บทที่ 2

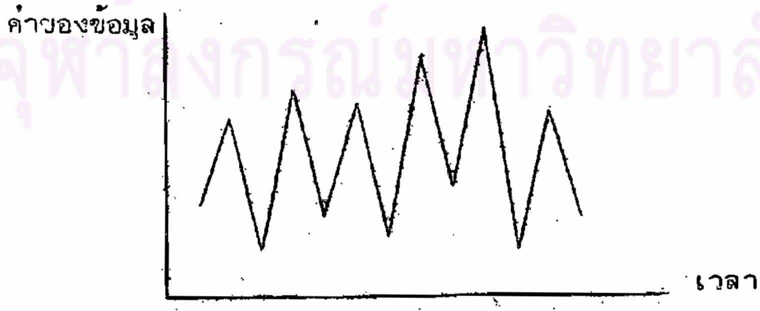
ระเบียบวิธีการทางสถิติที่ใช้ในการวิจัย

ค่าของข้อมูลที่เกิดขึ้นในระยะเวลาต่าง ๆ กัน ซึ่งได้รับการจัดเรียงลำดับโดยคำนึงถึงเวลาแห่งการเกิดขึ้นหรือปรากฏนั้นเป็นเกณฑ์ โดยจุดบอกเวลาจะอยู่ห่างจากกันแต่ละคู่เป็นระยะเท่า ๆ กัน เช่น ยอดขายสินค้าเป็นรายเดือน ข้อมูลนี้เรียกว่าข้อมูลอนุกรมเวลา (Time Series) ซึ่งแบ่งออกได้เป็น 3 ประเภท คือ

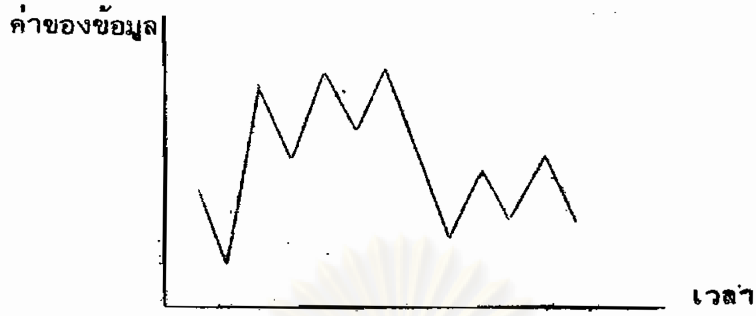
1. อนุกรมเวลาคงที่ (Stationary Time Series) ข้อมูลจะเคลื่อนไหวไปรอบ ๆ ค่าตัวกลาง (Mean) และการเคลื่อนไหวจะเป็นไปในลักษณะคงที่ พอค่าหนึ่งเพิ่มขึ้นค่าที่ตามมาจะลดลงดังรูปที่ 2.1

2. อนุกรมเวลาไม่คงที่ (Non Stationary Time Series) ข้อมูลจะเคลื่อนไหวไม่แน่นอน และจะไม่สามารถหาค่าตัวกลางที่เป็นศูนย์กลางได้ดังรูปที่ 2.2

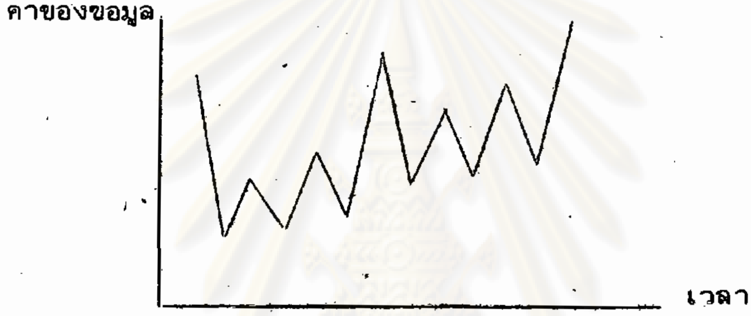
3. อนุกรมเวลาฤดูกาล (Seasonal Time Series) ข้อมูลจะมีการเคลื่อนไหวขึ้นลงตามระยะเวลาเป็นช่วงที่แน่นอน และลักษณะการเคลื่อนไหวในระยะเวลาหนึ่งจะคล้ายหรือเหมือนกันกับช่วงเวลาอื่น ๆ ซ้ำ ๆ กัน ดังรูปที่ 2.3



รูปที่ 2.1 ลักษณะของอนุกรมเวลาคงที่



รูปที่ 2.2 ลักษณะของอนุกรมเวลาไม่คงที่



รูปที่ 2.3 ลักษณะของอนุกรมเวลาฤดูกาล

ในการวิเคราะห์อนุกรมเวลา (Time Series Analysis) แบ่งเป็น 2 ส่วนใหญ่ คือ การวิเคราะห์หารูปแบบ (Models) ที่เหมาะสมกับข้อมูลในอดีต เพื่อนำมาใช้คาดคะเนค่าในอนาคต (Forecast - Values) อันเป็นจุดประสงค์แรกและสำคัญในการวิเคราะห์ทางสถิติ ในการวิจัยนี้จะได้นำวิธีการของบ็อกซ์และเจนกิ้นส์และวิธีการของ Classical Time Series มาใช้ ในส่วนแรกนี้จะกล่าวถึงขั้นตอนในการวิเคราะห์โดยวิธีการของบ็อกซ์และเจนกิ้นส์

ขั้นตอนการวิเคราะห์ส่วนที่ 1 เป็นการเลือกรูปแบบที่เหมาะสม

ขั้นที่ 1 การกำหนดรูปแบบ (Model Identification)

ขั้นที่ 2 การประมาณค่าพารามิเตอร์ (Parameter Estimation)

ขั้นที่ 3 การทดสอบความเหมาะสมของรูปแบบ (Model Diagnostic Checking)

ส่วนที่ 2 จะเป็นการนำรูปแบบที่เหมาะสมนั้นมาหาค่าคาดคะเนในอนาคตต่อไป

2.1 การกำหนดรูปแบบ

เป็นวิธีการขั้นแรกในการวิเคราะห์ รูปแบบที่จะนำมาพิจารณาว่ามีความเหมาะสมกับอนุกรมเวลาแต่ละชุดนั้นมีเป็นจำนวนมากแบ่งเป็น 3 ประเภทใหญ่ ๆ คือ Autoregressive Models, Moving Average Models และ Autoregressive - Moving Average Models ใช้สัญญลักษณ์แทนรูปแบบ คือ ARIMA

อนุกรมเวลาคงที่

Autoregressive Model อันดับที่ P ใช้สัญญลักษณ์ AR (P)

$$\dot{Y}_t = \phi_1\dot{Y}_{t-1} + \phi_2\dot{Y}_{t-2} + \cdots + \phi_p\dot{Y}_{t-p} + e_t$$

โดยที่ $\dot{Y}_t = Y_t - \mu$

Moving Average Model อันดับที่ q ใช้สัญญลักษณ์ MA (q)

$$\dot{Y}_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Autoregressive - Moving Average Model อันดับที่ p และ q ใช้สัญญลักษณ์ ARMA (p,q)

$$\dot{Y}_t - \phi_1\dot{Y}_{t-1} - \cdots - \phi_p\dot{Y}_{t-p} = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

ถ้าเป็นอนุกรมเวลาไม่คงที่ จะต้องทำให้เป็นอนุกรมเวลาคงที่โดยการหาผลต่าง (Differencing) ดังนี้

ผลต่างครั้งที่ 1 : $D^1Y_t = DY_t = Y_t - Y_{t-1}$

ผลต่างครั้งที่ 2 : $D^2Y_t = D(DY_t) = D(Y_t - Y_{t-1}) = DY_t - DY_{t-1}$

$$= Y_t - 2Y_{t-1} + Y_{t-2}$$

ผลต่างครั้งที่ 3 : $D^3Y_t = D(D^2Y_t) = Y_t - 3Y_{t-1} + 3Y_{t-2} - Y_{t-3}$

.
.
.

ผลต่างครั้งที่ d : $D^dY_t = \binom{d}{0}Y_t - \binom{d}{1}Y_{t-1} + \binom{d}{2}Y_{t-2} \cdots\cdots + \binom{d}{d}Y_{t-d}$

เมื่อได้ผลต่างแต่ละครั้งตั้งแต่ครั้งที่ 1,2, ...,d จะต้องพิจารณาว่าอนุกรมนั้นเป็นชนิดคงที่หรือยัง ถ้าเป็นก็จะหยุดเพียงนั้น รูปแบบสำหรับอนุกรมเวลาไม่คงที่ก็จะเป็นลักษณะเดียวกับอนุกรมเวลาคงที่แล้วใช้สัญญลักษณ์ W_t แทน $\dot{Y}_t$ โดย $W_t = Y_t - Y_{t-1}$ สัญญลักษณ์แทนรูปแบบ คือ ARIMA (p,d,q)

ในการวิจัยนี้จะนำรูปแบบมาใช้พิจารณาเพียงระดับที่ p = 2, q = 2 ส่วน d นั้น ในทางปฏิบัติจริง ๆ คงจะไม่เกินครั้งที่ 2 อนุกรมเวลาไม่คงที่จะเป็นชนิดคงที่ได้ ลักษณะของรูปแบบจะเป็นดังนี้

2.1.1 ARIMA (1,0,0) หรือ AR (1)

รูปแบบคือ $\dot{Y}_t = \phi_1\dot{Y}_{t-1} + e_t$ โดยที่ $\dot{Y}_t = Y_t - \mu$, $|\phi_1| < 1$

$t = 1,2,\ldots\ldots N$

เมื่อ Y_t คือข้อมูลอนุกรมเวลาที่ต้องการศึกษา ณ เวลา t

μ, ϕ_1 เป็นพารามิเตอร์ของรูปแบบ

e_t คือค่าความคลาดเคลื่อนของความสัมพันธ์ระหว่างข้อมูล Y_t และ Y_{t-1} หรือเรียกว่า "Shock"

2.1.2 ARIMA (2,0,0) หรือ AR (2)

รูปแบบคือ $\dot{Y}_t = \phi_1 \dot{Y}_{t-1} + \phi_2 \dot{Y}_{t-2} + e_t$ โดยที่ $\phi_2 + \phi_1 < 1$

