

บทที่ 2

ทฤษฎีที่ใช้ในงานวิจัย

การศึกษาจำนวนผู้โดยสารเครื่องบิน รถประจำทาง และรถไฟ ระหว่างเชียงใหม่และกรุงเทพมหานคร มีวัตถุประสงค์เพื่อสร้างแบบจำลองการเดินทางของผู้โดยสารในรูปของสมการถดถอย ศึกษาถึงปัจจัยที่มีอิทธิพลต่อปริมาณการเดินทางและการตัดสินใจเลือกรูปแบบในการเดินทางของผู้โดยสาร ระดับการแข่งขันกันระหว่างระบบขนส่งสาธารณะ นอกจากนี้ยังสร้างแบบจำลองพยากรณ์ปริมาณการเดินทางของผู้โดยสารทั้ง 3 ประเภทในระบบขนส่งสาธารณะโดยวิธีการวิเคราะห์อนุกรมเวลาของ Box-Jenkins

2.1 การวิเคราะห์การถดถอย (Regression Analysis)

การวิเคราะห์การถดถอยเป็นการอธิบายความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม การวิเคราะห์จะกำหนดให้ฟังก์ชันของค่าเฉลี่ยเป็นเส้นตรงหรือไม่ก็ได้ ถ้าการวิเคราะห์กำหนดให้ฟังก์ชันของค่าเฉลี่ยเป็นเส้นตรง เรียกว่า การถดถอยเชิงเส้น แต่ถ้ากำหนดให้ฟังก์ชันของค่าเฉลี่ยไม่เป็นเส้นตรง เรียกว่า การถดถอยที่ไม่เป็นเชิงเส้น หากความสัมพันธ์ของตัวแปรตามขึ้นอยู่กับตัวแปรอิสระเพียงตัวเดียวและฟังก์ชันของค่าเฉลี่ยเป็นเส้นตรงก็เรียกว่า การถดถอยเชิงเส้นแบบง่าย แต่ถ้าหากขึ้นอยู่กับตัวแปรหลายตัวก็เรียกว่าการถดถอยเชิงซ้อน การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติที่แสดงความสัมพันธ์ของตัวแปรตั้งแต่สองตัวขึ้นไปออกมาในรูปแบบของสมการทางคณิตศาสตร์ ความสัมพันธ์เชิงเส้นของสองตัวแปรสามารถแทนได้ด้วยสมการ

$$y = \beta_0 + \beta_1 x + e \quad (2.1)$$

เมื่อ

y คือ ตัวแปรตาม (Dependent Variable)

x คือ ตัวแปรอิสระ (Independent Variable)

β_0, β_1 คือ ค่าพารามิเตอร์

e คือ ค่าคลาดเคลื่อน

ส่วนมากการวิเคราะห์การถดถอยที่นำมาใช้ในการวิเคราะห์ปริมาณการเดินทางในระบบขนส่งนั้นจะเป็นสมการถดถอยเชิงซ้อน ซึ่งสามารถแทนได้ด้วยสมการดังนี้

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (2.2)$$

จะประมาณค่า y ได้ดังนี้คือ

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + \dots + b_k x_k \quad (2.3)$$

โดยที่

$$\hat{\beta}_0 = a, \hat{\beta}_1 = b_1, \hat{\beta}_2 = b_2, \dots, \hat{\beta}_k = b_k$$

ค่าคลาดเคลื่อนในการประมาณค่า y ด้วย $\hat{y}$ คือ $y - \hat{y} = e$ หรือเรียกว่า residual การประมาณค่าพารามิเตอร์ $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ ด้วยค่า $a, b_1, b_2, \dots, b_k$ ทำได้โดยวิธีกำลังสองน้อยที่สุด (Least Square Method) นั่นคือการประมาณค่า $a, b_1, b_2, \dots, b_k$ ที่ทำให้ผลบวกกำลังสองของค่าคลาดเคลื่อนมีค่าน้อยที่สุด หรือ

$$\sum_{i=1}^k e_i^2 = \sum_{i=1}^k (y_i - \hat{y}_i)^2 \quad (2.4)$$

2.1.1 สมมติฐานในการวิเคราะห์ความถดถอยเชิงซ้อน

1. ค่าคลาดเคลื่อน e เป็นตัวแปรที่มีการแจกแจงแบบปกติ
2. ค่าคลาดเคลื่อน e เป็นตัวแปรที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนเป็นค่าคงที่ที่ไม่ทราบค่า $Var(e) = \sigma_e^2$
3. e_i และ e_j เป็นอิสระต่อกัน นั่นคือ $Cov(e_i, e_j) = 0$; $i \neq j$
4. ตัวแปรอิสระ x_i และ x_j ต้องเป็นอิสระกัน

2.1.2 การทดสอบสมมติฐาน

ก. ทดสอบสมมติฐานเกี่ยวกับความเหมาะสมของสมการความถดถอยเชิงซ้อน

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \text{มี } \beta_i \text{ อย่างน้อย 1 ค่าที่ } \neq 0 ; i = 1, 2, \dots, k$$

$$\text{สถิติทดสอบ } F = \frac{MSR}{MSE}$$

เมื่อ MSR (Mean Square of Regression) คือ ค่าความแปรปรวนของ y เนื่องจากอิทธิพลของตัวแปรอิสระ และ MSE (Mean Square of Error) คือ ค่าความแปรปรวนของ y เนื่องจากอิทธิพลอื่นๆ หรือเรียกว่าค่าแปรปรวนอย่างสุ่ม

โดยจะปฏิเสธสมมติฐาน H_0 เมื่อ $F > F_{k, n-k-1, 1-\alpha}$ และจะยอมรับสมมติฐาน H_0 เมื่อ $F \leq F_{k, n-k-1, 1-\alpha}$

ข. ทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์ความถดถอย

หลังจากที่ทดสอบสมมติฐานโดยวิเคราะห์ความแปรปรวน $F-test$ พบว่ามี β_i อย่างน้อย 1 ค่าที่ไม่เท่ากับศูนย์ จะต้องทำการทดสอบต่อไปว่ามี β_i ตัวใดบ้างที่ไม่เท่ากับศูนย์ หรือมี x ตัวใดบ้างที่สัมพันธ์กับ y โดยการทดสอบสมมติฐานต่อไปนี้

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \quad ; i = 1, 2, \dots, k$$

$$\text{สถิติทดสอบ} \quad t = \frac{b_i - 0}{S_{b_i}}$$

เมื่อ S_{b_i} คือ ค่าเบี่ยงเบนมาตรฐานของ b_i

โดยจะปฏิเสธสมมติฐาน H_0 เมื่อ $|t| > t_{1-\alpha/2, n-k-1}$ และจะยอมรับสมมติฐาน H_0 เมื่อ $|t| \leq t_{1-\alpha/2, n-k-1}$

การวิเคราะห์การถดถอยแตกต่างกับการวิเคราะห์อนุกรมเวลา เนื่องจากการวิเคราะห์การถดถอยเป็นการพิจารณาค่าของข้อมูลในช่วงเวลาใดช่วงเวลาหนึ่งเท่านั้น (Cross sectional data) การพยากรณ์โดยวิธีการถดถอยจะใช้ตัวแปรอิสระเป็นปัจจัยหรือต้นเหตุของการเปลี่ยนแปลงค่าสังเกต ส่วนการพยากรณ์โดยวิธีการวิเคราะห์อนุกรมเวลาจะคำนึงว่าการเปลี่ยนแปลงของค่าสังเกตหรือข้อมูลเป็นผลเนื่องมาจากการเปลี่ยนแปลงเวลาเพียงอย่างเดียวเท่านั้น

2.2 การวิเคราะห์อนุกรมเวลาของ Box-Jenkins

การพยากรณ์อนุกรมเวลาของ Box-Jenkins เป็นวิธีการพยากรณ์ค่าในอนาคตที่พัฒนาโดยนักสถิติผู้มีชื่อเสียงสองท่านคือ George E. P. และ Gwilym M. Jenkins การพยากรณ์จะใช้อนุกรมเวลาในอดีตในการพยากรณ์อนุกรมเวลาในอนาคต การพยากรณ์โดยวิธีนี้จะมีความละเอียดและถูกต้อง

(Accuracy) สูงกว่าวิธีอื่นในการพยากรณ์ระยะสั้นและระยะกลาง แต่เป็นวิธีที่ค่อนข้างจะยุ่งยากในการวิเคราะห์เพราะจะต้องใช้ความรู้เกี่ยวกับรูปแบบ ARMA (Autoregressive Moving Average) และใช้เวลาในการวิเคราะห์คำนวณมาก

