวถ่วงน้ำหนักนี้จะเป็นค่าที่ทำให้ตัวประมาณที่ได้มีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด (Minimum Mean Square Error) มีรายละเอียดดังนี้

สมมติให้อนุกรม $\{Z_t\}$; $t = 1, \_\_\_ ,T$ มีข้อมูลดังนี้คือ $Z_1, \_\_\_ ,Z_r, Z_{r+m+1}, \_\_\_ ,Z_T$ นั่นหมายความว่าอนุกรมเวลา $\{Z_t\}$ มีข้อมูลสูญหาย $Z_{r+1}, \_\_\_, Z_{r+m}$ ทั้งหมด m ค่า ดังนั้นเราจะต้องทำการประมาณค่าของ Z_{r+l} ; $l=1, \_\_\_, m$

จากสมการที่ (1.1) และ บ๊อกซ์-เจนกินส์ (1976) จะได้ค่าประมาณที่เหมาะสมของ Z_{r+1} คือ $Z_r(1)$ โดยมีความคลาดเคลื่อนในการประมาณ (Forecast error หรือ Prediction error) คือ

$$u_{r+l} = Z_{r+l} - Z_r(l) \qquad ; \quad l = 1,2,\_\_\_,m$$
$$= \sum_{k=0}^{l-1} \psi_k a_{r+l-k} \qquad ; \quad \psi_0 = 1$$

และมีความแปรปรวน

$$Var(u_{r+l}) = \sigma_a^2 \sum_{k=0}^{l-1} \psi_k^2$$

ในทำนองเดียวกันจากสมการที่ (1.3) จะได้ค่าประมาณที่เหมาะสมของ Z_{r+l} คือ $Z_{r+m+1}(m+1-l)$ โดยมีความคลาดเคลื่อนในการประมาณค่า (Backforecast error) คือ

$$v_{r+l} = Z_{r+l} - Z_{r+m+1}(m+1-l) \qquad ; \quad l = 1,\_\_\_,m$$
$$= \sum_{k=0}^{m-l} \psi_k e'_{r+l+k}$$

และมีความแปรปรวน

$$Var(v_{r+l}) = \sigma_e^2 \sum_{k=0}^{m-l} \psi_k^2$$

ดังนั้นค่าประมาณที่เหมาะสมที่สุดของ Z_{r+l} คือ $Z_{r,r+m+1}(l)$

จาก Damsleth (1980) จะได้

$$Z_{r,r+m+1}(l) = c_l Z_r(l) + d_l Z_{r+m+1}(m+1-l) \qquad \text{-------- (2.1)}$$

โดย $l = 1,\_\_\_,m$

c_l, d_l : คือตัวถ่วงน้ำหนัก (Weighted)

และมีความคลาดเคลื่อนในการประมาณค่า (Between - Forecast error)

$$W_{r+l} = Z_{r+l} - Z_{r,r+m+1}(l) \qquad \text{-------- (2.2)}$$

สำหรับการกำหนดค่า c_l และ d_l จะเลือกค่า c_l และ d_l ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าน้อยที่สุด (Minimize Mean Square error)