$\phi_2 - \phi_1 < 1$

$-1 < \phi_2 < 1$

2.1.3 ARIMA (0,0,1) หรือ MA (1)

รูปแบบคือ $\dot{Y}_t = e_t - \theta_1 e_{t-1}$ โดยที่ $|\theta_1| < 1$

เมื่อ θ_1 คือค่าพารามิเตอร์แสดง proportion ของ e_{t-1} ที่มีต่อ $\dot{Y}_t$

e_t คือ current shock

e_{t-1} คือ previous shock

2.1.4 ARIMA (0,0,2) หรือ MA (2)

รูปแบบคือ $\dot{Y}_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$ โดยที่ $\theta_2 + \theta_1 < 1$

$\theta_2 - \theta_1 < 1$

$-1 < \theta_2 < 1$

2.1.5 ARIMA (1,0,1) หรือ ARMA (1,1)

รูปแบบคือ $\dot{Y}_t - \phi_1 \dot{Y}_{t-1} = e_t - \theta_1 e_{t-1}$ โดยที่ $|\phi_1| < 1$, $|\theta_1| < 1$

2.1.6 ARIMA (1,1,0) หรือ ARI (1,1)

รูปแบบคือ $W_t = \phi_1 W_{t-1} - e_t$ โดยที่ $W_t = Y_t - Y_{t-1}$, $|\phi_1| < 1$

2.1.7 ARIMA (2,1,0) หรือ ARI (2,1)

รูปแบบคือ $W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} - e_t$ โดยที่ $\phi_2 + \phi_1 < 1$

$\phi_2 - \phi_1 < 1$

$-1 < \phi_2 < 1$

2.1.8 ARIMA (0,1,1) หรือ IMA (1,1)

รูปแบบคือ $W_t = e_t - \theta_1 e_{t-1}$ โดยที่ $|\theta_1| < 1$

2.1.9 ARIMA (0,1,2) หรือ IMA (1,2)

รูปแบบคือ $W_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$ โดยที่ $\theta_2 + \theta_1 < 1$

$\theta_2 - \theta_1 < 1$

$-1 < \theta_2 < 1$

2.1.10 ARIMA (1,1,1)

รูปแบบคือ $W_t - \phi_1 W_{t-1} = e_t - \theta_1 e_{t-1}$ โดยที่ $|\phi_1| < 1$, $|\theta_1| < 1$

2.1.11 ARIMA (0,1,0) หรือ Random Walk Process

รูปแบบคือ $Y_t - Y_{t-1} = e_t$

ทั้งนี้แต่ละรูปแบบอยู่ภายใต้ข้อสมมติดังนี้

1. $E(e_t) = 0$, Variance $(e_t) = \sigma_e^2$ ทุกค่าของ t
2. $E(e_t e_{t'}) = 0$ สำหรับ $t \neq t'$
3. $E(e_t Y_{t'}) = 0$ สำหรับ $t' < t$

ค่าสถิติที่จะคำนวณเพื่อใช้ในการเลือกรูปแบบก็คือ ค่าสัมประสิทธิ์ออโตคอริเลชั่นของตัวอย่าง (Sample Autocorrelation coefficient - r_k) มีสูตรดังนี้

$$2.1.12 \quad r_k = \frac{\frac{1}{N-k}\sum_{t=1}^{N-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\frac{1}{N}\sum_{t=1}^{N}(Y_t - \bar{Y})^2}$$

$$= \frac{\sum_{t=1}^{N-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{N}(Y_t - \bar{Y})^2}$$

โดยที่ Y_t คือข้อมูลอนุกรมเวลา $Y_1, Y_2 \ldots, Y_N$

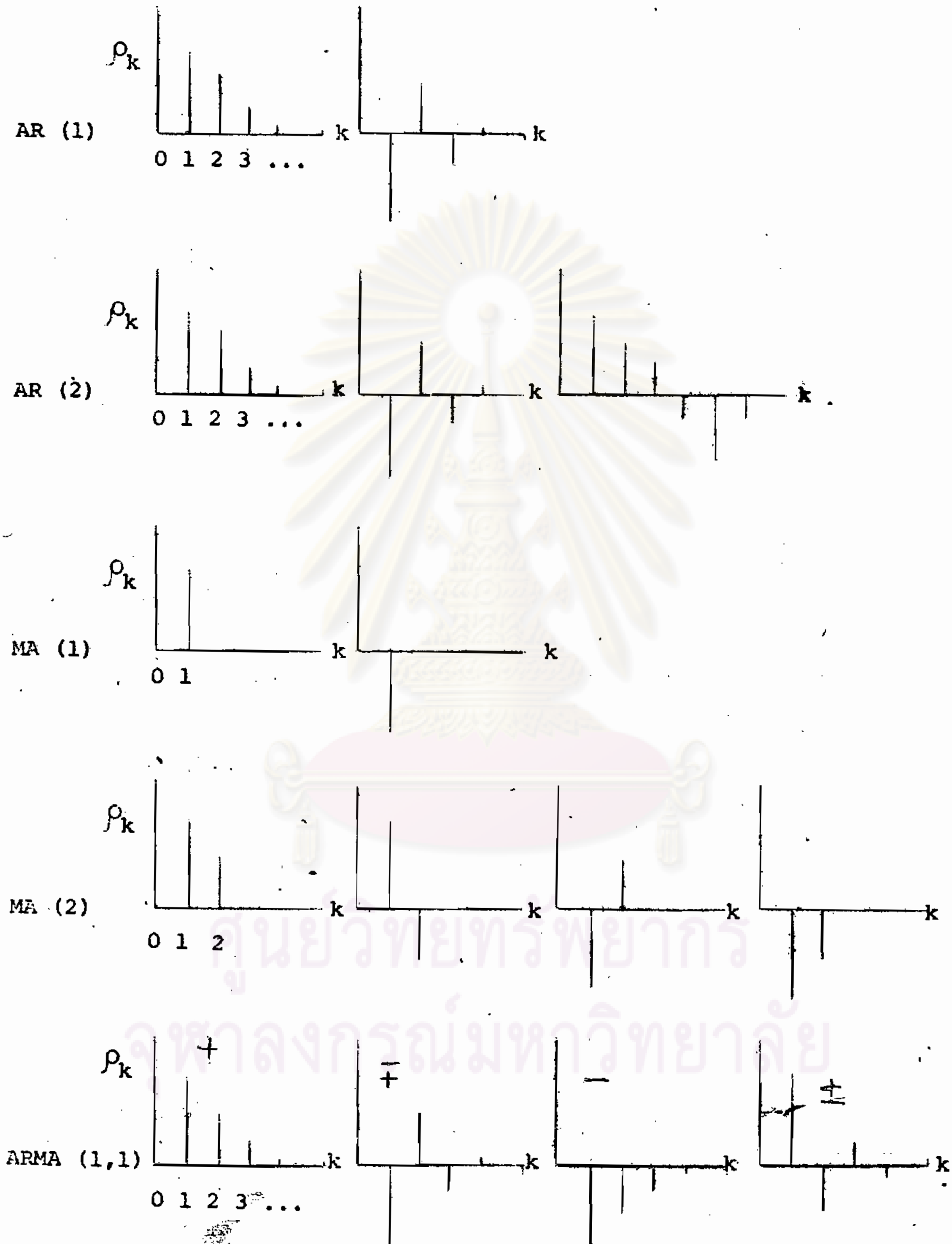
$$\bar{Y} = \frac{1}{N}\sum_{t=1}^{N} Y_t$$

และ $k = 1,2, \ldots, K$ (นิยมให้ $K < N/4$)

ในการวิเคราะห์การถดถอย (Regression Analysis)

ของตัวอย่าง (Sample Correlation coefficient) เพื่อดูความสัมพันธ์ของข้อมูลอย่างน้อย 2 ชุดขึ้นไป เช่น ข้อมูล X กับข้อมูล Y แต่สำหรับ r_k ดังสมการ 2.1.12 จะเป็นการแสดงความสัมพันธ์เชิงเส้นตรงของข้อมูลแต่ละคู่ภายในอนุกรมเวลาชุดเดียวกัน คือ ระหว่าง Y_t กับ Y_{t+k} โดย $t = 1,2,\ldots, N-k$ เมื่อ $k = 1,2,\ldots,K$ ซึ่งในการคำนวณค่า r_k นี้จะสะดวกรวดเร็วและถูกต้องโดยใช้เครื่องคอมพิวเตอร์ จากค่า r_k นี้นำไปพิจารณาว่าอนุกรมเวลานั้น ๆ เป็นอนุกรมเวลาคงที่หรือไม่ โดยถ้าอนุกรมเวลาคงที่ค่า r_k จะเปลี่ยนแปลงอย่างรวดเร็วเมื่อ $k = 1,2\ldots, K$ หากค่า r_k มีการเปลี่ยนแปลงทีละน้อยแสดงว่าเป็นอนุกรมเวลาไม่คงที่ ต้องทำให้เป็นอนุกรมเวลาคงที่โดยการหาผลต่าง

(ดังกล่าวข้างต้น) แล้วคำนวณหาค่า r_k และพิจารณาเช่นเดียวกัน ทำจนกว่าจะได้เป็นอนุกรมเวลาคงที่ นำค่า r_k นี้ไปสร้างกร๊าฟโดยให้ k ; $k = 1,2,\ldots\ldots,K$ เป็นแกนนอน ให้ค่า r_k เป็นแกนตั้ง เพื่อนำไปเปรียบเทียบกับกร๊าฟของค่าสัมประสิทธิ์ออโตคอริเลชั่น (Theoritical Autocorrelation Coefficient; ρ_k) ทั้ง 5 แบบ ดังรูปที่ 2.4 ว่าสอดคล้องกับแบบใด ซึ่งอาจจะสอดคล้องมากกว่า 1 แบบก็ได้ แสดงว่ารูปแบบที่จะเลือกใช้ได้มีมากกว่า 1 รูปแบบ แต่อย่างไรก็ตามจะต้องคำนึงถึงความสะดวกของรูปแบบนั้นด้วย ถ้ามีความเหมาะสมพอ ๆ กัน ก็ควรจะเลือกใช้รูปแบบที่มีจำนวนพารามิเตอร์น้อยที่สุดเพื่อสะดวกและถูกต้องในการคำนวณต่อไป