2.2.1 ลักษณะที่สำคัญของการพยากรณ์ของ Box-Jenkins

1. การพยากรณ์ระยะสั้น การพยากรณ์ของ Box-Jenkins เหมาะสำหรับการพยากรณ์ระยะสั้น เพราะรูปแบบของการพยากรณ์จะให้ความสำคัญกับอนุกรมเวลาที่อยู่ใกล้เวลาพยากรณ์มากกว่าอนุกรมเวลาที่อยู่ไกลเวลาพยากรณ์ ดังนั้นการพยากรณ์ระยะยาวจะให้ความเชื่อถือได้น้อยกว่าการพยากรณ์ระยะสั้น
2. ชนิดของอนุกรมเวลา อนุกรมเวลาที่ใช้ในการพยากรณ์ของ Box-Jenkins จะต้องเป็นจำนวนจริง และอนุกรมเวลาจะต้องเกิดขึ้นในช่วงเวลาที่เท่ากัน
3. ขนาดอนุกรมเวลา Box-Jenkins ได้เสนอว่าควรใช้อนุกรมเวลาอย่างน้อย 50 ตัว แต่บางครั้งอาจใช้น้อยกว่านี้ก็ได้ สำหรับอนุกรมเวลาที่มีฤดูกาลควรจะใช้อนุกรมเวลาจำนวนมากๆ
4. อนุกรมเวลาที่เป็นสเตชันนารี อนุกรมเวลาที่ใช้ในการพยากรณ์ของ Box-Jenkins จะต้องเป็นอนุกรมเวลาที่เป็นสเตชันนารี ที่มีค่าเฉลี่ยและความแปรปรวนของข้อมูลมีค่าคงที่

2.2.2 ฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Functions)

ฟังก์ชันสหสัมพันธ์ในตัวเอง ρ_k เป็นฟังก์ชันที่ใช้วัดความสัมพันธ์ในตัวเองของอนุกรมเวลา จากการกำหนดว่าอนุกรมเวลาจะต้องมีคุณสมบัติเป็นสเตชันนารี จะได้ว่าความแปรปรวนร่วมในตัวเอง (Autocovariance) ของอนุกรมเวลาที่มีช่วงห่างเวลาเท่ากันจะไม่มีเปลี่ยนแปลง ซึ่งความแปรปรวนร่วมในตัวเองของ Z_t และ Z_{t+k} ที่ช่วงห่าง k สามารถนิยามได้ดังสมการ (2.5)

$$\gamma_k = \text{Cov}(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu) \quad (2.5)$$

เมื่อ

Z_t คือ อนุกรมเวลาที่เวลา t

Z_{t+k} คือ อนุกรมเวลาที่เวลา $t+k$

γ_k คือ ความแปรปรวนร่วมในตัวเอง (Autocovariance)

μ คือ ค่าเฉลี่ยของอนุกรมเวลา

สำหรับความแปรปรวนร่วมในตัวเองใช้งานยากเพราะค่าของความแปรปรวนร่วมขึ้นอยู่กับค่าของข้อมูล จึงควรเปลี่ยนให้เป็นสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (autocorrelation coefficient) ใช้สัญลักษณ์ ρ_k โดยที่

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{Cov}(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)}\sqrt{\text{Var}(Z_{t+k})}} ; k = 0, 1, 2, \dots \quad (2.6)$$

โดยจะประมาณค่า ρ_k และ γ_k ด้วย $\hat{\rho}_k$ และ $\hat{\gamma}_k$ ตามลำดับ ดังนี้

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{i=1}^{n-k} (Z_i - \bar{Z})(Z_{i+k} - \bar{Z})}{\sum_{i=1}^n (Z_i - \bar{Z})^2} ; k = 0, 1, 2, \dots \quad (2.7)$$

$$\hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^{n-k} (Z_i - \bar{Z})(Z_{i+k} - \bar{Z}) \quad (2.8)$$

โดยที่
$$\hat{\gamma}_0 = \frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n}$$

การประมาณค่าความคลาดเคลื่อนมาตรฐานของฟังก์ชันสหสัมพันธ์ในตัวเอง (Standard Error of Autocorrelation Functions) Bartlett(1946) ได้ประมาณค่าความคลาดเคลื่อนมาตรฐานของ $\hat{\rho}_k ; k = 0, 1, 2, \dots$ ดังนี้

$$SE(\hat{\rho}_k) = \sqrt{\frac{1 + 2 \sum_{j=1}^{k-1} \hat{\rho}_j^2}{n}} \quad (2.9)$$

$\hat{\rho}_k$ มีการแจกแจงแบบที (t-distribution) เพราะฉะนั้นการทดสอบสมมติฐานของ $\hat{\rho}_k$ ที่ระดับนัยสำคัญ 0.05 ทำได้โดยการทดสอบสมมติฐาน

$$H_0 : \hat{\rho}_k = 0$$

$$H_1 : \hat{\rho}_k \neq 0$$

$$\text{สถิติทดสอบ } t = \frac{\hat{\rho}_k - \rho_k}{SE(\hat{\rho}_k)}$$

ดังนั้น $\hat{\rho}_k$ จะมีนัยสำคัญทางสถิติ คือ $\hat{\rho}_k \neq 0$ เมื่อ $|\hat{\rho}_k| > 1.96 SE(\hat{\rho}_k)$ และ $\hat{\rho}_k = 0$ เมื่อ $|\hat{\rho}_k| \leq 1.96 SE(\hat{\rho}_k)$

2.2.3 ฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Functions)

ฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วนใช้วัดความสัมพันธ์ระหว่าง Z_t กับ Z_{t+k} โดยรวมที่อยู่ระหว่าง t กับ $t+k$ เช่นต้องการวัดความสัมพันธ์ระหว่าง Z_t กับ Z_{t+2} โดยรวมความสัมพันธ์ของ Z_{t+1} และ Z_{t+2} ด้วย

สหสัมพันธ์ในตัวเองบางส่วนใช้สัญลักษณ์ ϕ_{kk} การหาค่าสหสัมพันธ์ในตัวเองบางส่วนสามารถหาได้จาก สมการยูล-วอล์กเกอร์ (Yule-Walker Equation) ซึ่งสามารถเขียนอยู่ในรูปเมตริกซ์ ได้ดังนี้

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.10)$$

สามารถแก้สมการหาค่าสหสัมพันธ์ในตัวเองบางส่วน ϕ_{kk} ; $k=0,1,2,\dots$ ได้โดยการประมาณค่า ρ_k ด้วย $\hat{\rho}_k$ และ $\hat{\phi}_{kk}$ เป็นค่าประมาณของ ϕ_{kk} จะได้ว่า

$$\begin{aligned} \hat{\phi}_{11} &= \hat{\rho}_1 \\ \hat{\phi}_{22} &= \frac{\begin{bmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & \hat{\rho}_2 \end{bmatrix}}{\begin{bmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 \end{bmatrix}} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} \end{aligned}$$

และสามารถประมาณค่า $\hat{\phi}_{kk}$ ได้จากสมการของ Recursive Formular ดังนี้

$$\hat{\phi}_{k+1,k+1} = \frac{\rho_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \quad (2.11)$$

และ

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j} \quad , j=1,2,3,\dots,k \quad (2.12)$$

และประมาณค่าความคลาดเคลื่อนมาตรฐานของ $\hat{\phi}_{kk}$ ได้ดังนี้

$$SE(\hat{\phi}_{kk}) = \frac{1}{\sqrt{n}} \quad (2.13)$$

$\hat{\phi}_{kk}$ มีการแจกแจงแบบที (t-distribution) เพราะฉะนั้นการทดสอบสมมติฐานของ $\hat{\phi}_{kk}$ ที่ระดับนัยสำคัญ 0.05 ทำได้โดยการทดสอบสมมติฐาน

$$H_0 : \hat{\phi}_{kk} = 0$$

$$H_1 : \hat{\phi}_{kk} \neq 0$$

$$\text{สถิติทดสอบ } t = \frac{\hat{\rho}_k - \rho_k}{SE(\hat{\rho}_k)}$$

ดังนั้น $\hat{\phi}_{kk}$ จะมีนัยสำคัญทางสถิติ นั่นคือ $\hat{\phi}_{kk} \neq 0$ เมื่อ $|\hat{\phi}_{kk}| > 1.96SE(\hat{\phi}_{kk})$ และ $\hat{\phi}_{kk} = 0$ ก็ต่อเมื่อ $|\hat{\phi}_{kk}| \leq 1.96SE(\hat{\phi}_{kk})$

2.2.4 การสแตชันนารีของอนุกรมเวลา (Stationary of Time Series)

การสแตชันนารีของอนุกรมเวลา หมายถึง อนุกรมเวลาอยู่ในสมดุลเชิงสถิติ (Statistical Equilibrium) ซึ่งคุณสมบัติทางสถิติของอนุกรมเวลาไม่เปลี่ยนแปลงเมื่อเวลาเปลี่ยนไป กำหนดให้

1. $Z_t, Z_{t+1}, Z_{t+2}, \dots, Z_{t+k}$ เป็นอนุกรมเวลาที่เวลา $t, t+1, t+2, \dots, t+k$
2. $Z_{t+m}, Z_{t+m+1}, Z_{t+m+2}, \dots, Z_{t+m+k}$ เป็นอนุกรมเวลาที่เวลา $t+m, t+m+1, \dots, t+m+k$
2. $P(Z_t, Z_{t+1}, Z_{t+2}, \dots, Z_{t+k})$ เป็นการแจกแจงความน่าจะเป็นร่วม (joint probability distribution) ของ $Z_t, Z_{t+1}, Z_{t+2}, \dots, Z_{t+k}$
- และ 4. $P(Z_{t+m}, Z_{t+m+1}, Z_{t+m+2}, \dots, Z_{t+m+k})$ เป็นการแจกแจงความน่าจะเป็นร่วมของ $Z_{t+m}, Z_{t+m+1}, Z_{t+m+2}, \dots, Z_{t+m+k}$