$$E[w^2_{r+l}] = E[Z_{r+l} - Z_{r,r+m+1}(l)]^2$$

จาก Abraham (1981) จะพบว่า

$$c_l = L_{22}(L_{11} - L_{12}) / H \qquad \text{-------- (2.3)}$$

$$d_l = L_{11}(L_{22} - L_{12}) / H \qquad \text{-------- (2.4)}$$

โดยที่ $H = L_{11}L_{22} - L_{12}^2$

$L_{11} = \sigma_z^2 - \sigma_1^2$

$L_{22} = \sigma_z^2 - \sigma_2^2$

$L_{12} = \sigma_z^2 + \sigma_{12} - \sigma_1^2 - \sigma_2^2$

$\sigma_1^2 = Var(u_{r+1}) = (\sum_{j=0}^{l-1} \psi_j^2)\,\sigma_a^2$

$\sigma_2^2 = Var(v_{r+1}) = (\sum_{j=0}^{m-1} \psi_j^2)\,\sigma_a^2$

$\sigma_z^2 = Var(Z_{r+1}) = (\sum_{j=0}^{\infty} \psi_j^2)\,\sigma_a^2$

$\sigma_{12} = COV(u_{r+1}, v_{r+1}) = \Psi_1' D \Psi_2$

$\Psi_1' = (1\ \psi_1 \cdots \psi_{l-1})$

$\Psi_2 = (1\ \psi_1 \cdots \psi_{m-1})$

$D = (d_{ij})$ เป็นเมตริกซ์ขนาด $l \times (m+1-l)$

$d_{ij} = E[a_{r+l+1-i}\ e_{r+l-1+j}]$; $i = 1,\_\_\_,l$; $l = 1,\_\_\_,m$

$= \sum_{h=0}^{\infty} (-\pi_h)\,\psi_{h+i+j-2}$ $j = 1,\_\_\_,m+1-l$

จากสมการที่ (2.3) และ (2.4) Abraham (1981) ได้ทำการศึกษาใน บางสถาณการณ์ ซึ่งสามารถสรุปได้ดังตารางต่อไปนี้

ตารางที่ 2.1 แสดงค่าตัวถ่วงน้ำหนักสำหรับข้อมูลสูญหายจำนวน 1 ค่า ; m = 1

$\sigma_1^2 = \sigma_2^2 = \sigma_a^2$; $\sigma_{12} = (1 - \sum_{h=1}^{\infty} \pi_h \psi_h)\,\sigma_a^2$	
รูปแบบ	$c_1 = d_1$
ARMA(1,0)	$1 / (1 + \phi^2)$
ARMA(0,1)	1
ARMA(1,1)	$(1 - \phi\theta) / (1 + \phi^2 - 2\phi\theta)$

ตารางที่ 2.2 แสดงค่าตัวถ่วงน้ำหนักสำหรับข้อมูลสูญหายจำนวน 2 ค่า ; $m = 2$

เมื่อ $l = 1$; $\sigma_{1}^{2} = \sigma_{a}^{2}$; $\sigma_{\varepsilon}^{2} = (1+\psi_{1}^{2})\,\sigma_{a}^{2}$

$$\sigma_{1\varepsilon} = \{(1 - \sum_{h=1}^{\infty} \pi_h \psi_h) + \psi_1(\psi_1 - \sum_{h=1}^{\infty} \pi_h \psi_{h+1})\}\,\sigma_{a}^{2}$$

เมื่อ $l = 2$; $\sigma_{1}^{2} = (1+\psi_{1}^{2})\,\sigma_{a}^{2}$; $\sigma_{\varepsilon}^{2} = \sigma_{a}^{2}$

$\sigma_{1\varepsilon} = \sigma_{1\varepsilon}$ ในกรณีที่ $l = 1$

รูปแบบ	l	c_l	d_l
ARMA(1,0)	1	$(1+\phi^{2})/(1+\phi^{2}+\phi^{4})$	$1/(1+\phi^{2}+\phi^{4})$
	2	d_1	c_1
ARMA(1,1)	1	$\dfrac{(1+\phi^{2}-\phi\theta)(1-\phi\theta)}{(1+\phi+\phi^{2}-\phi\theta)(1-\phi+\phi^{2}-\phi\theta)}$	$\dfrac{(1-\phi\theta)}{(1+\phi+\phi^{2}-\phi\theta)(1-\phi+\phi^{2}-\phi\theta)}$
	2	d_1	c_1

ดังนั้นเราจะได้ค่าประมาณค่าสูญหาย คือ

$$X_{r,r+m+1}(l) = Z_{r,r+m+1}(l) \qquad \text{---------- (2.5)}$$

2.1.2 Fixed - Point Smoothing Method

วิธีการนี้ได้อาศัยเทคนิคการย้อนกลับ (recursion) เริ่มตั้งแต่ $t = 0,1,$ ---,T โดยในแต่ละเวลาที่เปลี่ยนแปลงไปจะได้ค่าประมาณของข้อมูล ณ เวลานั้น ๆ ซึ่งค่าประมาณเหล่านี้จะอาศัย Kalman Filter และเมื่อพบค่าสูญหายก็จะสามารถประมาณค่าสูญหายโดยวิธี Fixed-Point Smoothing ซึ่งมีรายละเอียดดังนี้