แสดงลักษณะ fnc. ของ ρ_k ; $(k \geqslant 0)$ ของอนุกรมเวลาคงที่

สัมประสิทธิ์ออโตคอรีเลชั่น (Theoritical Autocorrelation Coefficient) สัญญลักษณ์ คือ ρ_k นี้ในทางปฏิบัติไม่สามารถคำนวณหาค่าได้มีสูตรดังนี้

$$2.1.13 \quad \rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)\ Var(Y_{t+k})}} = \frac{\gamma_k}{\sqrt{\gamma_0^2}} = \frac{\gamma_k}{\gamma_0} \quad \text{ทุก ๆ ค่าของ } t$$

ฉะนั้นจะได้ใช้ค่า r_k เป็นค่าประมาณของ $\rho_k) = \hat{\rho}_k$

เมื่อพิจารณา r_k เปรียบเทียบกับ ρ_k ในรูปที่ 2.4 เมื่อ $k = 1,2,\ldots,K$ เลือกรูปแบบที่เหมาะสมในขั้นต้นได้ จากรูปแบบนี้เราสามารถหาค่าประมาณเบื้องต้นของพารามิเตอร์ (Initial Parameters Estimation) ทั้งนี้โดยอาศัยความสัมพันธ์ระหว่าง ρ_k กับพารามิเตอร์ของแต่ละรูปแบบนั้น ๆ หรือเปิดตาราง[2] A,B,C,D ซึ่งแสดงค่า ρ_k กับค่าพารามิเตอร์ของรูปแบบ MA(1), AR(2), MA(2) และ ARMA(1,1) ตามลำดับ ทั้งนี้จะใช้ค่า r_k แทนค่า ρ_k จากสมการ 2.1.13 ใช้หาค่าความสัมพันธ์ระหว่าง ρ_k และพารามิเตอร์ของแต่ละรูปแบบในอนุกรมเวลาคงที่ได้ดังนี้

$$AR(1) : \dot{Y}_t = \phi_1 \dot{Y}_{t-1} + e_t \quad \text{โดยที่} \quad \dot{Y}_t = Y_t - \mu, \ |\phi_1| < 1$$

$$\begin{aligned}
\gamma_0 = E\left[\dot{Y}_t \dot{Y}_t\right] &= E\left[\phi_1 \dot{Y}_{t-1} + e_t\right]^2 \\
&= E\left[\phi_1^2 \dot{Y}_{t-1} + 2\phi_1 \dot{Y}_{t-1} e_t + e_t^2\right] \\
&= \phi_1^2 E\left[\dot{Y}_{t-1}^2\right] + 2\phi_1 E\left[\dot{Y}_{t-1} e_t\right] + E\left[e_t^2\right] \\
&= \phi_1^2 \gamma_0 + \sigma_e^2 \quad \text{โดยที่ } E\left[\dot{Y}_{t-1} e_t\right] = 0 \text{ จาก}
\end{aligned}$$

สมมุติฐานข้อที่ 3

---

[2] ใน Part V : Collection of Tables and Charts หน้า 517 - 520 Box, G.E., P. and G.M. Jenkins, Time Series Analysis Forecasting and Control

$$\gamma_0(1-\phi_1^2) = \sigma_e^2$$

$$\gamma_0 = \sigma_e^2 \Big/ (1-\phi_1^2)$$

$$\gamma_k = \gamma_{-k} = E\left[\dot{Y}_t \dot{Y}_{t-k}\right] \qquad ; k \geq 1$$

$$= E\left[(\phi_1 \dot{Y}_{t-1} + e_t)\dot{Y}_{t-k}\right]$$

$$= E\left[\phi_1 \dot{Y}_{t-1}\dot{Y}_{t-k} + \dot{Y}_{t-k}e_t\right]$$

$$= \phi_1 E\left[\dot{Y}_{t-1}\dot{Y}_{t-k}\right] + E\left[\dot{Y}_{t-k}e_t\right]$$

ถ้า $k = 1; \gamma_1 = \gamma_{-1} = \phi_1 E\left[\dot{Y}_{t-1}\dot{Y}_{t-1}\right] + E\left[\dot{Y}_{t-1}e_t\right]$

$$= \phi_1 \gamma_0$$

$$= \phi_1 \sigma_e^2 \Big/ (1-\phi_1^2)$$

$k = 2; \gamma_2 = \gamma_{-2} = \phi_1 E\left[\dot{Y}_{t-1}\dot{Y}_{t-2}\right] + E\left[\dot{Y}_{t-2}e_t\right]$

$$= \phi_1 E\left[\phi_1 \dot{Y}_{t-2} + e_{t-1}\right]\dot{Y}_{t-2}$$

$$= \phi_1 E\left[\phi_1 \dot{Y}_{t-2}\dot{Y}_{t-2}\right] + \phi_1 E\left[\dot{Y}_{t-2}e_{t-1}\right]$$

$$= \phi_1^2 \gamma_0$$

นั่นคือ $\gamma_k = \gamma_{-k} = \phi_1^k \gamma_0 \qquad ; k \geq 1$

และ $\rho_k = \rho_{-k} = \dfrac{\gamma_k}{\gamma_0} = \phi_1^k \qquad ; k \geq 1$

ในทำนองเดียวกันจะได้ว่า

AR (2) : $\rho_1 = \dfrac{\phi_1}{1-\phi_2}$, $\rho_2 = \phi_2 + \dfrac{\phi_1^2}{1-\phi_2}$

และ $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$ เมื่อ $k \geq 3$

MA (1) : $$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$$

และ $$\rho_k = 0 \qquad \text{เมื่อ } k > 1$$

MA (2) : $$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

และ $$\rho_k = 0 \qquad \text{เมื่อ } k > 2$$

ARMA (1,1) : $$\rho_k = \frac{(\phi_1-\theta_1)(1-\phi_1\theta_1)\ \phi_1^{k-1}}{1-2\phi_1\theta_1+\theta_1^2} \qquad \text{เมื่อ } k \geqslant 1$$

แทนค่า ρ_k ด้วยค่า r_k ลงในรูปแบบที่เหมาะสม คำนวณหาค่าประมาณเบื้องต้นของพารามิเตอร์ จะได้รูปแบบตามต้องการ ซึ่งจะได้นำไปหาค่าประมาณพารามิเตอร์ที่แน่นอนในขั้นที่ 2 ต่อไป

2.2 การประมาณค่าพารามิเตอร์

จากขั้นที่ 1 ในข้อ 2.1 จะได้รูปแบบที่เหมาะสมและค่าประมาณเบื้องต้นของพารามิเตอร์ ในขั้นนี้จะเป็นการหาค่าประมาณของพารามิเตอร์ที่ดีที่สุด ซึ่งในอนุกรมเวลา คือ ค่าประมาณกำลังสองน้อยที่สุด (Least Square Estimates) คำนวณโดยให้ได้ผลบวกกำลังสองของส่วนเบี่ยงเบนน้อยที่สุด (Minimize the Sum of Square Deviatious) เพื่อจะได้รูปแบบที่เหมาะสมยิ่งขึ้น สำหรับใช้หาค่าคาดหมาย (Fitted Values)

จากรูปแบบทั่วไป ARIMA (p,d,q) ประกอบด้วยพารามิเตอร์จำนวน p+q+1 ตัว

คือ $\underset{\sim}{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)$, $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)$ และ μ

นั่นคือ $S(\mu, \underset{\sim}{\phi}, \underset{\sim}{\theta}) = \sum_{t=1}^{N} \left[Y_t - E(Y_t \mid \text{previous } Y\text{'s}) \right]^2$

หรือ $\sum_{t=1}^{N} \left[W_t - E(W_t \mid \text{previous } w\text{'s}) \right]^2$

โดยที่ Y_t คือ ข้อมูล ณ เวลา t ที่นำมาวิเคราะห์

$E\left[Y_t \mid \text{previous } Y's\right] = \hat{Y}_t$ คือ ค่าคาดหมายของ Y_t ในเวลาที่ t

จะเลือกค่าพารามิเตอร์ที่ให้ค่า $S(\mu, \underset{\sim}{\emptyset}, \underset{\sim}{\theta})$ น้อยที่สุด เพราะถ้า $S(\mu, \underset{\sim}{\emptyset}, \underset{\sim}{\theta})$ น้อยที่สุดแสดงว่า $\hat{Y}_t$ มีส่วนเบี่ยงเบนจาก Y_t น้อยที่สุดในทุก ๆ เวลาที่ t การคำนวณหา $S(\mu, \underset{\sim}{\emptyset}, \underset{\sim}{\theta})$ ที่น้อยที่สุดนี้ จะคำนวณโดยใช้เครื่องคอมพิวเตอร์ $S(\mu, \underset{\sim}{\emptyset}, \underset{\sim}{\theta})$ แต่ละรูปแบบ คือ

AR (1) : $Y_t - \mu = \emptyset_1(Y_{t-1} - \mu) + e_t$

$$Y_t = \mu + \emptyset_1(Y_{t-1} - \mu) + e_t$$

$$E\left[Y_t \mid \text{previous } Y's\right] = E\left[Y_t \mid Y_{t-1}\right] = \mu + \emptyset_1(Y_{t-1} - \mu) + E(e_t)$$

$$= \mu + \emptyset_1(Y_{t-1} - \mu)$$

จากสมมติฐานข้อ 1 $E(e_t) = 0$

$$\therefore S(\mu,\emptyset_1) = \sum_{t=1}^{N}\left[Y_t - \mu - \emptyset_1(Y_{t-1} - \mu)\right]^2$$

$$= \sum_{t=2}^{N}\left[Y_t - \mu - \emptyset_1(Y_{t-1} - \mu)\right]^2$$

ทั้งนี้เพราะว่าเมื่อเริ่ม t=1 นั้น $Y_{t-1} = Y_0$ ซึ่งไม่มีค่าจริง

คำนวณหา $S(\mu, \emptyset_1)$ โดยแทนค่า Y_t, Y_{t-1}, μ, $\emptyset_1$

AR (2) : $S(\mu,\emptyset_1,\emptyset_2) = \sum_{t=3}^{N}\left[Y_t - \mu - \emptyset_1(Y_{t-1} - \mu) - \emptyset_2(Y_{t-2} - \mu)\right]^2$

MA (1) : $S(\mu,\theta_1) = \sum_{t=2}^{N}\left[Y_t - \mu + \theta_1 e_{t-1}\right]^2$