อนุกรมเวลา Z_t จะเรียกว่าเป็นอนุกรมเวลาสแตชันนารีแบบเข้ม (Strong Stationary Time Series) เมื่อ

$$P(Z_t, Z_{t+1}, Z_{t+2}, \dots, Z_{t+k}) = P(Z_{t+m}, Z_{t+m+1}, Z_{t+m+2}, \dots, Z_{t+m+k})$$

และอนุกรมเวลา Z_t จะเรียกว่าเป็นอนุกรมเวลาเสถียรแบบอ่อน (Weak Stationary Time Series) เมื่อลักษณะการแจกแจงความน่าจะเป็นร่วมไม่เปลี่ยนแปลงเมื่อเวลาเปลี่ยนไป เช่น ค่าเฉลี่ย (mean) ความแปรปรวน (variance) และความแปรปรวนร่วม (covariance) ของอนุกรมเวลามีค่าคงที่ทุกเวลา นั่นคือ

1. ค่าเฉลี่ยของ Z ที่เวลา t เท่ากับค่าเฉลี่ยของ Z ที่เวลา $t+m$
หรือ $E(Z_t) = E(Z_{t+m})$
2. ความแปรปรวนของ Z ที่เวลา t เท่ากับความแปรปรวนของ Z ที่เวลา $t+m$
หรือ $E(Z_t - \mu)^2 = E(Z_{t+m} - \mu)^2$
3. ความแปรปรวนร่วมของ Z ที่เวลา t เท่ากับความแปรปรวนร่วมของ Z ที่เวลา $t+m$ หรือ $E(Z_t - \mu)(Z_{t+k} - \mu) = E(Z_{t+m} - \mu)(Z_{t+m+k} - \mu)$

การประมาณค่าเฉลี่ย (μ) และความแปรปรวน (σ_z^2) จะประมาณด้วยค่าเฉลี่ย ($\bar{Z}$) และความแปรปรวน (S_z^2) ของอนุกรมเวลา Z_t โดย

$$\bar{Z} = \frac{\sum_{i=1}^n Z_i}{n} \quad (2.14)$$

$$S_z^2 = \frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n} \quad (2.15)$$

อนุกรมเวลาที่ไม่มีความสมบัติเป็นเสถียรแบบอ่อนจะไม่สามารถหารูปแบบให้กับอนุกรมเวลาได้ จะต้องแปลงอนุกรมเวลานั้นให้เป็นอนุกรมเวลาใหม่ที่มีความสมบัติเสถียรแบบอ่อน (Stationary) เสียก่อนจึงจะหารูปแบบให้กับอนุกรมเวลาใหม่ได้ การแปลงอนุกรมเวลาให้เป็นอนุกรมเวลาใหม่ที่เป็นเสถียรแบบอ่อนสามารถทำได้ด้วยวิธีการต่างๆ ดังนี้

1. การหาผลต่าง (Differencing)

เนื่องจากรูปแบบของอนุกรมเวลาของ Box-Jenkins จะเขียนอยู่ในรูปของอนุกรมเวลา ย้อนหลัง จึงกำหนดสัญลักษณ์ย้อนหลังเป็น B เรียกว่า Backward shift operator

$$\text{เมื่อ } BZ_t = Z_{t-1}$$

$$B^m Z_t = Z_{t-m}$$

กำหนดให้ ∇ เป็นผลต่างครั้งที่ 1 และ ∇^d เป็นผลต่างครั้งที่ d จะได้ว่า

$$\begin{aligned}\nabla Z_t &= (1-B)Z_t \\ &= Z_t - Z_{t-1} \\ \nabla^d Z_t &= (1-B)^d Z_t\end{aligned}$$

เมื่ออนุกรมเวลาไม่เป็นสเตชันนารีจะหาผลต่างครั้งที่ 1 ก่อนแล้วพิจารณาว่าอนุกรมเวลาของผลต่างครั้งที่ 1 เป็นสเตชันนารีหรือไม่ ถ้าอนุกรมเวลายังไม่สเตชันนารีก็จะหาผลต่างครั้งที่ 2 ต่อไปจนกระทั่งได้อนุกรมเวลาที่เป็นสเตชันนารี แต่ถ้าพบว่าอนุกรมเวลาเป็นสเตชันนารีแล้วก็ไม่จำเป็นต้องหาผลต่างเพราะผลต่างของอนุกรมเวลาที่เป็นสเตชันนารีจะเป็นอนุกรมเวลาที่เป็นสเตชันนารีด้วย

2. การหาผลต่างฤดูกาล (Seasonal Differencing)

เมื่ออนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้องก็จะต้องแปลงอนุกรมเวลาเดิมให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาลโดยการหาผลต่างของฤดูกาล ซึ่งจะคล้ายกับการหาผลต่างของข้อมูลอนุกรมเวลาที่ไม่มีฤดูกาล แต่จะหาผลต่างของข้อมูลที่มีช่วงห่างกัน S หน่วยเวลา และ D เป็นจำนวนครั้งในการหาผลต่างของฤดูกาล เพื่อให้อนุกรมเวลาที่ไม่เป็นสเตชันนารีเนื่องจากฤดูกาลเป็นอนุกรมเวลาที่เป็นสเตชันนารี

$$\begin{aligned}\text{โดย } \nabla_S Z_t &= (1-B^S)Z_t \\ &= Z_t - Z_{t-S} \\ \nabla_S^D Z_t &= (1-B^S)^D Z_t\end{aligned}$$

3. การแปลงเป็นค่ารากที่สอง (Square Root Transformation)

เมื่อค่าสังเกต Z_t ของอนุกรมเวลาไม่มีคุณสมบัติเป็นสเตชันนารี สามารถแปลงอนุกรมเวลาโดยการหารากของค่าสังเกต Z_t ได้จากสมการ

$$W_t = (Z_t + a)^{0.5} \quad (2.16)$$

เมื่อ

W_t คือ อนุกรมเวลาใหม่ที่เป็นสเตชันนารี โดยการหารากของค่าสังเกต Z_t

a คือ ค่าคงที่ , $a > \{|Z_t|, \text{ for } Z_t < 0\}$

4. การแปลงเป็นค่าลอการิทึม (Logarithmic Transformation)

การแปลงอนุกรมเวลาเดิมให้เป็นอนุกรมเวลาใหม่โดยการหาค่าลอการิทึมของค่าสังเกต Z_t เมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ สำหรับเวลา t ใดๆ

$$W_t = \log(Z_t + a) \quad (2.17)$$

เมื่อ

a คือ ค่าคงที่ , $a > \{|Z_t|, \text{ for } Z_t < 0\}$

5. การแปลงโดยวิธีบอกซ์-คอกซ์ (Box-Cox Transformations)

เมื่อการแปลงอนุกรมเวลาโดยวิธีการหารากของค่าสังเกตและค่าลอการิทึมของค่าสังเกต ไม่สามารถแปลงอนุกรมเวลาให้เป็นสเตชันนารีได้ สามารถแปลงเป็นอนุกรมที่สเตชันนารีได้โดยวิธีของ Box-Cox ดังสมการ (2.18)

$$W_t = (Z_t + a)^\lambda, \quad -2 < \lambda < 2 \quad (2.18)$$

6. การแปลงแบบนอนพารามेटริก (Non-Parametric Transformation)

เมื่ออนุกรมเวลามีอิทธิพลของแนวโน้มหรือฤดูกาลและการหาผลต่างฤดูกาลของอนุกรมเวลาไม่สามารถแปลงอนุกรมเวลาให้เป็นอนุกรมเวลาที่สเตชันนารีได้ สามารถแปลงอนุกรมเวลาได้โดยวิธี Non-Parametric Transformation ดังนี้

$$W_t = \frac{Z_{ij} - ZM_j}{SD_j} \quad (2.19)$$

เมื่อ

Z_{ij} คือ ข้อมูลของเดือนที่ j และปีที่ i

ZM_j คือ ค่าเฉลี่ยของเดือนที่ j

SD_j คือ ค่าส่วนเบี่ยงเบนมาตรฐานของเดือนที่ j

และ $t = (12)(i - 1) + j$

7. การวิเคราะห์ฮาร์โมนิก (Harmonic Analysis)

เป็นการแปลงอนุกรมเวลาอย่างหนึ่งที่สามารถกำจัดอิทธิพลเนื่องจากแนวโน้มหรือฤดูกาลให้ออกจากข้อมูล โดยมีวิธีการดังนี้

$$W_t = Z_t - \left[\mu + \sum_{j=1}^M (a_j \cos(\frac{2\pi t}{T_j}) + b_j \sin(\frac{2\pi t}{T_j})) \right] \quad (2.20)$$

เมื่อ

$$a_j = \frac{2}{N} \sum_{i=1}^N Z_i \cos\left(\frac{2\pi i}{T_j}\right); \quad j = 1, 2, 3, \dots, M$$

$$b_j = \frac{2}{N} \sum_{i=1}^N Z_i \sin\left(\frac{2\pi i}{T_j}\right); \quad j = 1, 2, 3, \dots, M$$

