# บทที่ 3

## สถิติที่ใช้ในการวิเคราะห์

<u>อนุกรมเวลาบ๊อกซ์-เจนกินส์</u> (Box-Jenkins Methodology)

การพยากรณ์เป็นกระบวนการทางคณิตศาสตร์และสถิติอย่างหนึ่งที่นิยมใช้กันอย่างกว้างขวางในหลายวงการ โดยเฉพาะวงการธุรกิจ การพยากรณ์ในบางครั้งอาจไม่จำเป็นต้องให้เกิดความแม่นยำสูงนักซึ่งมักจะเป็นการพยากรณ์ในระยะยาว แต่ในบางครั้งก็ต้องการให้เกิดความแม่นยำสูงลักษณะเช่นนี้จะเป็นการพยากรณ์ในระยะสั้น เทคนิคการพยากรณ์มีหลายวิธี บ๊อกซ์-เจนกินส์เทคนิค เป็นเทคนิคที่มีประสิทธิภาพสูงเทคนิคหนึ่งสำหรับการพยากรณ์ในระยะสั้น ซึ่งได้รับความนิยมมากในระยะหลังโดยอาศัยความสัมพันธ์จากข้อมูลในอดีตของตัวเองเพื่อหารูปแบบการเปลี่ยนแปลงมาเป็นแนวทางในการกำหนดพฤติกรรมในอนาคต บ๊อกซ์และเจนกินส์ได้เสนออนุกรมเวลาแบบไม่ต่อเนื่อง (Discrete Time Series) $\{ Z_t ; t = 0,1,2,\ldots\}$ ซึ่งเป็นกระบวนการเชิงความน่าจะเป็น (Stochastic Process) ที่อยู่ในสภาวะสมดุลย์เชิงสถิติ (Statistical Equilibrium) และเป็นกระบวนการที่มีคุณสมบัติไม่เปลี่ยนแปลงเมื่อจุดเริ่มต้นหรือเวลาเปลี่ยนแปลงไปที่เรียกว่า Stationary Process ซึ่งการแจกแจงความน่าจะเป็น $f(Z_t)$ ของข้อมูล $Z_1, Z_2, \ldots, Z_t$ มีค่าคาดหมายและความแปรปรวนของกระบวนการคงที่เขียนได้เป็น

$$E(Z_t) = \int_{-\infty}^{\infty} Z\, f(Z)\, dZ$$

และ $$V(Z_t) = E(Z_t - \mu)^2 = \int_{-\infty}^{\infty} (Z-\mu)^2 f(Z)\, dZ$$

ซึ่งในทางปฏิบัติไม่อาจทราบค่าที่แท้จริงได้จึงต้องประมาณจากค่าสังเกต $Z_1, Z_2, \ldots Z_N$ ด้วย

$$\bar{Z} = \frac{1}{N} \sum_{t=1}^{N} Z_t \qquad \hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^{N} (Z_t - \bar{Z})^2$$

เทคนิคการพยากรณ์โดยทั่วไปจะกำหนดรูปแบบความสัมพันธ์ขึ้นก่อนทำการวิเคราะห์ ซึ่งมักประสบปัญหาการกำหนดรูปแบบฟังก์ชั่นต่าง ๆ แต่อนุกรมเวลาบ๊อกซ์-เจนกินส์ได้ขจัดปัญหาดังกล่าวโดยไม่มีการกำหนดรูปแบบขึ้นตายตัวก่อนทำการวิเคราะห์ รูปแบบความสัมพันธ์จะถูกกำหนดขึ้นจากขั้นตอนการวิเคราะห์ความสัมพันธ์ของข้อมูลในตัวเอง แล้วจึงเลือกรูปแบบมาประมาณค่าพารามิเตอร์ ก่อนที่จะพยากรณ์จะต้องทำการตรวจสอบความเหมาะสม

ของตัวแบบก่อน ถ้าตัวแบบที่เลือกไม่เหมาะสมซึ่งหมายความว่าความคลาดเคลื่อนที่เกิดขึ้นไม่ได้เป็นไปโดยสุ่ม ก็จะอาศัยความสัมพันธ์ของความคลาดเคลื่อนนี้เป็นแนวทางในการเลือกรูปแบบใหม่จนกว่าจะพบตัวแบบที่เหมาะสมแล้วจึงพยากรณ์ค่า การวิเคราะห์อนุกรมเวลาบ๊อกซ์-เจนกินส์ต้องอาศัยคุณสมบัติการไม่เปลี่ยนแปลงของขบวนการเมื่อเวลาเปลี่ยนไป ดังนั้น Stationarity ของอนุกรมเวลาจึงเป็นสิ่งจำเป็น แต่ในทางปฏิบัติแล้ว ข้อมูลที่วิเคราะห์อาจเป็น nonstationary จะต้อง transform ข้อมูลให้มีคุณสมบัติเป็น Stationary วิธีหนึ่งที่ทำได้โดยการหาผลต่างลำดับต่าง ๆ ของข้อมูลจนกว่าจะเป็น Stationary แบ่งขั้นตอนในการวิเคราะห์อนุกรมเวลาบ๊อกซ์-เจนกินส์สำหรับ Stationary Process ได้ดังนี้