หรือ $= \sum_{t=2}^{N} e_t^2$

$$MA\ (2) : S(\mu,\theta_1,\theta_2) = \sum_{t=3}^{N}\left[Y_t - \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2}\right]^2$$

$$\text{หรือ} = \sum_{t=3}^{N} e_t^2$$

$$ARMA\ (1,1) : S(\mu,\phi_1,\theta_1) = \sum_{t=2}^{N}\left[Y_t - \mu - \phi_1(Y_{t-1} - \mu) + \theta_1 e_{t-1}\right]^2$$

$$ARI\ (1,1) : W_t = \phi_1 W_{t-1} + e_t$$

โดยที่ $W_t = Y_t - Y_{t-1}$ และ $t = 2,3,\ldots\ldots,N$

$$E\left[W_t \mid \text{previous } W's\right] = E\left[W_t \mid W_{t-1}\right] = \phi_1 W_{t-1} + E(e_t)$$

$$= \phi_1 W_{t-1}$$

โดยข้อสมมติ $E(e_t) = 0$

$$S(\mu,\phi_1) = \sum_{t=2}^{N}\left[W_t - \phi_1 W_{t-1}\right]^2$$

$$ARI\ (2,1) : S(\mu,\phi_1,\phi_2) = \sum_{t=3}^{N}\left[W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2}\right]^2$$

$$IMA\ (1,1) \quad S(\mu,\theta_1) = \sum_{t=2}^{N}\left[W_t + \theta_1 e_{t-1}\right]^2$$

$$IMA\ (2,1) \quad S(\mu,\theta_1,\theta_2) = \sum_{t=3}^{N}\left[W_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}\right]^2$$

$$ARIMA\ (1,1,1) : S(\mu,\phi_1,\theta_1) = \sum_{t=2}^{N}\left[W_t - \phi_1 W_{t-1} + \theta_1 e_{t-1}\right]^2$$

เมื่อคำนวณได้ค่า $S(\mu, \phi, \underset{\sim}{\theta})$ ที่น้อยที่สุดของอนุกรมเวลาแต่ละชุดแล้ว นำค่าพารามิเตอร์หรือค่าประมาณกำลังสองน้อยที่สุด $(\hat{\mu}, \hat{\phi}, \hat{\underset{\sim}{\theta}})$ ที่ได้นั้นไปแทนลงในรูปแบบนั้น ๆ ก็จะได้รูปแบบที่เหมาะสม (Fitted Model) ในขั้นต้น และจากรูปแบบนี้ก็นำไปหาค่าคาดหมายของ $Y_t; t=1,2,\ldots\ldots,N$ (Fitted values หรือ Estimated of Y_t) ใช้สัญลักษณ์ $\hat{Y}_t$ การพิจารณาดูว่ารูปแบบเหมาะหรือไม่ อย่างคร่าว ๆ ในขั้นนี้ก็โดยการสร้างกราฟ

เปรียบเทียบระหว่าง Y_t (Observed Values) และ $\hat{Y}_t$ (Fitted Values) ค่าทั้งสองอยู่ในแกนตั้ง ช่วงเวลา t; t=1,2,.....,N เป็นแกนนอนจะได้นำค่าผลต่างระหว่าง Y_t และ $\hat{Y}_t$ นี้ไปใช้ในการทดสอบความเหมาะสมของรูปแบบอย่างละเอียดต่อไปในขั้นที่ 3

2.3 การทดสอบความเหมาะสมของรูปแบบ

ในขั้นนี้จะเป็นการทดสอบว่ารูปแบบที่เลือกนั้นมีความถูกต้อง และเหมาะสมกับข้อมูลชุดนั้นหรือไม่ ถ้าหากว่าไม่จะต้องหาวิธีวิเคราะห์ว่า รูปแบบลักษณะไหนจึงจะนำมาใช้ได้ดีต่อไป ค่าสถิติที่จะใช้ในการทดสอบก็คือ ค่าความคลาดเคลื่อน (Residuals) สัญญลักษณ์ คือ $\hat{e}_t$ ซึ่งเป็นผลต่างระหว่างค่าข้อมูลจริงกับค่าคาดหมาย เท่ากับ $Y_t - \hat{Y}_t$ จะได้จากขั้นที่ 2 โดยที่รูปแบบที่เหมาะสมจะต้องมีคุณลักษณะดังนี้

2.3.1 $\hat{e}_t$ จะต้องไม่มีสหสัมพันธ์ต่อกัน (Uncorrelated) ทุกค่า lag k คือ $r_k(\hat{e}) = 0$

2.3.2 $\hat{e}_t$ จะมีการกระจายแบบปกติ (Normal Distribution) ด้วยค่าเฉลี่ย (Mean) = 0 ค่าแปรปรวน (Variance) = σ_e^2

ทั้งนี้จะต้องทดสอบว่า $E(e_t) = 0$

$E(e_t e_{t'}) = 0 \;;\; t \neq t'$ หรือ $r_k(\hat{e}) = 0$

2.3.3 **การทดสอบ** $E(\hat{e}_t) = \bar{\hat{e}}_t = 0$ โดยการเปรียบเทียบค่ามัชฌิมเลขคณิต (Arithmetic Mean) และค่าประมาณของส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของ Resudual นั้น ๆ

ค่ามัชฌิมเลขคณิตของ Residual : $\bar{\hat{e}}_t = \frac{1}{N-1}\sum_{t=2}^{N} \hat{e}_t$

ค่าประมาณของส่วนเบี่ยงเบนมาตรฐานของ Residual : $\frac{\hat{\sigma}_e}{\sqrt{N}}$

$$\hat{\sigma}_e^2 = \frac{\sum_{t=1}^{N} (Y_t - \hat{Y}_t)^2}{N - a} = \frac{\sum_{t=1}^{N} \hat{e}_t^2}{N - a}$$

โดยที่ N คือจำนวนของ Residuals

a คือจำนวนของพารามิเตอร์ในรูปแบบนั้น ๆ

ทั้งนี้ ถ้า $\bar{\hat{e}} < \hat{\sigma}_e / \sqrt{N}$ แล้วแสดงว่า $E(e_t) = 0$

2.2.4 <u>การทดสอบว่า</u> $E(e_t e_{t'}) = 0 \quad t \neq t'$ โดยการหาค่าสัมประสิทธิ์ออโตคอริเลชั่น ของ Residual มีสูตรในการคำนวณดังนี้

$$r_k(\hat{e}) = \frac{\sum_{t=1}^{N-k} \hat{e}_t \hat{e}_{t+k}}{\sum_{t=1}^{N} \hat{e}_t^2}$$

โดยที่ N คือ จำนวนของ Residuals และ k = 1,2,....., K (ตามกฎของ Thumb ให้ $K \leq \sqrt{N}$) และจากทฤษฎีว่าด้วยการกระจายของออโตคอริเลชั่นของ Residual ในรูปแบบ ARIMA ของ Box และ Pierce คือ

2.2.4.1 $E\left[r_k(\hat{e})\right] = 0$,

2.2.4.2 $Var\left[r_k(\hat{e})\right]$ ของแต่ละรูปแบบคือ

AR (1) และ ARI(1,1) : $Var\left[r_1(\hat{e})\right] \approx \frac{\phi_1^2}{N}$

$$Var\left[r_2(\hat{e})\right] \approx \frac{1 - \phi_1^2 + \phi_1^4}{N}$$

$Var\left[r_k(\hat{e})\right] \approx \frac{1}{N}$ โดยที่ $k \geq 3$

AR (2) และ ARI (2,1) : $\mathrm{Var}\left[r_1(\hat{e})\right] \approx \frac{\phi_2^2}{N}$

$$\mathrm{Var}\left[r_2(\hat{e})\right] \approx \frac{\phi_2^2 + \phi_1^2(1 + \phi_2)^2}{N}$$

$$\mathrm{Var}\left[r_k(\hat{e})\right] \approx \frac{1}{N} \qquad \text{โดยที่ } k \geqslant 3$$

MA (1) และ IMA (1,1) : $\mathrm{Var}\left[r_1(\hat{e})\right] \approx \frac{\theta_1^2}{N}$

$$\mathrm{Var}\left[r_2(\hat{e})\right] \approx \frac{1 - \theta_1^2 + \theta_1^4}{N}$$

$$\mathrm{Var}\left[r_k(\hat{e})\right] \approx \frac{1}{N} \qquad \text{โดยที่ } k \geqslant 3$$

MA (2) และ IMA (1,2) : $\mathrm{Var}\left[r_1(\hat{e})\right] \approx \frac{\theta_2^2}{N}$

$$\mathrm{Var}\left[r_2(\hat{e})\right] \approx \frac{\theta_2^2 + \theta_1^2\,(1 + \theta_2)^2}{N}$$

$$\mathrm{Var}\left[r_k(\hat{e})\right] \approx \frac{1}{N} \qquad \text{โดยที่ } k \geqslant 3$$

ARMA (1,1)และ ARIMA (1,1,1) : $\mathrm{Var}\left[r_1(\hat{e})\right] \approx \frac{\theta_1^2}{N}$

$$\mathrm{Var}\left[r_2(\hat{e})\right] \approx \frac{1 - \theta_1^2 + \theta_1^4}{N}$$

$$\mathrm{Var}\left[r_k(\hat{e})\right] \approx \frac{1}{N} \qquad \text{โดยที่ } k \geqslant 3$$

ดำเนินการทดสอบสมมติฐาน $H_0 : r_k(\hat{e}) = 0$ โดยการเปรียบเทียบค่า $r_k(\hat{e})$ กับค่าประมาณเบี่ยงเบนมาตรฐานของ $r_k(\hat{e})$ คือ $\hat{\sigma}_{r_k(\hat{e})}$ ซึ่งเท่ากับ $+\sqrt{\hat{Var}\left[r_k(\hat{e})\right]}$

ทั้งนี้จะยอมรับสมมติฐาน $H_0 : r_k(\hat{e}) = 0$ ถ้า $\left|r_k(\hat{e})\right| \leq Z_{\alpha/2}\ \hat{\sigma}_{r_k(\hat{e})}$ ทุก ๆ ค่า k ณ ระดับนัยสำคัญ α