จากสมการที่ (1.1) และ Akaike's Markovian ได้กำหนด state ของขบวนการ ARMA(p,q) ได้ดังนี้

ให้ $m' = \mathrm{Max}\,(p,q+1)$

$$Z(t) = \begin{bmatrix} x(t/t) \\ x(t+1/t) \\ \vdots \\ x(t+m'-1/t) \end{bmatrix}$$

$x(t+j/t)$: คือโปรเจคชั่น (projection) ของ x_{t+j} บนค่าของอนุกรมเวลาเมื่อเวลา t

$x(t/t)$: คือค่าสังเกตของอนุกรมเวลาเมื่อเวลา t นั่นคือ $x(t/t) = x_t$

$Z(t)$: คือเมตริกซ์ของค่าสังเกต มีขนาด $m' \times 1$

จะได้ state space representation คือ

$$Z(t+1) = FZ(t) + Ga_{t+1}$$

$$y(t) = HZ(t) + v(t)$$

โดย

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ \phi_{m'} & \dots & & \phi_2 & \phi_1 \end{bmatrix}_{m' \times m'}$$

$$G = \begin{bmatrix} 1 \\ g_2 \\ \vdots \\ g_{m'} \end{bmatrix}_{m' \times 1}$$

โดย $g_1 = 1$

$$g_j = \phi_{j-1} + \sum_{k=1}^{j-1} \phi_k g_{j-k} \quad ; j = 2, \dots, m'$$

$$H = [\,1 \; 0 \; \dots \; 0\,]_{1 \times m'}$$

$y(t)$: คือข้อมูลอนุกรมเวลาที่เวลา t

a_{t+1} : คือตัวแปรสุ่มอิสระที่มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวนคงที่ σ_a^2

$v(t)$: คือตัวแปรสุ่มที่แสดงถึง observation error ที่เป็นอิสระ ค่าเฉลี่ย 0 ความแปรปรวน $R = E[v(t)]^2$

จาก Kalman recursive estimation เราสามารถที่จะสรุปเป็นขั้นตอนได้ดังนี้

ให้ $\in(t+1)$ คือความคลาดเคลื่อนของการพยากรณ์ (Forecast error) นั่นคือ

$$\in(t+1) = y(t+1) - y(t+1/t)$$

ขั้นตอนการประมาณค่าสูญหายมีดังนี้

<u>ขั้นตอนที่ 0</u> ให้ $x(0/0) = 0$

$$P(0/0) = (p_{ij}(0/0))$$

$P(0/0)$: คือเมตริกซ์ความแปรปรวนร่วม (covariance matrix) ขนาด $m' \times m'$ เมื่อ $t=0$

เมื่อ $j > 0$ จะได้ $p_{0j}(0/0) = C(j)$

เมื่อ $j >= i > 0$ จะได้ว่า $p_{ij}(0/0) = C(j-i) - \sum_{k=0}^{i-1} \gamma_k \gamma_{k+j-i}$

ให้ γ_k คือความแปรปรวนร่วม (cross covariances) ระหว่างอนุกรมเวลากับความคลาดเคลื่อน

C_k คือความแปรปรวนร่วม (covariances)

โดย $\gamma_0 = 1$

$$\gamma_1 = \phi_1\gamma_0 + \theta_1$$

$$\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 + \theta_2$$

$$\vdots$$

และ $C_k = E\{x_{t+k}\, x_t\}$

$$= \sum_{j=1}^{p} \phi_j C_{k-j} + \sum_{j=k}^{q} \theta_j \gamma_{j-k}$$

$$C_0 = \phi_1 C_1 + \phi_2 C_2 + \ldots + \phi_p C_p + \gamma_0 + \theta_1\gamma_1 + \theta_2\gamma_2 + \ldots$$

$$C_1 = \phi_1 C_0 + \phi_2 C_{-1} + \ldots + \phi_p C_{1-p} + \theta_1\gamma_0 + \theta_2\gamma_1 + \ldots$$

$$\vdots$$

$$C_p = \phi_1 C_{p-1} + \phi_2 C_{p-2} + \ldots + \phi_p C_0 + \theta_p\gamma_0 + \theta_{p+1}\gamma_1 + \ldots$$