1. การกำหนดรูปแบบ (Identification)
2. การประมาณค่าพารามิเตอร์ (Parameter Estimation)
3. การตรวจสอบความเหมาะสมของตัวแบบ (Diagonostic Checking)
4. การพยากรณ์ค่า (Forecasting)

1. <u>การกำหนดรูปแบบ</u>

อาศัยความสัมพันธ์ในตัวเองของข้อมูลจากฟังก์ชั่นสหสัมพันธ์ในตัวเอง (Autocorrelation Function) กับฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function) เป็นตัวบอกถึงรูปแบบของกระบวนการและบอกถึงลำดับหรือจำนวนเทอมของข้อมูลที่จะพิจารณาย้อนหลัง ฟังก์ชั่นสหสัมพันธ์ในตัวเองที่ช้ากว่ากัน j หน่วยเวลา นิยามให้เท่ากับ

$$\rho_j = \frac{\gamma_j}{\gamma_o}$$

เมื่อ $\gamma_j = \text{cov}\,(Z_t, Z_{t+j}) = E(Z_t - \mu)(Z_{t+j} - \mu)$ ; $j = 0,1,2,\ldots,K$ ในทางปฏิบัติค่า K เท่ากับ 25 ก็พอเพียงที่จะเห็นลักษณะของฟังก์ชั่น

$\gamma_j$ เรียกสหความแปรปรวน (Autocovariance) ระหว่าง $Z_t$ กับ $Z_{t+j}$ จะประมาณ $\gamma_j$ ด้วย $C_j$ ซึ่ง

$$C_j = \frac{1}{N-j} \sum_{t=1}^{N-j} (Z_t - \bar{Z})(Z_{t+j} - \bar{Z}) \quad j = 0,1,2,\ldots,K$$

ดังนั้นประมาณ $\rho_j$ ด้วย $r_j = \frac{C_j}{C_o}$ ซึ่งมีค่าอยู่ระหว่าง -1 ถึง 1 Bartlett ให้สูตรประมาณในการกำหนดค่าความแปรปรวนของ $r_j$ เมื่อ j มีค่ามากกว่า q ไว้ ดังนั้น

การคำนวณความคลาดเคลื่อนมาตรฐาน (Standard Error) ของ $r_j$ ใช้

$$SE\ (r_j) \approx \frac{1}{\sqrt{N}}\ (1 + 2 \sum_{v=1}^{q} r_v^2)\ ,\quad j > q$$

ซึ่งจะใช้ในการทดสอบนัยสำคัญของ $\rho_j$ ว่าเท่ากับศูนย์หรือไม่

สำหรับการคำนวณฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วน คำนวณจากสมการยูล-วอล์กเกอร์ (Yule - Walker eqvation)

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{K-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{K-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{K-1} & \rho_{K-2} & \rho_{K-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \phi_{22} \\ \vdots \\ \phi_{KK} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_K \end{bmatrix}$$

เมื่อ $\phi_{KK}$ เป็นสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนที่ช้ากว่ากัน K หน่วย หรือประมาณ $\phi_{KK}$ ด้วย $\hat{\phi}_{KK}$ จาก Recursive Relation

$$\hat{\phi}_{KK} = \begin{cases} r_1 & ;\ K = 1 \\ \dfrac{r_K - \sum_{j=1}^{K-1} \hat{\phi}_{K-1,j} \cdot r_{K-j}}{1 - \sum_{j=1}^{K-1} \hat{\phi}_{K-1,j} \cdot r_j} & ;\ K = 2, 3, \ldots K \end{cases}$$

$$\hat{\phi}_{Kj} = \hat{\phi}_{K-1,j} - \hat{\phi}_{kk} \cdot \hat{\phi}_{K-1,K-j} \quad ;\ j = 1,2,\ldots,k-1$$