M คือ จำนวนของฮาร์โมนิก

T_j คือ คาบที่สอดคล้องกับฮาร์โมนิกที่ j , หาก $\left\{ \frac{T_j}{i} \right\}$

2.2.5 รูปแบบอนุกรมเวลาของ Box-Jenkins

อนุกรมเวลาของ Box-Jenkins สามารถจำแนกออกเป็นอนุกรมเวลาที่มีคุณสมบัติเป็นสเตชันนารี และอนุกรมเวลาที่ไม่เป็นสเตชันนารีซึ่งจะต้องแปลงอนุกรมเวลาให้เป็นสเตชันนารีเสียก่อน รูปแบบอนุกรมเวลาของ Box-Jenkins สามารถแยกได้เป็น 4 รูปแบบดังนี้

1. Autoregressive Processes, AR(p)

รูปแบบ AR(p) เป็นรูปแบบที่ค่าสังเกต Z_t ขึ้นอยู่กับค่าสังเกตที่ผ่านมา $Z_{t-j}; j=1, 2, 3, \dots, p$ หรือ

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{Z}_t = a_t \quad (2.21)$$

$$\phi_p(B) \tilde{Z}_t = a_t$$

เมื่อ

$\tilde{Z}_t = Z_t - \mu$ เป็นอนุกรมเวลาที่เบี่ยงเบนไปจากค่าเฉลี่ย

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

สมการที่ (2.21) เรียกว่ากระบวนการถดถอยในตัวเองลำดับที่ p (Autoregressive Processes of Order p) เมื่อ a_t เป็นสิ่งรบกวนอย่างสุ่ม (White noise) มีการแจกแจงแบบปกติ (normal) ที่เหมือนกันทุกเวลาและเป็นอิสระต่อกัน โดยมีค่าเฉลี่ยเป็นศูนย์และความแปรปรวนเป็น σ_a^2 เนื่องจาก a_t มีการแจกแจงแบบปกติที่เป็นอิสระต่อกัน ดังนั้นจะไม่มีสหสัมพันธ์ในตัวเองและไม่สามารถใช้ค่า a_t ในอดีตในการพยากรณ์ค่า a_t ในอนาคตได้ รูปแบบ AR(p) มีพารามิเตอร์ที่ไม่ทราบค่า $p+2$ ตัว คือ $\mu, \phi_1, \phi_2, \dots, \phi_p, \sigma_a^2$

2. Moving Average Processes, MA(q)

เป็นรูปแบบของอนุกรมเวลาที่มีคุณสมบัติเป็นสเตชันนารีที่ค่าสังเกต Z_t ขึ้นอยู่กับค่าความคลาดเคลื่อน $a_{t-j}; j=1,2,3,\dots,q$ หรือขึ้นอยู่กับความคลาดเคลื่อนที่อยู่ก่อนหน้านั้น q ค่า ดังสมการที่ (2.22)

$$\begin{aligned}\tilde{Z}_t &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \\ \tilde{Z}_t &= \theta_q(B) a_t\end{aligned}\tag{2.22}$$

เมื่อ

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

สมการที่ (2.22) เรียกว่ากระบวนการเฉลี่ยเคลื่อนที่ลำดับที่ q (Moving Average Processes of Order q) รูปแบบ MA(q) มีพารามิเตอร์ที่ไม่ทราบค่า $q+2$ ตัว คือ $\mu, \theta_1, \theta_2, \dots, \theta_q, \sigma_a^2$

3. Autoregressive Moving Average ARMA(p, q) Processes

เป็นรูปแบบของอนุกรมเวลาที่มีคุณสมบัติเป็นสเตชันนารีที่ค่าสังเกต Z_t จะขึ้นอยู่กับค่าสังเกตและค่าความคลาดเคลื่อน a_t ในอดีตที่เกิดขึ้นก่อนหน้า นั่นคือเป็นรูปแบบผสมของ AR และ MA ดังนี้

$$\begin{aligned}(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{Z}_t &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \\ \phi_p(B) \tilde{Z}_t &= \theta_q(B) a_t\end{aligned}\tag{2.23}$$

สมการที่ (2.23) เรียกว่ากระบวนการผสมการถดถอยในตัวเองและการเฉลี่ยเคลื่อนที่ลำดับที่ p, q (Autoregressive Moving Average Processes of Order p, q) รูปแบบ ARMA(p, q) มีพารามิเตอร์ที่ไม่ทราบค่า $p+q+2$ ตัว คือ $\mu, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma_a^2$

4. Autoregressive Integrated Moving Average ARIMA(p, d, q) Processes

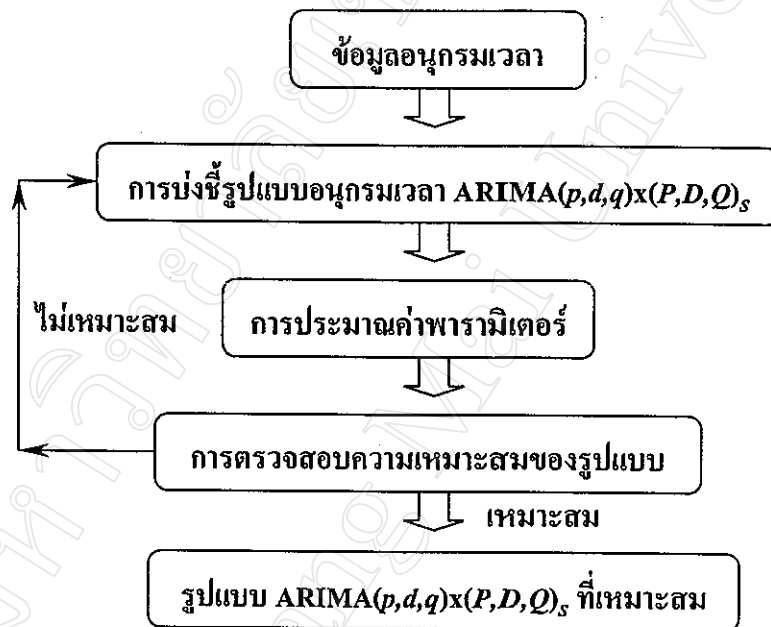
เป็นรูปแบบอนุกรมเวลาที่ไม่มีความสัมพันธ์เป็นสเตชันนารี เมื่อจะหารูปแบบ ARMA(p, q) ให้แปลงเป็นอนุกรมเวลาที่มีความสัมพันธ์สเตชันนารีเสียก่อน การแปลงอนุกรมเวลาเดิมให้เป็นอนุกรมเวลาที่เป็นสเตชันนารีทำได้โดยการหาผลต่าง ซึ่งจะอยู่ในรูปแบบ ARIMA (p, d, q) เมื่อ d เป็นจำนวนครั้งในการหาผลต่าง หรือ

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d \tilde{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (2.24)$$

$$\phi_p(B)(1 - B)^d \tilde{Z}_t = \theta_q(B) a_t$$

2.2.6 ขั้นตอนในการหารูปแบบอนุกรมเวลาของ Box-Jenkins

การวิเคราะห์อนุกรมเวลาของ Box-Jenkins มีขั้นตอนในการหารูปแบบอนุกรมเวลาอยู่ด้วยกัน 3 ขั้นตอน คือ การบ่งชี้รูปแบบ การประมาณค่าพารามิเตอร์ และตรวจสอบรูปแบบ ดังแสดงในรูปที่ 2.1



รูปที่ 2.1 ขั้นตอนการสร้างรูปแบบอนุกรมเวลาของ Box-Jenkins
ที่มา : Lin Lee-Yu (1982)

ก. การบ่งชี้รูปแบบ (Identification)

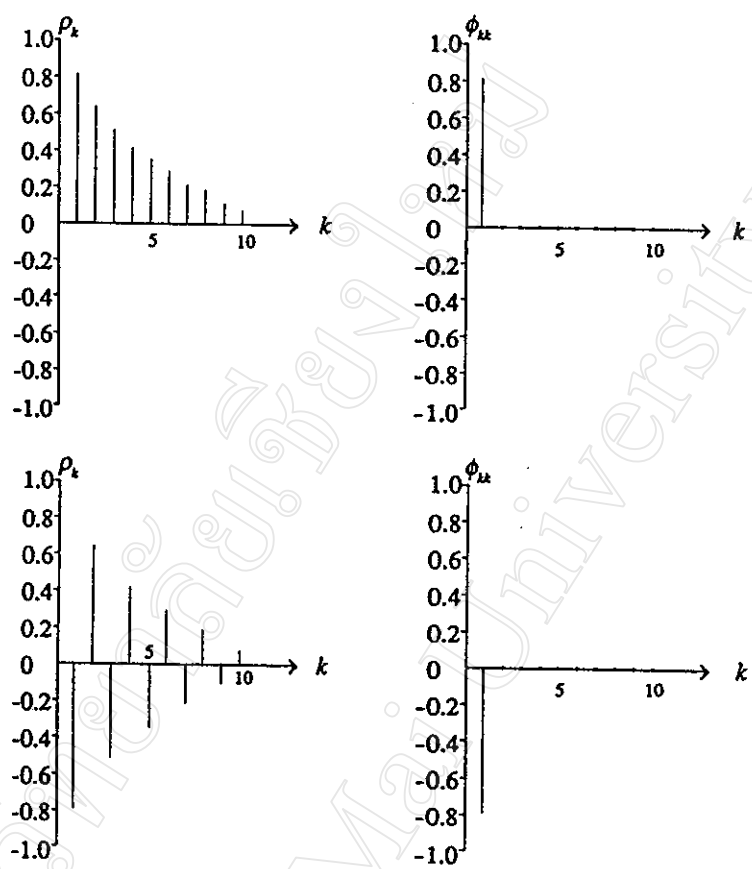
การบ่งชี้รูปแบบเป็นการหารูปแบบที่เหมาะสมของอนุกรมเวลาโดยพิจารณาจากฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function, ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function, PACF) โดยมีขั้นตอนดังต่อไปนี้