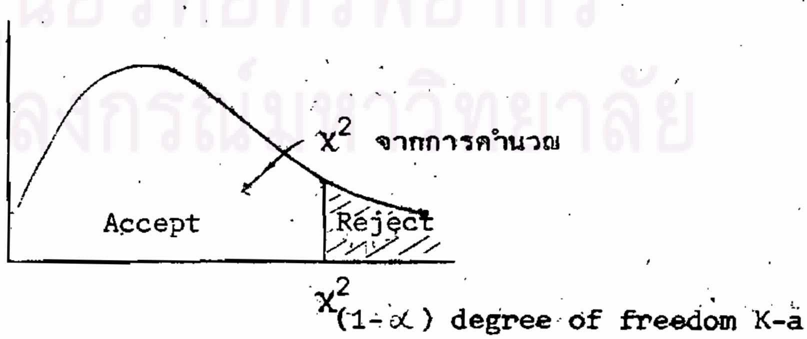
นั่นคือถ้าได้ว่า $r_k(\hat{e}) = 0$ ก็แสดงว่า $\hat{e}_t$ ไม่มีสหสัมพันธ์ต่อกันหรือ $E\left[e_t e_{t'}\right] = 0$

จากการทดสอบในข้อ 2.2.3 และ 2.2.4 ถ้าเป็นไปตามกฎเกณฑ์ดังกล่าวแล้ว รูปแบบที่เลือกใช้สำหรับข้อมูลอนุกรมเวลานั้น จะมีความเหมาะสมเพียงพอที่จะใช้ในการหาค่าคาดคะเนของข้อมูลในอนาคตได้ดี

ยังมีการทดสอบว่ารูปแบบที่เลือกมานั้นมีความเหมาะสมหรือไม่อีกวิธีหนึ่ง คือ การทดสอบโดยใช้ไค - สแควร์ : χ^2 - Test ซึ่งโดยวิธีนี้ $\hat{e}_t$ ไม่จำเป็นต้องมีการกระจายแบบปกติ คำนวณหา χ^2 ได้ดังนี้

$$\chi^2 = N \sum_{k=1}^{K} r_k^2(\hat{e}) \quad ; \text{Degree of Freedom} = K - a$$

ถ้ารูปที่เลือกได้นั้นมีความเหมาะสมกับข้อมูลแล้ว χ^2 จากการคำนวณจะน้อยกว่า $\chi^2_{(1-\alpha)}$ จากตารางสถิติโดยมี degree of freedom = K - a



2.4 การคาดคะเนค่าในอนาคตของอนุกรมเวลา

จะเป็นวิธีการหาค่าคาดคะเนในอนาคต โดยใช้รูปแบบที่เหมาะสมจากวิธีการข้างต้น และจาก Observed data ของข้อมูลอนุกรมเวลานั้น ๆ ค่าคาดคะเนที่จะคำนวณหา คือ $Y_{N+\ell}$ โดยที่ ; ℓ เรียกว่า Lead Time ซึ่ง $\ell = 1,2,\ldots,q$ และ Y_N จะเป็นค่าสุดท้ายของอนุกรมเวลา ซึ่งการคาดคะเนนี้จะใช้ช่วงเวลา $t = N$ เป็นจุดเริ่มต้นของการคาดคะเน (Forecast Origin) จะใช้สัญลักษณ์แทนค่าคาดคะเน $Y_{N+\ell}$ คือ $E_N(Y_{N+\ell})$ หรือ $\hat{Y}_N(\ell)$ โดยที่ N จะแสดงว่าเป็นจุดเริ่มต้นของการคาดคะเน และ ℓ ค่าในวงเล็บจะแสดงถึงช่วงเวลาที่เท่าใดของค่าคาดคะเน

$E_N(Y_{N+\ell})$ หรือ $\hat{Y}_N(\ell)$ จะเป็นค่าคาดคะเนที่ดีที่สุดของ $Y_{N+\ell}$ ก็ต่อเมื่อค่าคาดหมาย ณ เวลาที่ N ของผลต่างกำลังสองของ $\hat{Y}_N(\ell)$ กับ $Y_{N+\ell}$ [Expected (mean) Square Error] มีค่าน้อยมาก : $E_N\left[\hat{Y}_N(\ell) - Y_{N+\ell}\right]^2$

<u>วิธีการหาค่าคาดคะเนที่ดีที่สุดสำหรับรูปแบบ ARIMA โดยทั่ว ๆ ไปมีดังนี้</u>

2.4.1 หาค่า $\hat{Y}_N(\ell)$ ได้จากรูปแบบที่เหมาะสมที่เลือกใช้นั้น ๆ โดยที่ N คือ จุดเริ่มต้นของการคาดคะเน (Forecast Origin) และ ℓ คือ ช่วงเวลาที่ต้องการคาดคะเน (Lead Time)

2.4.2 การคาดหมาย ณ จุดเริ่มต้นของการคาดคะเนที่ N สำหรับเทอมของ Y และค่าความคลาดเคลื่อน สรุปได้ดังนี้

$$E_N(Y_{N+j}) = \begin{cases} Y_N(j) & \text{เมื่อ } j > 0 \quad \text{(เป็นค่าคาดคะเน)} \\ Y_{N+j} & \text{เมื่อ } j \leq 0 \quad \text{(ค่าจากการสังเกต)} \end{cases}$$

$$E_N(e_{N+j}) = \begin{cases} 0 & \text{เมื่อ } j > 0 \quad \text{(เพราะว่าไม่ทราบค่า)} \\ e_{N+j} & \text{เมื่อ } j \leq 0 \quad \text{(เป็นความคลาดเคลื่อนของ } Y_{N+j} - \hat{Y}_{N+j} \text{ เมื่อ } j \leq 0 \text{ เรียกว่า Observed Error)} \end{cases}$$

2.4.3 ค่าพารามิเตอร์จะได้จากการประมาณกำลังสองน้อยที่สุด คือ $\hat{\mu}$, $\hat{\underset{\sim}{\phi}}$, $\hat{\underset{\sim}{\theta}}$ และใช้ $\hat{e}_{N+j}$; $j \leq 0$ เป็นค่าประมาณของ e_{N+j} คำนวณหา $\hat{Y}_N(\ell)$

ฉะนั้นการหาค่าคาดคะเน $\hat{Y}_N(\ell)$; $\ell = 1,2,\ldots,q$ ของแต่ละรูปแบบหาได้ดังนี้

AR (1) : $\dot{Y}_t = \phi_1 \dot{Y}_{t+1} + e_t$

โดยที่ $\dot{Y}_t = Y_t - \mu$ และให้ $t = N + \ell$

$$Y_{N+\ell} = \mu + \phi_1 (Y_{N+\ell-1} - \mu) + e_{N+\ell}$$

โดยการใช้ค่าคาดคะเน ณ จุดเริ่มต้นของการคาดคะเนที่ N

$$\begin{aligned} E_N(Y_{N+\ell}) &= E_N\left[\mu + \phi_1 (Y_{N+\ell-1} - \mu) + e_{N+\ell}\right] \\ &= \hat{\mu} + \hat{\phi}_1 E_N(Y_{N+\ell-1} - \mu) + E_N(e_{N+\ell}) \\ &= \hat{\mu} + \hat{\phi}_1 E_N(Y_{N+\ell-1} - \mu) \end{aligned}$$

โดยที่ $E_N(e_{N+\ell}) = 0$; $\ell > 0$ จาก 2.4.2

เมื่อ $\ell = 1$: $\hat{Y}_N(1) = E_N(Y_{N+1}) = \hat{\mu} + \hat{\phi}_1 (Y_N - \hat{\mu})$;

Y_N เป็น observed value

และ $\ell \geq 2$: $\hat{Y}_N(\ell) = E_N(Y_{N+\ell}) = \hat{\mu} + \hat{\phi}_1\left[\hat{Y}_N(\ell-1) - \hat{\mu}\right]$;

$\hat{Y}_N(\ell-1)$ เป็นค่าคาดคะเนก่อนหน้า $\hat{Y}_N(\ell)$

ในทำนองเดียวกันรูปแบบอื่น ๆ จะได้ดังนี้

AR (2) :

$$\hat{Y}_N(1) = \hat{\mu} + \hat{\phi}_1\left[Y_N - \hat{\mu}\right] + \hat{\phi}_2\left[Y_{N-1} - \hat{\mu}\right]$$

$$\hat{Y}_N(2) = \hat{\mu} + \hat{\phi}_1\left[\hat{Y}_N(1) - \hat{\mu}\right] + \hat{\phi}_2\left[Y_N - \hat{\mu}\right]$$

และ $\hat{Y}_N(\ell) = \hat{\mu} + \hat{\phi}_1\left[\hat{Y}_N(\ell-1) - \hat{\mu}\right] + \hat{\phi}_2\left[\hat{Y}_N(\ell-2) - \hat{\mu}\right]$ เมื่อ $\ell \geqslant 3$

MA (1) : $\hat{Y}_N(1) = \hat{\mu}$

และ $\hat{Y}_N(\ell) = \hat{\mu} - \hat{\theta}_1\hat{e}_N$ เมื่อ $\ell \geqslant 2$

MA (2) : $\hat{Y}_N(1) = \hat{\mu} - \hat{\theta}_1\hat{e}_N - \hat{\theta}_2\hat{e}_{N-1}$

$\hat{Y}_N(2) = \hat{\mu} - \hat{\theta}_2\hat{e}_N$

และ $\hat{Y}_N(\ell) = \hat{\mu}$ เมื่อ $\ell \geqslant 3$

ARMA (1,1) : $\hat{Y}_N(1) = \hat{\mu} + \hat{\phi}_1(Y_N - \hat{\mu}) - \hat{\theta}_1\hat{e}_N$

และ $\hat{Y}_N(\ell) = \hat{\mu} + \hat{\phi}_1\left[\hat{Y}_N(\ell-1) - \hat{\mu}\right]$ เมื่อ $\ell \geqslant 2$

ARI (1,1) : $\hat{Y}_N(1) = Y_N + \hat{\phi}_1\left[Y_N - Y_{N-1}\right]$

$\hat{Y}_N(2) = \hat{Y}_N(1) + \hat{\phi}_1\left[\hat{Y}_N(1) - Y_N\right]$

และ $\hat{Y}_N(\ell) = \hat{Y}_N(\ell-1) + \hat{\phi}_1\left[\hat{Y}_N(\ell-1) - \hat{Y}_N(\ell-2)\right]$

เมื่อ $\ell \geqslant 3$

ARI (2,1) : $\hat{Y}_N(1) = Y_N + \hat{\phi}_1\left[Y_N - Y_{N-1}\right] + \hat{\phi}_2\left[Y_{N-1} - Y_{N-2}\right]$

หรือ $= (1 + \hat{\phi}_1)\, Y_N - (\hat{\phi}_1 - \hat{\phi}_2)\, Y_{N-1} - \hat{\phi}_2 Y_{N-2}$