<u>ขั้นตอนที่ 1</u> โดย ordinary kalman filter

$$Z(t+1/t) = FZ(t/t)$$

$$P(t+1/t) = FP(t/t)F' + GG'$$

$$y(t+1/t) = HZ(t+1/t)$$

เมื่อ $y(t+1)$ สูญหายให้ไปยัง ขั้นตอนที่ 2

$y(t+1)$ ไม่สูญหายให้ไป ขั้นตอนที่ 3

<u>ขั้นตอนที่ 2</u>

$$Z(t+1/t+1) = Z(t+1/t)$$

$$P(t+1/t+1) = P(t+1/t)$$

กลับไป ขั้นตอนที่ 1

<u>ขั้นตอนที่ 3</u>

$$\epsilon(t+1) = y(t+1) - y(t+1/t)$$

$$Var(\epsilon(t+1)) = P_{11}(t+1/t)$$

$$Z(t+1/t+1) = Z(t+1/t) + P(t+1/t)H' \epsilon(t+1)[P_{11}(t+1/t)]^{-1}$$

$$P(t+1/t+1) = P(t+1/t) + P(t+1/t)H'HP(t+1/t)[P_{11}(t+1/t)]^{-1}$$

จากขั้นตอนที่ 2 พบว่า y(t+1) สูญหายไป ดังนั้นเราสามารถที่ประมาณค่า y(t+1) ได้โดย Fixed-Point Smoothing Algorithm ดังรายละเอียดต่อไปนี้

ให้ข้อมูลสูญหาย ณ เวลา $\tau = \tau_j$; j=1,---,m และตัวประมาณของ $y(\tau)$ คือ $y(\tau/T)$ จาก R.Kohn and C.F.Ansley (1985) สรุปได้ดังนี้

<u>ขั้นตอนที่ 1'</u> สำหรับ $\tau = \tau_1,---,\tau_m$ ถ้า $\tau <= t$ แล้ว

$$\nu(t+1/t;\tau) = F\nu(t/t;\tau)$$

<u>ขั้นตอนที่ 2'</u> y(t+1) is missing For $\tau = \tau_1,---,\tau_m$

ถ้า $\tau <= t$

$$y(\tau/t+1) = y(\tau/t)$$

$$s(\tau/t+1) = s(\tau/t)$$

$$\nu(t+1/t+1;\tau) = \nu(t+1/t;\tau)$$

กลับไป ขั้นตอนที่ 1'

ถ้า $\tau = t+1$

$$y(t+1/t+1) = Z_1(t+1/t+1)$$

$$s(t+1/t+1) = P_{11}(t+1/t+1)$$

$$\nu(t+1/t+1;\tau) = P(t+1/t+1)H'$$

กลับไป ขั้นตอนที่ 1'

<u>ขั้นตอนที่ 3</u> ถ้า $\tau <= t$

$$y(\tau/t+1) = y(\tau/t) + \nu(t+1/t;\tau)\,\epsilon(t+1)[P_{11}(t+1/t+1)]^{-1}$$

$$s(\tau/t+1) = s(\tau/t) - \{\nu(t+1/t;\tau)\}^2\,[P_{11}(t+1/t+1)]^{-1}$$

$$\nu(t+1/t+1;\tau) = \nu(t+1/t;\tau) - \nu(t+1/t;\tau)P(t+1/t+1)H'[P_{11}(t+1/t+1)]^{-1}$$

โดยที่ $Z_1(t+1/t+1)$, $P(t+1/t+1)$, $\epsilon(t+1)$ เป็นค่าที่ได้มาจาก ordinary kalman filter

2.2 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าสูญหาย

เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าสูญหายคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error ; MSE)

$$MSE = \frac{\sum_{i=1}^{m} (z_i - \hat{z}_i)^2}{m}$$

z_i : คือค่าข้อมูลจริงที่สูญหายตัวที่ i จากแต่ละตัวอย่าง

$\hat{z}_i$: คือค่าประมาณของข้อมูลที่สูญหายตัวที่ i จากแต่ละตัวอย่าง

m : คือจำนวนข้อมูลสูญหาย x จำนวนรอบ