1. สร้างกราฟความสัมพันธ์ระหว่างอนุกรมเวลากับช่วงห่างเวลา k เพื่อดูลักษณะการเปลี่ยนแปลงของอนุกรมเวลาว่ามีแนวโน้มหรือมีการเปลี่ยนแปลงของฤดูกาลหรือไม่
2. สร้างกราฟความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (ACF) และสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วน (PACF) กับช่วงห่าง k หน่วยเวลา เพื่อใช้หารูปแบบของอนุกรมเวลา $ARIMA(p,d,q)x(P,D,Q)_s$
3. พิจารณาลักษณะ ACF และ PACF ของอนุกรมเวลา เมื่ออนุกรมเวลาไม่เป็นสเตชันนารีจะต้องแปลงอนุกรมเวลาให้เป็นสเตชันนารีเสียก่อน โดยการหาผลต่างหรือผลต่างของฤดูกาล จึงจะหารูปแบบให้กับอนุกรมเวลาได้
4. พิจารณาเลือกรูปแบบอนุกรมเวลาจากตารางที่ 2.1 และรูปที่ 2.2-2.6 แสดงคอเรลโรแกรม ACF และ PACF ของอนุกรมเวลารูปแบบ $ARIMA(p,d,q)x(P,D,Q)_s$

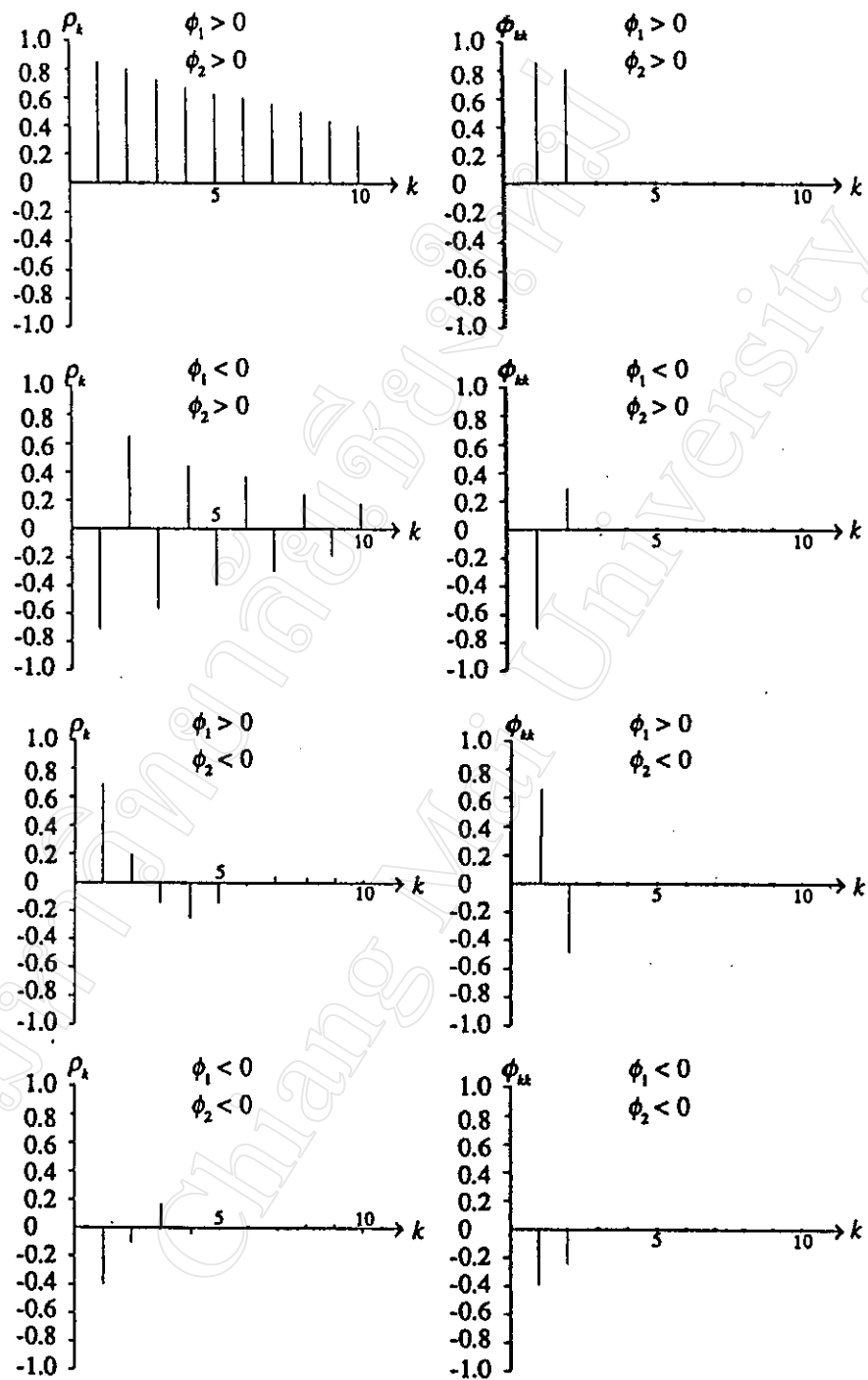
ตารางที่ 2.1 แสดงลักษณะของ ACF และ PACF สำหรับรูปแบบ ARIMA(p, d, q)

	Order				
	(1, d, 0)	(0, d, 1)	(2, d, 0)	(0, d, 2)	(1, d, 1)
Behavior of ρ_k	Decays exponentially	Only ρ_1 nonzero	Mixture of exponentials or damped sine wave	Only ρ_1 and ρ_2 nonzero	Decays exponentially from first lag
Behavior of ϕ_k	Only ϕ_1 nonzero	Exponential dominates decay	Only ϕ_1 and ϕ_2 nonzero	Dominated by mixture of exponentials or damped sine wave	Dominated by exponential decay from first lag
Preliminary estimates from:	$\phi_1 = \rho_1$	$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$	$\phi_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2}$ $\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$	$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}$ $\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$	$\rho_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$ $\rho_2 = \rho_1\phi_1$ $-1 < \phi_1 < 1$ $-1 < \theta_1 < 1$
Admissible region	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$	$-1 < \phi_2 < 1$ $\phi_2 + \phi_1 < 1$ $\phi_2 - \phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$	

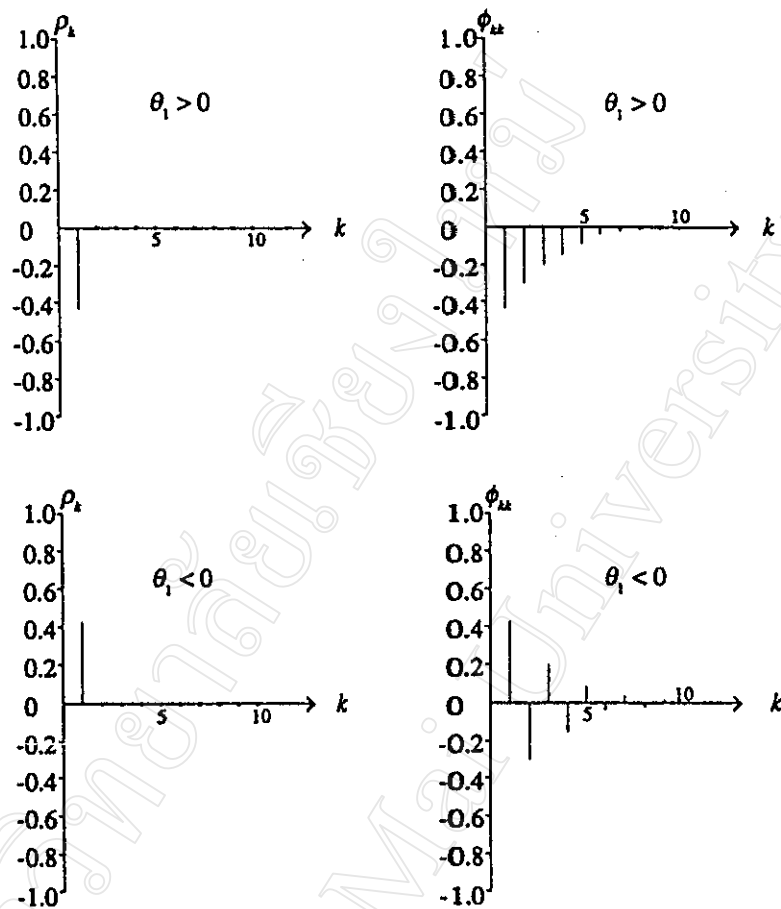
ที่มา: Box (1994)