$\hat{Y}_N(2) = \hat{Y}_N(1) + \hat{\phi}_1\left[\hat{Y}_N(1) - Y_N\right] + \hat{\phi}_2\left[Y_N - Y_{N-1}\right]$

หรือ $= (1 + \hat{\phi}_1)\hat{Y}_N(1) - (\hat{\phi}_1 - \hat{\phi}_2)Y_N - \hat{\phi}_2\, Y_{N-1}$

$$\hat{Y}_N(3) = \hat{Y}_N(2) + \hat{\phi}_1\left[\hat{Y}_N(2) - \hat{Y}_N(1)\right] + \hat{\phi}_2\left[\hat{Y}_N(1) - Y_N\right]$$

หรือ $= (1 + \hat{\phi}_1)\hat{Y}_N(2) - (\hat{\phi}_1 - \hat{\phi}_2)\hat{Y}_N(1) - \hat{\phi}_2 Y_N$

และ $\hat{Y}_N(\ell) = \hat{Y}_N(\ell-1) + \hat{\phi}_1\left[\hat{Y}_N(\ell-1) - \hat{Y}_N(\ell-2)\right] + \hat{\phi}_2\left[\hat{Y}_N(\ell-2) - \hat{Y}_N(\ell-3)\right]$ เมื่อ $\ell \geq 4$

หรือ $= (1 + \hat{\phi}_1)\hat{Y}_N(\ell-1) - (\hat{\phi}_1 - \hat{\phi}_2)\hat{Y}_N(\ell-2) - \hat{\phi}_2\hat{Y}_N(\ell-3)$

IMA (1,1) : $\hat{Y}_N(1) = Y_N - \hat{\theta}_1\hat{e}_N$

และ $\hat{Y}_N(\ell) = \hat{Y}_N(\ell-1)$ เมื่อ $\ell \geq 2$

IMA (1,2) : $\hat{Y}_N(1) = Y_N - \hat{\theta}_1\hat{e}_N - \hat{\theta}_2\hat{e}_{N-1}$

$\hat{Y}_N(2) = \hat{Y}_N(1) - \hat{\theta}_2\hat{e}_N$

และ $Y_N(\ell) = Y_N(\ell-1)$ เมื่อ $\ell \geq 3$

ARIMA (1,1,1) : $\hat{Y}_N(1) = Y_N + \hat{\phi}_1\left[Y_N - Y_{N-1}\right] - \hat{\theta}_1 e_N$

หรือ $= (1 + \hat{\phi}_1)Y_N - \hat{\phi}_1 Y_{N-1} - \hat{\theta}_1\hat{e}_N$

$\hat{Y}_N(2) = \hat{Y}_N(1) + \hat{\phi}_1\left[\hat{Y}_N(1) - Y_N\right]$

หรือ $= (1 + \hat{\phi}_1)\hat{Y}_N(1) - \hat{\phi}_1 Y_N$

และ $\hat{Y}_N(\ell) = \hat{Y}_N(\ell-1) + \hat{\phi}_1\left[\hat{Y}_N(\ell-1) - \hat{Y}_N(\ell-2)\right]$ เมื่อ $\ell \geq 3$

หรือ $= (1 + \hat{\phi}_1)\hat{Y}_N(\ell-1) - \hat{\phi}_1\hat{Y}_N(\ell-2)$

2.5 ขอบเขตที่น่าจะเป็นของค่าคาดคะเน (Probability Limits)

เป็นการหาจุดสูงสุดและต่ำสุดของค่าคาดคะเนในอนาคตแต่ละตัว ซึ่งจะทำให้เราทราบขอบเขตของค่าคาดคะเนนั้น ๆ จากค่าคาดคะเนในอนาคตที่กล่าวมาแล้ว คือ $\hat{Y}_N(\ell)$; $\ell = 1,2,\ldots,q$ และถ้าเรามีค่าจากการสังเกต $Y_{N+\ell}$ ก็จะหาค่าความคลาดเคลื่อนในอนาคตได้ (Forecast Error), $e_N(\ell)$ ซึ่งเท่ากับ $Y_{N+\ell} - \hat{Y}_N(\ell)$ แต่ทั้งนี้ค่า $Y_{N+\ell}$ เป็นค่าที่ยังไม่เกิดขึ้นจริง ฉะนั้น $e_N(\ell)$ จะเป็นตัวแปรเชิงสุ่ม (Random Variable) ในการหาค่าความแปรปรวนของ $e_N(\ell)$ จะหาจากรูปแบบ ARIMA ซึ่งโดยทั่วไปจะประกอบด้วยเทอมของความคลาดเคลื่อนในอดีต คือ e_t, e_{t-1}, ... สามารถเขียนเป็นสมการได้ดังนี้

2.5.1 $$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \ldots$$

โดยที่ $\psi_1, \psi_2, \ldots$ เป็นค่าสัมประสิทธิ์ ซึ่งเป็นฟังก์ชั่นของพารามิเตอร์ใน ARIMA ดังจะได้กล่าวต่อไป และถ้าเป็นอนุกรมเวลาคงที่ ข้างซ้ายของสมการก็จะเป็น $\dot{Y}_t = Y_t - \mu$

ให้ $t = N + \ell$ จาก 2.5.1 จะได้

2.5.2 $$Y_{N+\ell} = e_{N+\ell} + \psi_1 e_{N+\ell-1} + \psi_2 e_{N+\ell-2} + \ldots + \psi_{\ell-1} e_{N+1} + \psi_\ell e_N + \ldots$$

โดยที่ ณ ช่วงเวลาที่ N ; เทอมของ e_N; e_{N-1}, e_{N-2}, ... จะคือค่าความคลาดเคลื่อนจากการสังเกต (Observed Error) และเทอมของ $e_{N+\ell}$, $e_{N+\ell-1}$, $e_{N+\ell-2}$, ... เป็นค่าความคลาดเคลื่อนในอนาคต ซึ่งค่าคาดคะเนของเทอมเหล่านี้จะเป็นศูนย์

ค่าคาดคะเนของสมการ 2.5.2 ณ จุดเริ่มต้นของการคาดคะเนที่ N คือ

$$2.5.3 \quad \hat{Y}_N(\ell) = \underset{N}{E}(Y_{N+\ell}) = \psi_\ell e_N + \psi_1 e_{N+\ell-1} + \psi_2 e_{N+\ell-2} + \ldots\ldots$$

นำเอาสมการ 2.5.2 ลบด้วยสมการ 2.5.3จะได้ค่าความคลาดเคลื่อนในอนาคต เป็นดังนี้

$$e_N(\ell) = Y_{N+\ell} - \hat{Y}_N(\ell) = e_{N+\ell} + \psi_1 e_{N+\ell-1} + \psi_2 e_{N+\ell-2} + \ldots\ldots + \psi_{\ell-1} e_{N+1}$$

เพราะฉะนั้น $\quad \mathrm{Var}\left[e_N(\ell)\right] = (1 + \psi_1^2 + \psi_2^2 + \ldots\ldots + \psi_{\ell-1}^2)\,\sigma_e^2$

ทั้งนี้ภายใต้ข้อสมมติที่ว่าความคลาดเคลื่อนเหล่านี้ไม่มีความสัมพันธ์ต่อกัน และค่าความแปรปรวนของความคลาดเคลื่อนในอนาคต (The Variance of the Forecast Error) นี้ จะนำไปใช้สร้างขอบเขตของค่าคาดคะเนได้ โดยถือว่าภายใต้ข้อสมมติที่ว่า $Y_{N+\ell}$ มีการกระจายแบบปกติ แล้วค่าความแปรปรวนของ $\hat{Y}_N(\ell)$ จะเท่ากับความแปรปรวนของค่าความคลาดเคลื่อนในอนาคต ฉะนั้นค่า $Y_{N+\ell}$ ที่เกิดขึ้นจะอยู่ภายในขอบเขตนี้ ณ ระดับนัยสำคัญ α คือ

$$2.5.4 \quad \hat{Y}_N(\ell) \pm Z_{\alpha/2}\sqrt{1 + \psi_1^2 + \psi_2^2 + \ldots\ldots + \psi_{\ell-1}^2}\;\sigma_e$$

<u>การหาค่าสัมประสิทธิ์ ψ ของแต่ละรูปแบบ</u>

จากสมการ $\quad Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \ldots\ldots$

สามารถหาค่า $\psi_1, \psi_2, \ldots\ldots$ ได้เมื่อให้ $t = 1, 2, \ldots\ldots$ ซึ่งจะสอดคล้องกับสมการโพลีโนเมียลของ X ใน ARIMA ดังนี้

$$2.5.5 \quad (1 - \phi_1 X - \phi_2 X^2 - \ldots\ldots \phi_p X^p)(1 - X)^d (1 + \psi_1 X + \psi_2 X^2 + \ldots\ldots) = (1 - \theta_1 X - \theta_2 X^2 - \ldots\ldots - \theta_q X^q)$$

โดยที่ p และ q เป็นอันดับที่ในรูปแบบ AR และ MA ตามลำดับ และ d คือจำนวนครั้งที่ได้ทำการหาผลต่าง (Differencing) เพื่อทำให้อนุกรมเวลาไม่คงที่เป็นอนุกรมเวลาคงที่

จาก 2.5.5 นี้จะนำมาหาค่า $\psi_1, \psi_2, \ldots\ldots$ ได้โดยการแทนค่า p,q และ d ตามลักษณะของรูปแบบนั้น

AR (1) หรือ ARIMA (1,0,0)

นั่นคือ $\quad p = 1,\ d = q = 0$

จาก 2.5.5 จะได้ว่า

$$(1 - \phi_1 X)(1 - X)^0 (1 + \psi_1 X + \psi_2 X^2 + \ldots\ldots) = 1$$

$$(1 - \phi_1 X)(1 + \psi_1 X + \psi_2 X^2 + \ldots\ldots) = 1$$

$$1 + (\psi_1 - \phi_1)X + (\psi_2 - \psi_1\phi_1)X^2 + (\psi_3 - \psi_2\phi_1)X^3 + \ldots = 1$$