ในการทดสอบนัยสำคัญของ $\phi_{KK}$ Quenouilli ให้สูตรประมาณในการคำนวณค่าความแปรปรวนของ $\hat{\phi}_{KK}$ ซึ่งมีค่ามากกว่า p การคำนวณค่าความคลาดเคลื่อนมาตรฐานของ $\hat{\phi}_{KK}$ ใช้

$$SE\ (\hat{\phi}_{KK}) \approx \frac{1}{\sqrt{N}} \quad ;\ K > p$$

บ็อกซ์และเจนกินส์ได้สร้างตัวแบบสำหรับการวิเคราะห์อนุกรมเวลาไว้ 3 ประเภท ดังนี้

1.1 Autoregressive Process of Order P (AR(P)) เป็นกระบวนการถดถอยในตัวเองแสดงความสัมพันธ์ของค่าอนุกรมเวลาในปัจจุบันกับค่าในอดีตพิจารณาเพียง P เทอมดังนี้

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + u_t$$

เมื่อ $\delta$ = ค่าคงที่ของกระบวนการ

$\phi_j$ = Autoregressive Parameter ; $j = 1,2,\dots,p$

$u_t$ = การแปรปรวน (Disturbance) สุ่มซึ่งเป็นอิสระต่อกันและมีการแจกแจงของความน่าจะเป็นเดียวกันมีค่าคาดหมายเท่ากับศูนย์และความแปรปรวนเท่ากับ $\sigma_u^2$

แบบจำลอง AR ไม่มีข้อจำกัดของพารามิเตอร์ที่ทำให้แบบจำลอง Invertible (หมายถึง ข้อจำกัดพารามิเตอร์ที่ทำให้เมตริกซ์ผกผัน (Invert Matrix) exist ในการประมาณค่า) แต่มีข้อจำกัดของพารามิเตอร์ที่ทำให้แบบจำลองเป็น Stationary ฟังก์ชั่นสหสัมพันธ์ในตัวเอง $r_K$ มีค่ามากในค่าแรก ๆ และลดลงอย่างช้า ๆ เมื่อ K มีค่ามากขึ้นเกือบถึงศูนย์หรือ tails off มีลักษณะ Damped Exponential หรือ Damped Sine Wave หรือทั้ง 2 แบบรวมกัน ฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วนจะมีค่าลดลงอย่างรวดเร็ว ค่า $\phi_{KK}$ มีค่าเป็นศูนย์หลัง lag p หรือ Cutts off แบบจำลอง AR จะมี Order เป็น p

$$\phi_{KK} \neq 0 \quad ; \quad K \leq p$$

และ $$\phi_{KK} = 0 \quad ; \quad K > p$$

ลำดับของกระบวนการ AR ในทางปฏิบัติมัก ไม่เกิน 2 นั่นคือ $P \leq 2$

1.1.1 First Order Autoregressive Model (AR(1))

$$z_t = \delta + \phi_1 z_{t-1} + u_t$$

ข้อจำกัดที่จะต้องมีสำหรับพารามิเตอร์ $\phi_1$ ที่ทำให้แบบจำลองเป็น Stationary คือ $|\phi_1| < 1$ ฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วน Cutts off หลัง Lag 1 และฟังก์ชั่นสหสัมพันธ์ในตัวเองลดลงแบบ Exponential

1.1.2 Second Order Autoregressive Model (AR(2))

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + u_t$$

ข้อจำกัดของพารามิเตอร์ที่ทำให้แบบจำลองเป็น Stationary คือ

$$\phi_1 + \phi_2 < 1$$
$$\phi_2 - \phi_1 < 1$$
$$|\phi_2| < 1$$

ฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วนจะ Cutts off หลัง Lag 2 และฟังก์ชั่นสหสัมพันธ์ในตัวเองลดลงในลักษณะของ Damped Exponentials และ/หรือ Damped Sine Waves

1.2 Moving Average of Order Q (MA(Q)) แสดงค่าอนุกรมเวลาอยู่ในเทอมของการแปรปรวน ($u_t$) โดยพิจารณาเพียง q เทอม

$$z_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} \cdots\cdots - \theta_q u_{t-q}$$

เมื่อ $\theta_i$ เป็น Moving Average Parameter ($i = 1,2,\ldots,q$)

ไม่มีข้อจำกัดของพารามิเตอร์ที่ทำให้แบบจำลองเป็น Stationary แต่มีข้อจำกัดของพารามิเตอร์ที่ทำให้แบบจำลอง Invertible ฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วนมีลักษณะ Damped Exponential หรือ Damped Sine Waves หรือทั้ง 2 แบบรวมกันเป็น tails off ฟังก์ชั่นสหสัมพันธ์ในตัวเองจะมีลักษณะ Cutts off ค่า $r_K$ ลดลงอย่างรวดเร็ว จะมีลำดับ q เมื่อหลัง lag q มีค่าเป็นศูนย์ ในทางปฏิบัติลำดับ q มักมีค่าไม่เกิน 2