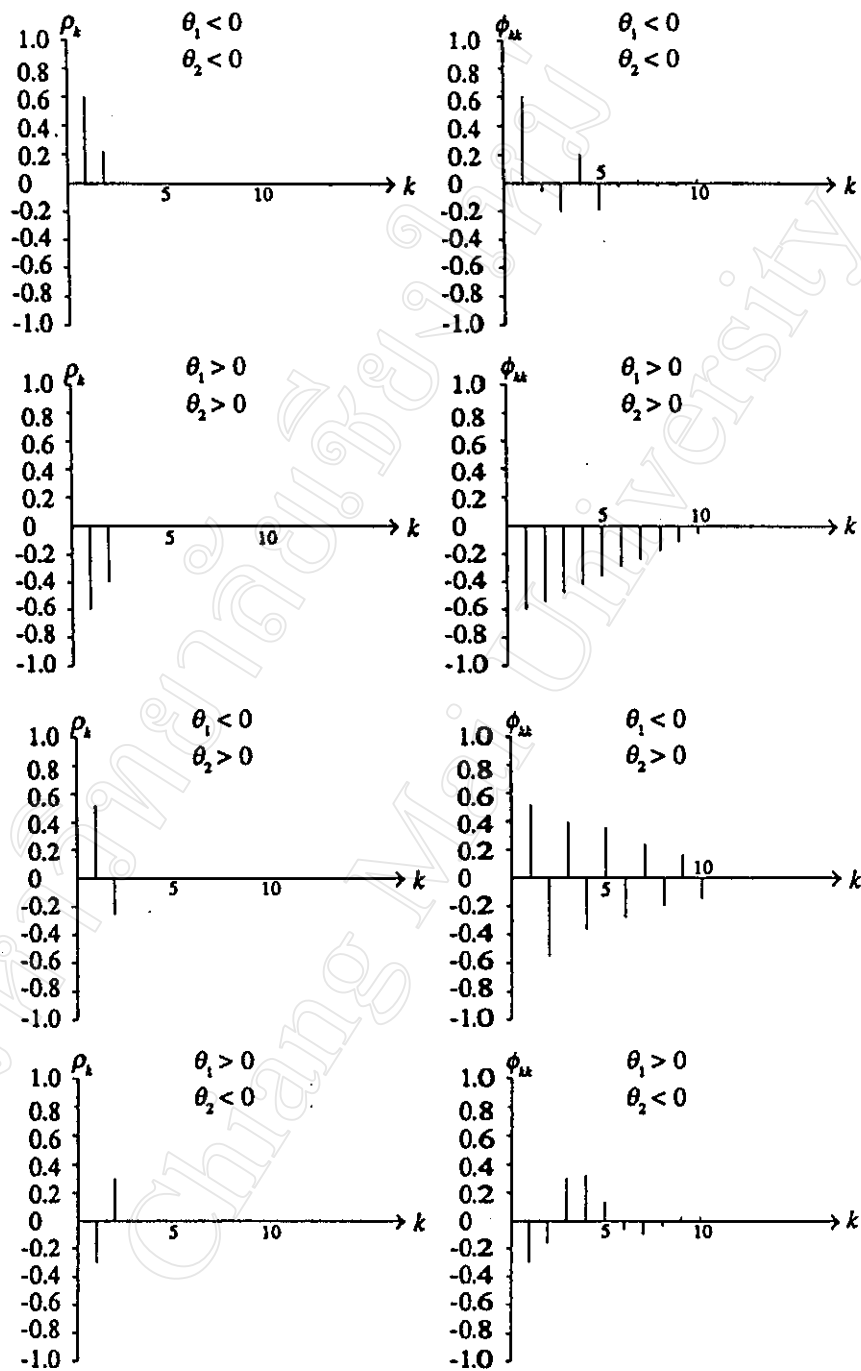
รูปที่ 2.2 แสดงลักษณะ ACF และ PACF ของรูปแบบ AR(1)



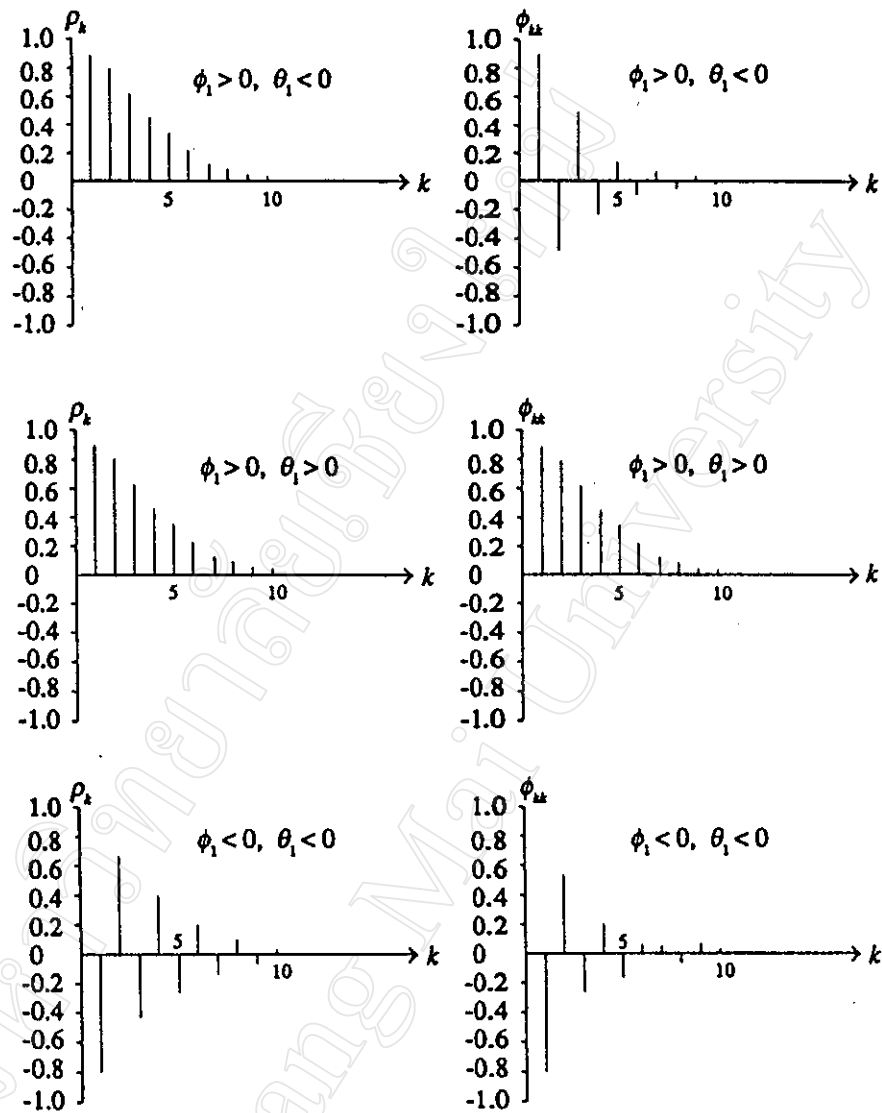
รูปที่ 2.3 แสดงลักษณะ ACF และ PACF ของรูปแบบ AR(2)



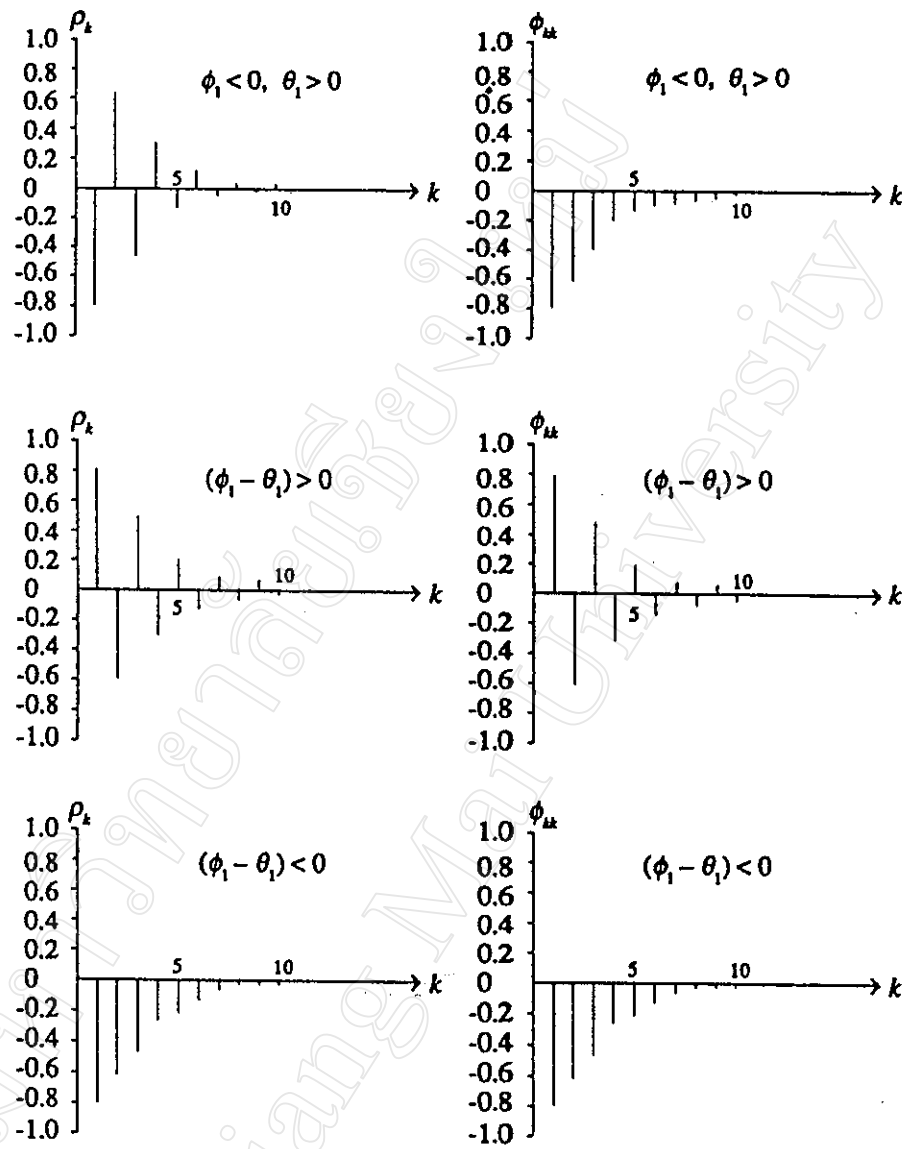
รูปที่ 2.4 แสดงลักษณะ ACF และ PACF ของรูปแบบ MA(1)