เปรียบเทียบสมการข้างซ้ายและขวาจะได้

$\psi_1 - \phi_1 = 0$ นั่นคือ $\psi_1 = \phi_1$

$\psi_2 - \psi_1\phi_1 = 0$ นั่นคือ $\psi_2 = \psi_1\phi_1 = \phi_1^2$

$\psi_3 - \psi_2\phi_1 = 0$ นั่นคือ $\psi_3 = \psi_2\phi_1 = \phi_1^3$

ฉะนั้นสามารถเขียนในรูปทั่วไปได้ดังนี้ $\psi_j = \phi_1^j$ เมื่อ $j \geq 1$

ในทำนองเดียวกัน

AR (2) หรือ ARIMA (2,0,0) คือ $p = 2$, $d = q = 0$ จากสมการ 2.5.5ผลจะได้ดังนี้

$$\psi_1 = \phi_1$$

$$\psi_2 = \phi_1^2 + \phi_2$$

$$\psi_j = \psi_{j-1}\phi_1 + \psi_{j-2}\phi_2 \qquad \text{เมื่อ } j \geqslant 3$$

MA (1) หรือ ARIMA (0,0,1) คือ $p = d = 0$; $q = 1$ จากสมการ 2.5.5 ผลจะได้ดังนี้

$$\psi_1 = -\theta_1$$

$$\psi_j = 0 \qquad \text{เมื่อ } j \geqslant 2$$

MA (2) หรือ ARIMA (0,0,2) คือ $p = d = 0$, $q = 2$ จากสมการ 2.5.5 ผลจะได้ดังนี้

$$\psi_1 = -\theta_1$$

$$\psi_2 = -\theta_2$$

$$\psi_j = 0 \qquad \text{เมื่อ } j \geqslant 3$$

ARMA (1,1) หรือ ARIMA (1,0,1) คือ $p = 1$, $d = 0$, $q = 1$ จากสมการ 2.5.5

$$\psi_1 = \phi_1 - \theta_1 \qquad \text{เมื่อ } j = 1$$

$$\psi_j = \phi_1 \psi_{j-1} \qquad \text{เมื่อ } j \geqslant 2$$

ARI (1,1) หรือ ARIMA (1,1,0) คือ $p = 1, d = 1, q = 0$ จากสมการ 2.5.5 ผลจะได้ดังนี้

$$\psi_j = 1 + \phi_1 + \phi_1^2 + \ldots\ldots + \phi_1^j \qquad \text{เมื่อ} \quad j \geq 1$$

ARI (2,1) หรือ ARIMA (2,1,0) คือ $p = 2, d = 1, q = 0$ จากสมการ 2.5.5 ผลจะได้ดังนี้

$$\psi_1 = 1 + \phi_1$$
$$\psi_2 = (1 + \phi_1)^2 - (\phi_1 - \phi_2)$$
$$\psi_3 = (1 + \phi_1)\psi_2 - (\phi_1 - \phi_2)\psi_1 - \phi_2$$
$$\psi_j = (1 + \phi_1)\psi_{j-1} - (\phi_1 - \phi_2)\psi_{j-2} - \phi_2\psi_{j-3} \qquad \text{เมื่อ} \quad j \geq 4$$

IMA (1,1) หรือ ARIMA (0,1,1) คือ $p = 0, d = 1, q = 1$ จากสมการ 2.5.5 ผลจะได้ดังนี้

$$\psi_j = (1 - \theta_1) \qquad \text{เมื่อ} \quad j \geq 1$$

IMA (1,2) หรือ ARIMA (0,1,2) คือ $p = 0, d = 1, q = 2$ จากสมการ 2.5.5 ผลจะได้ดังนี้

$$\psi_1 = 1 - \theta_1$$
$$\psi_j = 1 - \theta_1 - \theta_2 \qquad \text{เมื่อ} \quad j \geq 2$$

ARIMA (1,1,1) คือ $p = d = q = 1$ จากสมการ 2.5.5 ผลจะได้ดังนี้

$$\psi_1 = 1 + \phi_1 - \theta_1$$
$$\psi_2 = (1 + \phi_1)^2 - \theta_1(1 + \phi_1) - \phi_1$$
$$\psi_j = (1 + \phi_1)\psi_{j-1} - \phi_1\psi_{j-2} \quad \text{เมื่อ } j \geq 3$$

ทั้งนี้โดยที่ ψ_0 ของทุก ๆ รูปแบบมีค่าเท่ากับ 1 และจะเห็นได้ว่าค่าสัมประสิทธิ์ ψ_j นี้ จะเป็นฟังก์ชั่นของพารามิเตอร์ $\underset{\sim}{\phi}$ และ/หรือ $\underset{\sim}{\theta}$ ซึ่งจาก 2.2.2 จะได้ค่าประมาณกำลังสองน้อยที่สุดของพารามิเตอร์เหล่านี้ $\hat{\underset{\sim}{\phi}}, \hat{\underset{\sim}{\theta}}$ นำมาหาค่า $\hat{\psi}_j$ ซึ่งจะได้ใช้เป็นค่าประมาณของ ψ_j และใช้ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน $\hat{\sigma}_e$ เป็นค่าประมาณของ σ_e เพื่อหาค่าขอบเขตของ $\hat{Y}_N(\ell)$ ดังสมการ 2.5.4

ข้อสังเกต : สำหรับอนุกรมเวลาคงที่ ค่า ψ_j^2 จะตัดออกไปเมื่อ j มีค่ามาก ๆ และขอบเขตที่น่าจะเป็นนี้ จะมีลักษณะอยู่ในแนวนอนตลอด (Horizontal Bands)

สำหรับอนุกรมเวลาไม่คงที่ ค่า ψ_j^2 จะต้องนำมาคิดด้วยถึงแม้ j มีค่ามากขึ้น และขอบเขตที่น่าจะเป็น จะมีลักษณะขยายกว้างขึ้นเมื่อช่วงเวลาของการคาดคะเนเพิ่มขึ้น ซึ่งทำให้ความเชื่อในค่าคาดคะเนนั้น ๆ มีน้อยลง

2.6 การปรับปรุงค่าคาดคะเน (Updating Forecasts)

เป็นการปรับปรุงค่าคาดคะเนที่คำนวณได้แล้วให้ดียิ่งขึ้น ถ้าหากว่าได้ข้อมูลอนุกรมเวลาชุดนั้นเพิ่มขึ้นคือ $Y_{N+1}, Y_{N+2}, \ldots\ldots$ มีวิธีการคำนวณ 2 วิธีด้วยกัน คือ

วิธีการปรับปรุงค่าคาดคะเนโดยทั่ว ๆ ไป สำหรับทุกรูปแบบ ARIMA มีสูตรในการคำนวณดังนี้

$$\hat{Y}_{N+1}(\ell) = \hat{Y}_N(\ell + 1) + \psi_\ell e_{N+1} \quad ; \quad \ell = 1,2,\ldots\ldots$$

โดยที่ $\hat{Y}_{N+1}(\ell)$ คือ ค่าประมาณของค่าคาดคะเนที่เวลา $N + \ell + 1$ จากจุดเริ่มต้นของการคาดคะเนที่ $N + 1$

$\hat{Y}_N(\ell + 1)$ คือ ค่าปรับปรุงของค่าคาดคะเนที่เวลา $N + \ell + 1$ จากจุดเริ่มต้นของการคาดคะเนที่ N

e_{N+1} คือ ค่าความคลาดเคลื่อนระหว่าง $Y_{N+\ell} - \hat{Y}_N(\ell)$ หรือเรียกว่า One-Step ahead forecast error ณ ช่วงเวลาที่ N

ψ_ℓ คือ ค่าสัมประสิทธิ์ซึ่งเป็นฟังก์ชั่นของพารามิเตอร์ ดังสมการในข้อ 2.5

วิธีการนี้เป็นการปรับปรุงค่าคาดคะเนที่คำนวณได้เติมให้ถูกต้องยิ่งขึ้น โดยอาศัยค่าความคลาดเคลื่อนระหว่างข้อมูลจริงที่ได้เพิ่มขึ้น กับค่าคาดคะเนตัวแรก คือ e_{N+1} คูณกับ ψ_ℓ จากรูปแบบที่เลือกใช้ซึ่งถ้าหากว่าค่า e_{N+1} เป็นค่าลบ ค่าคาดคะเนหลังการปรับปรุงก็จะลดลง แต่ถ้าหากว่าเป็นค่าบวก ค่าคาดคะเนหลังการปรับปรุงก็จะเพิ่มขึ้น และค่าคาดคะเนหลังการปรับปรุงเกือบเท่ากันตลอดทุกช่วงเวลา ฉะนั้นเมื่อได้ข้อมูลจริงเพิ่มขึ้นแต่ละครั้ง ก็ต้องนำมาปรับปรุงค่าคาดคะเนตามสูตรข้างต้น โดยเปลี่ยนจุดเริ่มต้นของการคาดคะเน

อีกวิธีหนึ่งทำได้โดยตรงจากรูปแบบที่เลือกใช้กับข้อมูลชุดนั้น ๆ และใช้หาค่าคาดคะเน ณ จุดเริ่มต้นของการคาดคะเนที่ N เมื่อได้ค่าข้อมูลจริงที่เกิดขึ้นในช่วงเวลาที่ N+1 แล้วก็นำค่าตัวเลขใหม่นี้มาปรับค่าคาดคะเนที่ได้คำนวณไว้ จะได้ค่าคาดคะเนชุดใหม่ ณ จุดเริ่มต้นของการคาดคะเนที่ N+1 สัญลักษณ์คือ $\hat{Y}_{N+1}(\ell)$ และในทำนองเดียวกันเมื่อได้ค่าข้อมูลจริงเพิ่มขึ้นแต่ละครั้ง ก็ควรนำมาใช้ปรับปรุงค่าคาดคะเนใหม่ที่จุดเริ่มต้นของการคาดคะเนนั้น ๆ

นอกจากนี้ในระยะต่อมา เมื่อข้อมูลจริงเกิดขึ้นหลาย ๆ ค่าหลังจากช่วงเวลาที่ N แล้ว ควรจะได้นำมาวิเคราะห์ว่ารูปแบบที่นำมาใช้คาดคะเนค่าในอนาคตนั้น ยังมีความเหมาะสมอยู่อีกหรือไม่ หรือไม่เช่นนั้นก็ควรจะได้หาค่าประมาณจากพารามิเตอร์ใหม่จากรูปแบบนั้น ๆ ซึ่งอาจมีการเปลี่ยนแปลงจากเดิมก็ได้ ทั้งนี้เพื่อให้ได้รูปแบบที่มีความถูกต้องสมบูรณ์ยิ่งขึ้น