1.2.1 First Order Moving Average Model (MA(1))

$$z_t = \mu + u_t - \theta_1 u_{t-1}$$

ข้อจำกัดที่ทำให้แบบจำลองเป็น Invertible คือ $|\theta_1| < 1$

ฟังก์ชั่นสหสัมพันธ์ในตัวเอง Cutts off หลัง lag 1 ฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วนมีลักษณะ Damped Exponential

1.2.2 Second Order Moving Average Model (MA(2))

$$z_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}$$

ข้อจำกัดของพารามิเตอร์ทำให้แบบจำลอง Invertible ได้แก่

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$|\theta_2| < 1$$

ฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วนจะลดลงมีลักษณะผสมของ Damped Exponential และ/หรือ Damped Sine Waves และฟังก์ชั่นสหสัมพันธ์ในตัวเองจะ Cutts off หลัง lag 2

1.3 Mixed Autoregressive – Moving Average Process of Order P and Q (ARMA(P,Q)) เป็นตัวแบบที่มีลักษณะผสมของ AR และ MA ที่มีลำดับ P และ Q ตามลำดับ

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

โดยที่ข้อจำกัดที่ทำให้แบบจำลองเป็น Stationary สำหรับกระบวนการ AR ลำดับที่ P และ Invertibility สำหรับกระบวนการ MA ลำดับที่ Q ได้แก่ฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วนและฟังก์ชั่นสหสัมพันธ์ในตัวเองมีลักษณะ Damped Exponential หรือ Damped Sine Waves หรือทั้ง 2 แบบรวมกันเป็น tail off

1.3.1 Mixed Autoregressive-Moving Average Model of Order (1,1) (ARMA(1,1))

$$Z_t = \delta + \phi_1 Z_{t-1} + u_t - \theta_1 u_{t-1}$$

กระบวนการจะเป็น Stationary ถ้า $|\phi_1| < 1$ และจะ Invertible ถ้า $|\theta_1| < 1$

เมื่อกระบวนการเป็น Nonstationary จะทำอนุกรมเวลาให้เป็น Stationary โดยการหาผลต่างลำดับต่าง ๆ ของข้อมูล เมื่อเป็นผลต่างลำดับที่ 1 ได้ว่า

$$W_t = Z_t - Z_{t-1}$$

แล้วใช้ $W_t$ ในตัวแบบต่าง ๆ แทน $Z_t$ เรียกกระบวนการ ARMA ในรูปของ $W_t$ ว่า Autoregressive Integrated Moving Average (ARIMA) นิยามให้ ARIMA(p,d,q) เมื่อ d = 1 เป็น

$$W_t = \phi_1 W_{t-1}+\phi_2 W_{t-2}+\ldots+\phi_p W_{t-p}+u_t-\theta_1 u_{t-1}-\theta_2 u_{t-2}\ldots\theta_q u_{t-q}$$

หรือเขียนในเทอมของ $Z_t$ ได้เป็น

$$z_t = z_{t-1} - 1 + \phi_1(z_{t-1} - z_{t-2}) + \phi_2(z_{t-2} - z_{t-3}) + \ldots + \phi_p(z_{t-p} - z_{t-p-1})$$

$$+ u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \ldots \theta_q u_{t-q}$$

ถ้าผลต่างลำดับที่ 1 ของข้อมูลยังไม่ทำให้อนุกรมเวลาเป็น Stationary ก็หาผลต่างลำดับอื่น ๆ ไปจนกว่าจะเป็น Stationary ในทางปฏิบัติแล้วมักจะใช้ผลต่างลำดับที่ 0, 1 และอย่างมาก 2 ก็สามารถทำให้รูปแบบที่ได้เป็น Stationary กระบวนการ ARIMAหากไม่มีเทอมเฉลี่ยเคลื่อนที่หรือเทอมถดถอยในตัวเองเขียนได้เป็น

$$ARIMA(p,d,o) = ARI(p,d)$$

$$ARIMA(q,d,o) = IMA(q,d)$$

2. <u>การประมาณค่าพารามิเตอร์</u> หลังจากได้เลือกรูปแบบในขั้นตอนที่ 1 แล้วต่อไปจะประมาณค่าพารามิเตอร์ด้วยวิธีการของ Non-linear Programing แต่ตัวแบบที่เลือกนี้ยังใช้ไม่ได้ จนกว่าจะตรวจสอบแล้วว่าเหมาะสม โดยพิจารณาว่าความคลาดเคลื่อนเป็นไปโดยสุ่มหรือไม่ ถ้าไม่ต้องกลับสู่ขั้นตอนที่ 1 ใหม่จนกว่าจะพบความเหมาะสม ตัวแบบนั้นจึงจะนำไปพยากรณ์