รูปที่ 2.5 แสดงลักษณะ ACF และ PACF ของรูปแบบ MA(2)



รูปที่ 2.6 แสดงลักษณะ ACF และ PACF ของรูปแบบ ARMA(1,1)



รูปที่ 2.6(ต่อ) แสดงลักษณะ ACF และ PACF ของรูปแบบ ARMA(1,1)

ข. การประมาณค่าพารามิเตอร์ (Estimation)

หลังจากได้กำหนดรูปแบบให้กับอนุกรมเวลาแล้ว ขั้นตอนต่อไปเป็นหาค่าประมาณของพารามิเตอร์เพื่อจะนำไปใช้ในการพยากรณ์ การประมาณค่าพารามิเตอร์ที่ดีที่สุดเป็นการหาค่าของความคลาดเคลื่อนกำลังสองในการพยากรณ์ให้มีค่าน้อยที่สุด

วิธีการประมาณค่าพารามิเตอร์ของรูปแบบที่ดีที่สุด คือ วิธีแมกซิมัมไลกelihood (Maximum Likelihood Method) โดยที่ค่าคลาดเคลื่อน (residual) ก็คือผลต่างของค่าจริงกับค่าพยากรณ์ การคำนวณค่าประมาณของ a_t หรือ $\hat{a}_t$ มี 2 วิธี ดังนี้

ข.1 การคำนวณแบบมีเงื่อนไข (Conditional Likelihood Estimates)

สำหรับอนุกรมเวลารูปแบบ ARIMA(p, d, q) เมื่อ $W_t = \nabla^d Z_t$ และ $\tilde{W}_t = W_t - \mu$ โดยที่ $E(W_t) = \mu$ ซึ่งสามารถเขียนรูปแบบสมการได้ดังนี้

$$\begin{aligned}\tilde{W}_t &= \phi_1 \tilde{W}_{t-1} + \phi_2 \tilde{W}_{t-2} + \dots + \phi_p \tilde{W}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \\ \text{หรือ} \quad a_t &= \tilde{W}_t - \phi_1 \tilde{W}_{t-1} - \phi_2 \tilde{W}_{t-2} - \dots - \phi_p \tilde{W}_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}\end{aligned}$$

จากสมการข้างต้นในการคำนวณหาค่าความคลาดเคลื่อน a_t จะต้องทราบค่าเริ่มต้นของ W_* จำนวน p ตัวของอนุกรม W และค่าเริ่มต้น a_* จำนวน q ตัวของอนุกรม a_t ดังนั้น W_* และ a_* จึงเป็นเงื่อนไขในการคำนวณหาค่าความคลาดเคลื่อน a_t โดยกำหนดให้ค่าเริ่มต้น $W_* = 0$ และค่าเริ่มต้น $a_* = 0$ จากนั้นแทนค่าพารามิเตอร์ (ϕ, θ) และค่าเริ่มต้น (W_*, a_*) ก็จะสามารถคำนวณหาค่าความคลาดเคลื่อน $a_t(\phi, \theta / W_*, a_*, W); t = 1, 2, \dots, n$ ได้ เลือกค่าพารามิเตอร์ (ϕ, θ) ที่ให้ค่าความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด โดยพิจารณาจากฟังก์ชันค่าความคลาดเคลื่อนกำลังสองแบบมีเงื่อนไข (Conditional least square estimates) ดังนี้

$$S_*(\phi, \theta) = \sum_{t=1}^n a_t^2(\phi, \theta / W_*, a_*, W) \quad (2.25)$$

ข.2 การคำนวณแบบไม่มีเงื่อนไข (Unconditional Likelihood Estimates)

การประมาณค่าพารามิเตอร์โดยการคำนวณแบบไม่มีเงื่อนไข จะกำหนดให้ $S(\phi, \theta)$ เป็นฟังก์ชันค่าความคลาดเคลื่อนกำลังสองแบบไม่มีเงื่อนไข การคำนวณหาค่า $S(\phi, \theta)$ จะใช้วิธีการคำนวณแบบย้อนหลัง พิจารณาอนุกรมเวลารูปแบบ ARIMA(0,1,1)

$$\begin{aligned}\tilde{W}_t &= a_t - \theta_1 a_{t-1} \\ &= (1 - \theta_1 B) a_t\end{aligned}$$

เปลี่ยนรูปสมการจากรูปแบบ backward เป็นรูปแบบ forward ได้ดังนี้

$$\tilde{W}_t = (1 - \theta_1 F) e_t$$

โดยที่

$$FZ_t = Z_{t+1} \text{ และ } F^m Z_t = Z_{t+m}$$

สมมติให้ $\mu = E(W_t) = 0$

จะได้

$$W_t = e_t - \theta_1 e_{t+1}$$

$$[e_t] = [W_t] + \theta_1 [e_{t+1}]$$

$$[a_t] = [W_t] + \theta_1 [a_{t-1}]$$

เมื่อ $[W_t] = W_t$ สำหรับ $t = 1, 2, \dots, n$ และเป็นค่าพยากรณ์ย้อนหลัง (Back-forecast) ของ W_t สำหรับ $t \leq 0$

การคำนวณค่า e_t จะกำหนดให้ $e_t = 0; t = n+1$ และ $e_t = 0; t = 0, -1, -2, \dots$ ซึ่งจะสามารถหาค่าพยากรณ์ย้อนหลังของ W_t ได้ จากนั้นก็จะสามารถคำนวณค่าคลาดเคลื่อน a_t ได้โดยเลือกค่าพารามิเตอร์ (ϕ, θ) ที่ให้ค่าความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด โดยพิจารณาจากฟังก์ชันค่าความคลาดเคลื่อนกำลังสองแบบไม่มีเงื่อนไข (Unconditional least square estimates) ดังนี้

$$S(\phi, \theta) = \sum_{t=-\infty}^n [a_t / \phi, \theta, W]^2 \quad (2.26)$$

ค. การตรวจสอบรูปแบบ (Diagnostic Checking)

เมื่อได้รูปแบบและประมาณค่าพารามิเตอร์แล้ว จะต้องมีการตรวจสอบรูปแบบที่ได้ว่ามีความเหมาะสมที่จะนำไปใช้ในการพยากรณ์หรือไม่ รูปแบบเหมาะสมเมื่อค่า $\hat{a}_t$ เป็นอิสระต่อกัน หรือค่าสัมประสิทธิ์สหสัมพันธ์ของค่าคลาดเคลื่อนไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ โดยการทดสอบสมมติฐานดังนี้

$$H_0 : \rho_1(\hat{a}_t) = \rho_2(\hat{a}_t) = \dots = \rho_k(\hat{a}_t) = 0$$

และ $H_0 : \rho_k(\hat{a}_t) = 0$

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของค่าคลาดเคลื่อน $\rho_k(\hat{a}_t)$ จะคำนวณเช่นเดียวกับสัมประสิทธิ์สหสัมพันธ์ในตัวเองของอนุกรมเวลา ดังนี้

$$\hat{\rho}_k(\hat{a}_t) = \frac{\sum_{t=1}^{n-k} (a_t - \bar{a})(a_{t+k} - \bar{a})}{\sum_{t=1}^n (a_t - \bar{a})^2} \quad (2.27)$$

ค.1 การทดสอบสมมติฐาน $H_0 : \rho_1(\hat{a}_t) = \rho_2(\hat{a}_t) = \dots = \rho_k(\hat{a}_t) = 0$

เป็นการทดสอบว่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของค่าคลาดเคลื่อนทุกช่วงห่าง K หนึ่งเวลามีความเป็นอิสระต่อกันหรือไม่ โดยใช้

Modified portmanteau lack of fit test (Q)

เมื่อกำหนด Q นิยามดังสมการ (2.28)

$$Q = n(n+2) \sum_{j=1}^K \frac{\hat{\rho}_j^2(\hat{a}_t)}{n-j} \quad (2.28)$$

Q มีการแจกแจงแบบไคสแควร์ (Chi-square) $\chi_{\alpha, K-m}^2$ ที่ดัชนีความอิสระ $K-m$ เมื่อ m เป็นจำนวนพารามิเตอร์ที่ได้จากการประมาณ และ n เป็นจำนวนข้อมูลในอนุกรมเวลา