2.7 ในส่วนต่อไปนี้จะเป็นวิธีการและขั้นตอนในการวิเคราะห์ข้อมูลอนุกรมเวลาโดยทั่วไปหรือที่เรียกว่า Classical Time Series Analysis ซึ่งส่วนประกอบในอนุกรมเวลานั้นแบ่งออกได้เป็น 4 ชนิดด้วยกันคือ

ก. แนวโน้มตามลำดับเวลา (Secular Trend ; T) หมายถึง ลักษณะการเคลื่อนไหวขึ้นหรือลงของเส้นที่ยาวต่อเนื่องกันไปในช่วงระยะเวลายาว โดยไม่มีการหักมุม ณ ที่ใดของเส้นนั้น

ข. การเปลี่ยนแปลงตามฤดูกาล (Seasonal Movements ; S) หมายถึง พฤติการณ์ที่เกิดขึ้นอย่างเดียวกันหรือคล้าย ๆ กัน ในช่วงระยะเวลาอันสั้น เช่น เป็นรายวัน รายสัปดาห์ หรือรายเดือนเป็นต้น ในช่วงเวลาที่เหมือนกันข้อมูลมีการเปลี่ยนแปลงคล้ายคลึงกัน

ค. การเปลี่ยนแปลงตามวัฏจักร (Cyclical Movements ; C) หมายถึง พฤติการณ์ที่เกิดขึ้นเป็นระยะเวลาหลาย ๆ ปีคล้าย ๆ กับแนวโน้ม แต่มีรูปร่างประกอบด้วยระยะเวลารุ่งเรืองถึงจุดสูงสุดจนถึงต่ำสุด

ง. การเปลี่ยนแปลงเนื่องจากเหตุการณ์ผิดปกติ (Irregular Movements ; I) เช่น การเคลื่อนไหวที่เกิดขึ้นจากภัยธรรมชาติเป็นต้นว่า น้ำท่วม แผ่นดินไหว

ส่วนประกอบของอนุกรมเวลามี 2 ลักษณะ คือ ในเชิงบวก (Additive Models) และในเชิงคูณ (Multiplicative Models) ซึ่งส่วนใหญ่ข้อมูลทางเศรษฐกิจจะอยู่ในลักษณะผลคูณ ฉะนั้นในที่นี้จะใช้องค์ประกอบในเชิงผลคูณมีรูปสมการดังนี้

$$Y_i = T_i \times S_i \times C_i \times I_i$$

เมื่อ Y_i คือ ตัวแปรตาม (Dependent variable) ที่เราต้องการศึกษาในช่วงระยะเวลา i ซึ่งเราคาดว่าจะขึ้นอยู่กับแนวโน้ม (T) การเปลี่ยนแปลงเนื่องจากฤดูกาล (S) การเปลี่ยนแปลงตามวัฏจักร (C) และการเปลี่ยนแปลง เนื่องจากเหตุการณ์ผิดปกติ (I)

ในรูปแบบเชิงผลคูณนี้มักจะพบค่าแนวโน้ม และการเปลี่ยนแปลงเนื่องจากฤดูกาลปรากฏอยู่เสมอ ดังนั้นจะนำวิธีการวิเคราะห์หาค่าแนวโน้ม ซึ่งเป็นการศึกษาหาความสัมพันธ์ระหว่างเวลากับข้อมูลที่สนใจศึกษา โดยมีเวลา (X) เป็นตัวแปรอิสระ (Independent Variable) และข้อมูลที่สนใจศึกษา (Y) เป็นตัวแปรตาม (Dependent Variable) และหาค่าดัชนีฤดูกาล (Seasonal Index) เพื่อนำทั้งสองส่วนประกอบนี้มาใช้หาค่าคาดคะเนต่อไป

2.7.1 การวิเคราะห์หาแนวโน้มของข้อมูล ข้อมูลที่จะนำมาศึกษาเพื่อหาแนวโน้มนั้นจะต้องเป็นอนุกรมเวลาในระยะเวลายาวพอสมควร ส่วนมากตั้งแต่ 10 ช่วงเวลาขึ้นไป อาจจะเป็นเดือนหรือปีก็ได้ การสร้างแนวโน้มนี้อาจจะเป็นรูปอะไรก็ได้ ซึ่งแบ่งได้อย่างกว้าง ๆ เป็น 2 ลักษณะ คือ แนวโน้มที่เป็นเส้นตรง (Linear Trend) และไม่เป็นเส้นตรง (Non - Linear Trend) ในการคำนวณหาเส้นแนวโน้มมีหลายวิธีด้วยกัน แต่วิธีที่นิยมใช้กันมากที่สุดและง่ายในการคำนวณ คือ วิธีกำลังสองน้อยที่สุด (Least Squares Method) โดยมีหลักเกณฑ์ที่พยายามทำให้ผลรวมของผลต่างระหว่างค่าที่ได้จากเส้นแนวโน้มกับค่าข้อมูลจริงยกกำลังสองแล้วได้ค่าน้อยที่สุด

รูปแบบทั่วไปของสมการแนวโน้ม คือ สมการโพลิโนเมียล มีรูปแบบดังนี้

$$Y = a_0 + a_1X + a_2X^2 + a_3X^3 + \ldots\ldots + a_nX^n$$

โดยที่ Y เป็น ตัวแปรตาม

X เป็น ตัวแปรอิสระ

a_i เป็น ค่าคงที่หรือพารามิเตอร์ ; i = 0,1,2, ... n

เส้นกราฟที่ได้จากสมการโพลิโนเมียลแต่ละอันดับจะมีลักษณะแตกต่างกันไป คือ

อันดับที่ 1 จะเป็นรูปแบบของแนวโน้มเส้นตรง (Linear Trend)

$$Y_t = a_0 + a_1X$$

อันดับที่ 2 จะเป็นรูปแบบของแนวโน้มเส้นโค้งเดียว (Parabola Trend)

$$Y_t = b_0 + b_1X + b_2X^2$$

อันดับที่ 3 จะเป็นรูปแบบของแนวโน้มสองโค้ง (Hyperbola Trend)

$$Y_t = C_0 + C_1X + C_2X^2 + C_3X^3$$

การคำนวณหาแนวโน้มโดยสมการโพลิโนเมียล มักนิยมและเหมาะสมกับข้อมูลโดยทั่วไปเพียงอันดับสามเท่านั้นเพราะยิ่งกำลังสูงมากขึ้น การคำนวณก็จะยุ่งยากมากขึ้นตามไปด้วย และจากทั้งสามรูปแบบนี้จะหาค่าประมาณของพารามิเตอร์ด้วยวิธีการของ Least Squares และคำนวณจากสมการปกติ (Normal Equations) ซึ่งหามาได้จากการ Differentiate ผลรวมของผลต่างกำลังสองของข้อมูลทั้งหมด เทียบกับตัวคงที่ในรูปแบบนั้น ๆ

นอกจากสามรูปแบบที่กล่าวมาแล้ว จะได้นำรูปแบบของแนวโน้มที่ไม่เป็นเส้นตรงอีกลักษณะหนึ่งมาคำนวณ คือ รูปแบบแนวโน้มเอ็กซ์โพเนนเชียล ; $Y_t = ab^X$ ซึ่งการหาค่าประมาณของ a,b อาจทำให้ง่ายขึ้น โดยการแปรรูปเข้าช่วยเพื่อให้สมการเดิมเปลี่ยนรูปเป็นเส้นตรง จะได้ดังนี้

$$\log Y_t = \log a + X \log b$$

จากสมการปกติสามารถคำนวณค่าพารามิเตอร์ได้จากสูตร

$$\log a = \frac{\sum \log Y}{N}$$

$$\log b = \frac{\sum X \log Y}{X^2}$$

ในการวิจัยนี้ จะใช้วิธีการสำหรับการพิจารณาเลือกรูปแบบแนวโน้ม ลักษณะใดที่จะมีความเหมาะสมกับข้อมูลที่นำมาศึกษาแต่ละชุดดีที่สุดนี้จะเลือกรูปแบบที่ให้ค่าผลบวกของส่วนเบี่ยงเบนยกกำลังสองน้อยที่สุด และจากรูปแบบที่ได้นี้จะนำมาหาค่าแนวโน้มของแต่ละเดือนในปีต่าง ๆ ต่อไป

2.7.2 การวิเคราะห์หาดัชนีฤดูกาล ดัชนีฤดูกาลนี้จะใช้เป็นเครื่องมือกำจัด Seasonal Variationsออกจากข้อมูลอนุกรมเวลา และใช้ดัชนีฤดูกาลเพื่อประโยชน์ในการหาค่าคาดคะเนประกอบกับค่าแนวโน้มอีกด้วย การหาค่าดัชนีฤดูกาลนี้อาจทำได้หลายวิธีด้วยกันคือ

ก. Ratio-to-Moving Average Method

ข. Ratio-to-Trend Method ซึ่งจะได้กล่าวถึงขั้นตอนในการคำนวณในวิธีการนี้แต่เพียงอย่างเดียวดังนี้

1. จากการวิเคราะห์หาแนวโน้มของข้อมูลข้างต้น นำค่าแนวโน้ม (Y_t) ของแต่ละเดือนในปีต่าง ๆ กำจัดค่าแนวโน้มออกจากรายการโดยนำไปหารข้อมูลจริง จะได้ค่า Ratio-to-Trend ; $\left[\frac{Y_t}{\hat{Y}_t} \times 100 \right]$ ของแต่ละเดือนของทุกปี

2. กำจัด C และ I จากข้อมูลในข้อ 1 โดยวิธีหาค่าเฉลี่ย ซึ่งจะใช้ค่าเฉลี่ยเลขคณิตหรือมัธยฐานก็ได้

3. นำค่าเฉลี่ยที่ได้ในแต่ละเดือนเทียบกับค่าเฉลี่ยของค่าเฉลี่ยของทุก ๆ เดือนแล้วคูณด้วย 100 ซึ่งผลที่ได้ก็คือดัชนีฤดูกาลที่ต้องการ

เมื่อได้ค่าแนวโน้มและดัชนีฤดูกาลของอนุกรมเวลาแต่ละชุดแล้ว นำมาคำนวณหาค่าคาดหมาย (Fitted Values) เพื่อเปรียบเทียบกับค่าข้อมูลจริง พร้อมทั้งหาค่าผลบวกของส่วนเบี่ยงเบนยกกำลังสอง หลังจากนั้นก็คำนวณหาค่าคาดคะเน เพื่อช่วยในการพยากรณ์หรือทำนายเหตุการณ์ล่วงหน้าต่อไป