3. <u>การตรวจสอบความเหมาะสมของตัวแบบ</u> จากสหสัมพันธ์ในตัวเองของความคลาดเคลื่อนที่ช้ากว่ากันในหน่วยเวลาต่าง ๆ Box และ Pierce ได้เสนอวิธีทดสอบโดยสร้างตัวสถิติที่มีการแจกแจงแบบไคสแควร์ เรียกตัวสถิตินี้ว่า Box-Pierce Chi-Square Statistics ใช้สัญญาลักษณ์ Q

$$Q = (N-d) \sum_{i=1}^{k} r_i^2(u)$$

เมื่อ $N$ = จำนวนข้อมูลที่ใช้ในการวิเคราะห์อนุกรมเวลา

$d$ = ลำดับผลต่างของข้อมูลที่ทำให้เป็น Stationary

$k$ = จำนวน lag

$r_i^2(u)$ = กำลังสองของสหสัมพันธ์ในตัวเองของค่าความคลาดเคลื่อนที่ช้ากว่ากัน $i$ หน่วย

Q มีองศาความเป็นอิสระเท่ากับ k- จำนวนพารามิเตอร์ที่ต้องประมาณในตัวแบบ จะทำการทดสอบว่า $\rho_i$ ( ;i=1,2,...,K) มีค่าเป็นศูนย์หรือไม่โดยเทียบ Q กับค่าจากตาราง $(\chi^2_{\alpha,df})$ ถ้า $Q \geq \chi^2_{\alpha,df}$ แล้วแสดงว่า $\rho_i$ มีค่าความคลาดเคลื่อนยังมีสหสัมพันธ์กันอยู่ จึงต้องปรับปรุงตัวแบบใหม่โดยพิจารณาฟังก์ชั่นสหสัมพันธ์ในตัวเองของความคลาดเคลื่อนว่าควรเพิ่มเทอมใดลงไป ถ้า $Q < \chi^2_{\alpha,df}$ ก็แสดงว่าความคลาดเคลื่อนเป็นอิสระต่อกัน ตัวแบบที่

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

I 15456729

ได้ถือว่าเป็นตัวแบบที่เหมาะสมและจะนำไปใช้ในการพยากรณ์ต่อไป

4. <u>การพยากรณ์ค่า</u> อนุกรมเวลาบ๊อกซ์-เจนกินส์สามารถพยากรณ์ไปข้างหน้ากี่หน่วยเวลาก็ได้ แต่ยิ่งการพยากรณ์ออกไปไกลเท่าไรค่าพยากรณ์ที่ได้จะอาศัยสาระจากข้อมูลจริงน้อยลงมากเท่านั้น โดยนำค่าจากการพยากรณ์ในคาบเวลาที่ผ่านมามาใช้พยากรณ์ในเวลาข้างหน้า ความแม่นยำจึงลดลงมากถ้ายิ่งพยากรณ์ออกไปไกล และถ้าสหสัมพันธ์ในตัวเองมีเครื่องหมายบวกค่าพยากรณ์ในอนาคตจะมีแนวโน้มสูงขึ้นเรื่อย ๆ ในทำนองเดียวกันถ้าสหสัมพันธ์มีเครื่องหมายลบค่าพยากรณ์จะมีแนวโน้มลดลงเรื่อย ๆ ทั้ง ๆ ที่ในความเป็นจริงแล้วพฤติกรรมในอนาคตมีแนวโน้มที่เป็นได้ทั้งเพิ่มขึ้นและลดลง อนุกรมเวลาบ๊อกซ์ - เจนกินส์จึงเหมาะสำหรับการพยากรณ์ในระยะสั้น ๆ และควรพยากรณ์เพียง 1 หน่วยเวลาข้างหน้าเท่านั้นจึงเกิดความแม่นยำสูงที่สุด เนื่องจากได้อาศัยสาระของข้อมูลที่เกิดขึ้นจริงมากที่สุด

วิธีการพยากรณ์ค่าให้ $\hat{Z}_{T+t}(T)$ เป็นค่าพยากรณ์ t หน่วยเวลาข้างหน้า ณ เวลา T พบว่า $\hat{Z}_{T+t}(T)$ เป็นฟังก์ชั่นของ $Z_{t+j}$ ถ้า $j \leq 0$ แล้ว $Z_{t+j}$ จะเป็นค่าจริงของข้อมูล ถ้า $j > 0$ แล้ว $Z_{t+j}$ ต้องใช้ $\hat{Z}_{T+j}(T)$ ซึ่งเป็นค่าประมาณต่อไปพิจารณาเทอม $u_{T+j}$ ถ้า $j \leq 0$ จะประมาณ $u_{T+j}$ โดยใช้ค่าความคลาดเคลื่อน 1 คาบเวลาจากการพยากรณ์คือ $Z_{T+j} - \hat{Z}_{T+j}(T+j-1)$ ถ้า $j > 0$ แล้ว $u_{T+j}$ จะเท่ากับศูนย์ ยกตัวอย่างถ้าจะพยากรณ์ค่าจากตัวแบบ ARMA (1,1) ที่มีสมการเป็น

$$\hat{Z}_{t+1} = 0.345 + Z_t - 0.25\, u_t$$

หรือ $$\hat{Z}_{T+t}(T) = 0.345 + Z_{T+j} - 0.25\, u_{T+j}$$

ถ้าพยากรณ์ 1 หน่วยเวลาล่วงหน้า ณ คาบเวลาที่ 70 สมมติให้ $\hat{Z}_{70}(69) = 198$ และ $Z_{70} = 200$ ดังนั้น $t = 1,\ j = 0$ ได้ว่า

$$\hat{Z}_{71}(70) = 0.345 + Z_{70} - 0.25\,(Z_{70} - \hat{Z}_{70}(69))$$
$$= 0.345 + 200 - 0.25\,(200 - 198)$$
$$= 199.845$$

ถ้าพยากรณ์ 2 หน่วยเวลาล่วงหน้าแล้ว $t = 2,\ j = 1$ ได้ว่า

$$\hat{Z}_{72}(70) = 0.345 + \hat{Z}_{71}(70) - 0.25\,(0)$$

$= 0.345 + 199.845 - 0$

$= 200.19$

จะเห็นว่าการพยากรณ์ 1 หน่วยเวลาล่วงหน้าใช้สาระของข้อมูลจริงมากกว่าการพยากรณ์ 2 หน่วยเวลาล่วงหน้า เนื่องจากค่าพยากรณ์ 2 หน่วยเวลาล่วงหน้าเป็นฟังค์ชั่นของค่าที่ต้องพยากรณ์จาก 1 หน่วยเวลาล่วงหน้า

<u>การเฉลี่ยเคลื่อนที่ซ้ำสองครั้ง</u> (Double Moving Average)

การเฉลี่ยเคลื่อนที่เป็นเทคนิคการทำให้เรียบ (Smoothing Technique) แบบหนึ่ง โดยอาศัยสาระของข้อมูลในอดีตมาพยากรณ์ค่าในอนาคตด้วยการเฉลี่ยค่าสังเกตที่ใกล้เวลา ณ จุดนั้น และให้น้ำหนักของแต่ละข้อมูลที่ใช้เฉลี่ยเคลื่อนที่เท่ากันเป็นค่าเฉลี่ยเลขคณิต ใช้แนวความคิดในลักษณะดีเทอร์มินิสติค (Deterministic) ไม่ได้ใช้กระบวนการความน่าจะเป็น (Stochastic Process) เหมือนของอนุกรมเวลาบ๊อกซ์ - เจนกินส์ เทคนิคการเฉลี่ยเคลื่อนที่มีหลายลำดับขึ้นอยู่กับลักษณะการเปลี่ยนแปลงของข้อมูล ก่อนที่จะกล่าวถึงการเฉลี่ยเคลื่อนที่ซ้ำสองครั้ง จะกล่าวถึงการเฉลี่ยเคลื่อนที่แบบธรรมดาก่อนจากค่าสังเกต $X_1 , X_2, \ldots . X_N$ กำหนดรูปแบบเป็น

$$X_t = a + u_t \quad ; t = 1 , 2 , \ldots\ldots , N$$

เมื่อ $X_t$ = ค่าสังเกต ณ เวลา $t$

$a$ = พารามิเตอร์

$u_t$ = ค่าความคลาดเคลื่อน ณ เวลา $t$

รูปแบบนี้ข้อมูลกระจายอยู่ที่ค่าคงที่ค่าหนึ่ง สูตรในการหาค่าเฉลี่ยเคลื่อนที่ได้แก่

$$M_t = \frac{X_t + X_{t-1} + \text{-------} + X_{t-NM+1}}{NM}$$

หรือ $$M_t = M_{t-1} + \frac{X_t - X_{t-NM}}{NM}$$

เมื่อ $M_t$ = ค่าเฉลี่ยเคลื่อนที่ ณ เวลา $t$

$NM$ = จำนวนเทอมของการเฉลี่ยเคลื่อนที่

ค่าพยากรณ์ 1 หน่วยเวลาล่วงหน้า $(\hat{X}_{t+1})$ ใช้ $M_t$ เป็นค่าพยากรณ์ ค่า

ความคลาดเคลื่อนจากการพยากรณ์ ณ เวลา t เท่ากับ $M_t - X_{t+1}$ วิธีการเฉลี่ยเคลื่อนที่แบบธรรมดาเหมาะสำหรับข้อมูลที่ไม่ค่อยมีการเปลี่ยนแปลงจากค่าคงที่ แต่การเปลี่ยนแปลงของข้อมูลมีแนวโน้ม ควรใช้การเฉลี่ยเคลื่อนที่ซ้ำสองครั้งซึ่งมีรูปแบบ

$$X_t = a_0 + a_1 t + u_t$$

มีพารามิเตอร์เป็น $a_0 + a_1$ รูปแบบนี้คำสังเกตมีแนวโน้มด้วยเมื่อเวลาผ่านไป การคำนวณเหมือนการเฉลี่ยเคลื่อนที่แบบธรรมดา โดยเฉลี่ยเคลื่อนที่ครั้งที่สองจากค่าเฉลี่ยเคลื่อนที่ในครั้งแรก

$$M_t^{(2)} = \frac{M_t + M_{t-1} + \text{-----} + M_{t-NM+1}}{NM}$$

หรือ $$M_t^{(2)} = M_{t-1}^{(2)} + \frac{M_t - M_{t-NM}}{NM}$$

สามารถเขียน $M_t$ , $M_t^{(2)}$ อยู่ในเทอมของ $a_0$ , $a_1$ ได้ดังนี้

$$M_t = a_0 + (t - \frac{NM - 1}{2}) a_1$$

$$M_t^{(2)} = a_0 + (t - NM + 1) a_1$$

ค่าประมาณของ $a_o$ , $a_1$ ได้แก่

$$\hat{a}_o(t) = 2 M_t - M_t^{(2)}$$

$$\hat{a}_1(t) = \frac{2}{NM-1} (M_t - M_t^{(2)})$$

ค่าพยากรณ์ 1 หน่วยเวลาล่วงหน้าได้แก่

$$\hat{X}_{t+1} = \hat{a}_o(t) + \hat{a}_1(t)$$

และค่าพยากรณ์ j หน่วยเวลาล่วงหน้า ณ เวลา t ได้แก่

$$\hat{X}_{t+j} = \hat{a}_o(t) + \hat{a}_1(t) \cdot j$$

ปัญหาในการเฉลี่ยเคลื่อนที่คือควรจะใช้จำนวนเทอมในการเฉลี่ยเท่าไรจึงจะเหมาะสม จำนวนเทอมในการเฉลี่ยนี้ขึ้นกับลักษณะของข้อมูล ถ้าข้อมูลมีการเคลื่อนไหวน้อย NM ก็ควรใช้มาก แต่ถ้าข้อมูลเปลี่ยนแปลงเร็วก็ควรใช้ NM น้อย เนื่องจากการเปลี่ยนแปลงต้องปรับอย่างรวดเร็ว ในทางปฏิบัติจะเลือกค่า NM ที่ทำให้เกิดความคลาดเคลื่อนต่ำที่สุดไม่ว่าจะใช้เกณฑ์ความคลาดเคลื่อนกำลังสองหรือค่าสัมบูรณ์ของความคลาดเคลื่อน การใช้เกณฑ์ทั้งสองอาจ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ให้ค่า NM ที่ต่างกัน ในงานวิจัยนี้จะลองค่า NM ตั้งแต่ 2 ถึง 20 เพราะถ้าใช้ NM มากเกินไปแล้วจะทำให้เหลือจำนวนเทอมในการพยากรณ์น้อย การเปรียบเทียบระหว่างเทคนิคการพยากรณ์หรือการเลือก NM ตัวแบบที่ได้จะมาจากช่วงเวลาที่น้อยเกินไป โดยจำนวนเทอมที่เหลือในการเทียบค่าจริงกับค่าพยากรณ์จะเท่ากับ N - 2×NM

<u>การทำให้เรียบแบบเอกซ์โปแนนเชียลซ้ำสองครั้ง</u> (Double Exponential Smoothing)

การทำให้เรียบแบบเอกซ์โปแนนเชียลมีแนวความคิดต่างจากการเฉลี่ยเคลื่อนที่ โดยการเฉลี่ยเคลื่อนที่ให้ความสำคัญของข้อมูลที่ใช้เฉลี่ยเท่ากันหมด แต่การทำให้เรียบแบบเอกซ์โปแนนเชียลจะให้ความสำคัญแก่ข้อมูลที่อยู่ใกล้กับเวลาที่จะพยากรณ์มากกว่าและให้น้ำหนักข้อมูลก่อนหน้านี้ลดลงเรื่อย ๆ แบบเรขาคณิต การทำให้เรียบแบบเอกซ์โปแนนเชียลมีข้อสมมติว่าอนุกรมเวลาเป็นโพลิโนเมียลฟังค์ชั่นของเวลา

$$X_t = \sum_{p=0}^{P} a_p t^p + u_t$$

เมื่อ $X_t$ = ค่าสังเกต ณ เวลา

$a_p$ = พารามิเตอร์ของตัวแบบ

$p$ = ลำดับของโพลิโนเมียล

$u_t$ = ค่าความคลาดเคลื่อน

การทำให้เรียบแบบเอกซ์โปแนนเชียลครั้งเดียวเป็นตัวแบบโพลิโนเมียลลำดับที่ศูนย์

$$X_t = a_0 + u_t$$

ค่าทำให้เรียบแบบเอกซ์โปแนนเชียลได้แก่

$$S_t = a X_t + (1 - \alpha) S_{t-1}$$

หรือ $S_t = S_{t-1} + \alpha (X_t - S_t)$

เมื่อ $S_t$ = ค่าทำให้เรียบแบบเอกซ์โปแนนเชียล ณ เวลา $t$

$\alpha$ = น้ำหนักของการเฉลี่ยเป็นค่าคงที่ โดย $0 < \alpha < 1$

สมการ $S_t$ อยู่ในรูปของ Recursive Formular จะทราบค่า $S_t$ ได้ต้องทราบค่า $S_{t-1}$ ย้อนขึ้นไปเรื่อย ๆ จึงเกิดปัญหาในการกำหนดค่าเริ่มต้น $S_0$ ได้ มีผู้เสนอแนวความคิดในการกำหนดค่าเริ่มต้นหลายวิธี ในงานวิจัยนี้จะใช้ค่าเริ่มต้นตามวิธีของ Holt ซึ่งใช้ข้อมูลตัวแรกเป็นค่าเริ่มต้น ใช้ $S_t$ เป็นค่าพยากรณ์ 1 หน่วยเวลาล่วงหน้า ($\hat{X}_{t+1}$)

การทำให้เรียบแบบเอกซ์โปแนนเชียลครั้งเดียวเป็นตัวแบบที่ไม่มีแนวโน้ม ตัวแบบที่มีแนวโน้มเป็นตัวแบบโฮลิโนเมียลลำดับที่เรียกว่าการทำให้เรียบแบบเอกซ์โปเนนเชียลซ้ำสองครั้ง ซึ่งมีรูปแบบเป็น

$$X_t = a_o + a_1 t + U_t$$

$$S_t^{(2)} = \alpha S_t + (1 - \alpha) S_{t-1}^{(2)}$$

เมื่อ $S_t$ , $S_t^{(2)}$ เป็นค่าทำให้เรียบแบบเอกซ์โปเนนเชียลครั้งเดียวและซ้ำสองครั้ง ณ เวลา t

การประมาณค่าพารามิเตอร์ $a_o$ , $a_1$ สามารถเขียนอยู่ในรูปของการทำให้เรียบแบบเอกซ์โปเนนเชียลได้ดังนี้

$$\hat{a}_0(t) = 2 S_t - S_t^{(2)}$$

$$\hat{a}_1 (t) = \frac{\alpha}{1 - \alpha} (S_t - S_t^{(2)})$$

ค่าพยากรณ์ 1 หน่วยเวลาล่วงหน้า ณ เวลา t ได้แก่

$$X_{t+1} = \hat{a}_o (t) + \hat{a}_1 (t)$$

และค่าพยากรณ์ j หน่วยเวลาล่วงหน้า ณ เวลา t ได้แก่

$$\hat{X}_{t+j} = \hat{a}_o (t) + \hat{a}_1 (t) \cdot j$$

นอกเหนือจากปัญหาการกำหนดค่าเริ่มต้นแล้ว การทำให้เรียบแบบเอกซ์โปเนนเชียลยังมีปัญหาในการกำหนดค่า $\alpha$ ว่าควรให้น้ำหนักเท่าไรจึงจะเหมาะสม ถ้าข้อมูลมีความสัมพันธ์กับคาบใกล้ ๆ กันมากค่า $\alpha$ ควรใกล้ 1 แต่ถ้าข้อมูลมีความสัมพันธ์กับคาบใกล้ ๆ กัน น้อยค่า $\alpha$ ควรใกล้ศูนย์ การกำหนดค่า $\alpha$ จะใช้วิธีการค้นหา (Search) โดยลองค่า $\alpha$ ตั้งแต่ 0.01 ถึง 0.99 แล้วเลือกค่าที่ทำให้เกิดความคลาดเคลื่อนต่ำที่สุด

เทคนิคการทำให้เรียบทั้ง 2 วิธี ในงานวิจัยนี้จะศึกษาในลำดับที่ 2 เนื่องจากในทางปฏิบัติข้อมูลที่ใช้ในการพยากรณ์ระยะสั้นส่วนใหญ่มักจะสอดคล้องกับวิธีการทำให้เรียบไม่เกินลำดับที่ 2