โดยจะยอมรับสมมติฐาน เมื่อ $Q \leq \chi_{\alpha, K-m}^2$ นั่นคือค่าคลาดเคลื่อน $\hat{a}_t$ เป็นอิสระต่อกันที่ช่วงห่าง K และจะปฏิเสธสมมติฐานเมื่อ $Q > \chi_{\alpha, K-m}^2$ นั่นคือสัมประสิทธิ์สหสัมพันธ์ในตัวเองของค่าคลาดเคลื่อนมีค่าอย่างน้อยหนึ่งค่าไม่เท่ากับศูนย์ แสดงว่า $\hat{a}_t$ ไม่เป็นอิสระต่อกันที่ช่วงห่าง K และรูปแบบที่พิจารณานั้นไม่เหมาะสมจะต้องกำหนดรูปแบบให้กับอนุกรมเวลาใหม่ จนกว่าจะได้รูปแบบที่เหมาะสม

ค.2 การทดสอบสมมติฐาน $H_0 : \rho_k(\hat{a}_t) = 0$

การทดสอบความเป็นอิสระของค่าคลาดเคลื่อน $\hat{a}_t$ จะทดสอบโดยใช้สถิติ t ดังสมการที่ (2.29)

$$t = \frac{\rho_k(\hat{a}_t) - 0}{SE\{\rho_k(\hat{a}_t)\}} \quad (2.29)$$

เมื่อ $SE\{\rho_k(\hat{a}_t)\}$ เป็นส่วนเบี่ยงเบนมาตรฐานของสัมประสิทธิ์สหสัมพันธ์ในตัวเองของค่าคลาดเคลื่อน และจะยอมรับสมมติฐาน $H_0: \rho_k(\hat{a}_t) = 0$ เมื่อ $|t| \leq 1.96SE\{\rho_k(\hat{a}_t)\}$ แสดงว่าค่าคลาดเคลื่อนเป็นอิสระต่อกัน และรูปแบบที่กำหนดให้กับอนุกรมเวลามีความเหมาะสม

รูปแบบที่เหมาะสมอาจมีหลายรูปแบบ เพราะฉะนั้นหลังจากที่ได้รูปแบบที่เหมาะสมแล้วต่อไปจะเป็นการเลือกรูปแบบที่ดีที่สุดให้กับอนุกรมเวลา โดยการพิจารณาจากค่า Akaike Information Criteria (AIC) และ Schwartz' SBC Criteria (SBC) รูปแบบของอนุกรมเวลาที่ให้ค่า AIC และ SBC น้อยที่สุดจะเป็นรูปแบบของอนุกรมเวลาที่ดีที่สุด

1. Akaike Information Criteria (AIC) and Modified AIC

Akaike (1973, 1974a) นิยามค่า AIC ว่า

$$\begin{aligned} AIC(M) &= -2 \ln(\text{maximum likelihood}) + 2M \quad \text{หรือ} \\ AIC(M) &= n \ln \hat{\sigma}_a^2 + 2M \end{aligned} \quad (2.30)$$

Shibata (1976) ได้แสดงให้เห็นว่า AIC เป็นการประมาณค่าที่ให้ค่าสูงเกินจริงสำหรับสัมประสิทธิ์สหสัมพันธ์อันดับสูงๆ ดังนั้น Akaike (1978, 1979) จึงได้พัฒนาสมการให้มีความเหมาะสมยิ่งขึ้นเรียกว่า Modified AIC ดังสมการที่ (2.31)

$$\begin{aligned} BIC(M) &= n \ln \hat{\sigma}_a^2 - (n - M) \ln \left\{ 1 - \frac{M}{n} \right\} + M \ln n \\ &\quad + M_z \ln \left[\left(\frac{\hat{\sigma}_z^2}{\hat{\sigma}_a^2} - 1 \right) / M \right] \end{aligned} \quad (2.31)$$

เมื่อ

M คือ จำนวนพารามิเตอร์ , $M = p + d + P + Q$

n คือ จำนวนข้อมูล , $n = N - d - (S)(D)$

N คือ จำนวนข้อมูลทั้งหมด

d คือ จำนวนครั้งในการหาผลต่างของอนุกรมเวลา

S คือ ช่วงเวลาของฤดูกาล (Seasonal period)

D คือ จำนวนครั้งในการหาผลต่างของฤดูกาล

$\hat{\sigma}_a^2$ คือ ค่าประมาณความแปรปรวนของ σ_a^2 (Maximum likelihood)

$\hat{\sigma}_z^2$ คือ ค่าประมาณความแปรปรวนของตัวอย่าง (Sample variance of the series)

2. Schwartz Bayesian Criterion (SBC)

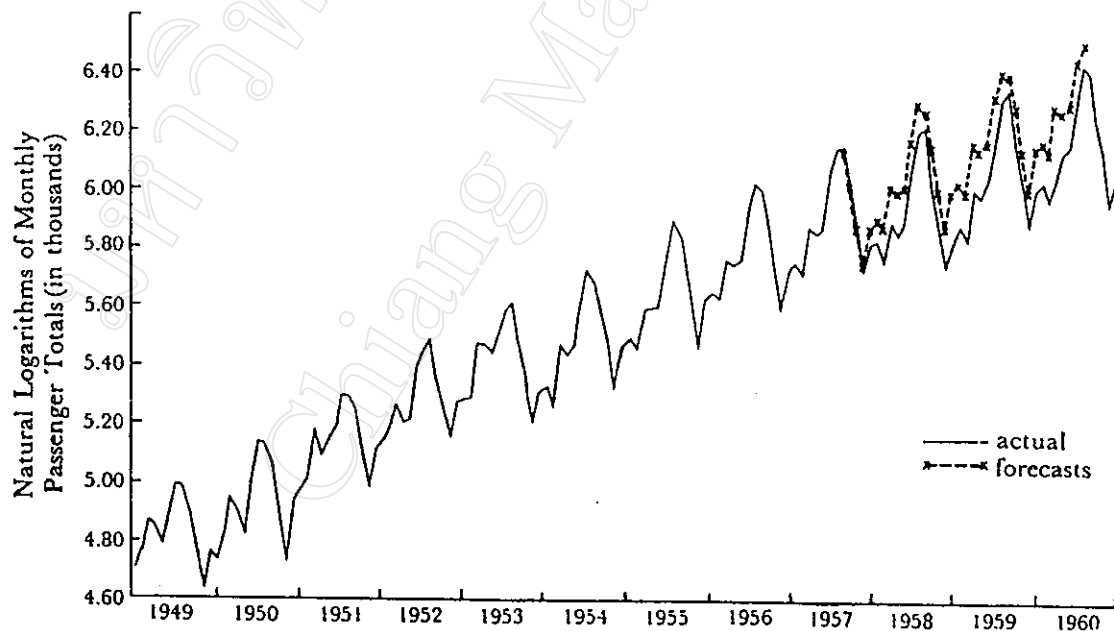
คล้ายกับรูปแบบของ Akaike Information Criteria(AIC), Schwartz (1978) ได้เสนอเกณฑ์ในการเลือกรูปแบบที่ดีที่สุดของอนุกรมเวลา ดังสมการ (2.32)

$$SBC = n \ln \hat{\sigma}_a^2 + M \ln n \quad (2.32)$$

2.2.7 การวิเคราะห์ความผันแปรตามฤดูกาล

การเปลี่ยนแปลงหรือความเคลื่อนไหวตามฤดูกาล (Seasonal Variation) เป็นส่วนประกอบหนึ่งของอนุกรมเวลาหลายๆ อนุกรม ในการพิจารณาความเปลี่ยนแปลงตามฤดูกาล หน่วยของระยะเวลาในการพิจารณาอาจเป็นรายเดือน หรือรายไตรมาส

ส่วนมากอนุกรมเวลาที่มีฤดูกาลจะมีลักษณะซ้ำกันเป็นช่วง (Periodic) ซึ่งจะมีลักษณะคล้ายกันทุกช่วงห่าง S หน่วยเวลา ลักษณะอนุกรมเวลาที่มีลักษณะเป็นช่วงๆ นี้เป็นการเปลี่ยนแปลงตามฤดูกาล โดยที่ S แทนความยาวของช่วงเวลา เช่น สำหรับข้อมูลรายเดือน $S = 12$ และข้อมูลรายไตรมาส $S = 4$ เป็นต้น รูปที่ 2.7 แสดงอนุกรมเวลาที่มีฤดูกาล



รูปที่ 2.7 แสดงลักษณะข้อมูลอนุกรมเวลาที่มีฤดูกาลของ International Air Travel
ที่มา Box (1994)

รูปแบบอนุกรมเวลาที่มีฤดูกาลของ Box-Jenkins ได้แก่วิธีแบบ Seasonal ARIMA(P,D,Q)_S (Seasonal Autoregressive Integrated Moving Average of order P,D,Q) โดย P เป็นอันดับของ AR (Seasonal Autoregressive) Q เป็นอันดับของ MA (Seasonal Moving Average) S เป็นช่วงเวลาของฤดูกาล (Seasonal period) และ D เป็นจำนวนครั้งในการหาผลต่างของฤดูกาลดังแสดงในสมการที่ (2.33)

$$\Phi_p(B^S)(1-B^S)^D \tilde{Z}_t = \Theta_q(B^S)a_t \quad (2.33)$$

เมื่อ

$$\Phi_p(B^S) = (1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_p B^{pS})$$

$$\Theta_q(B^S) = (1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_q B^{qS})$$

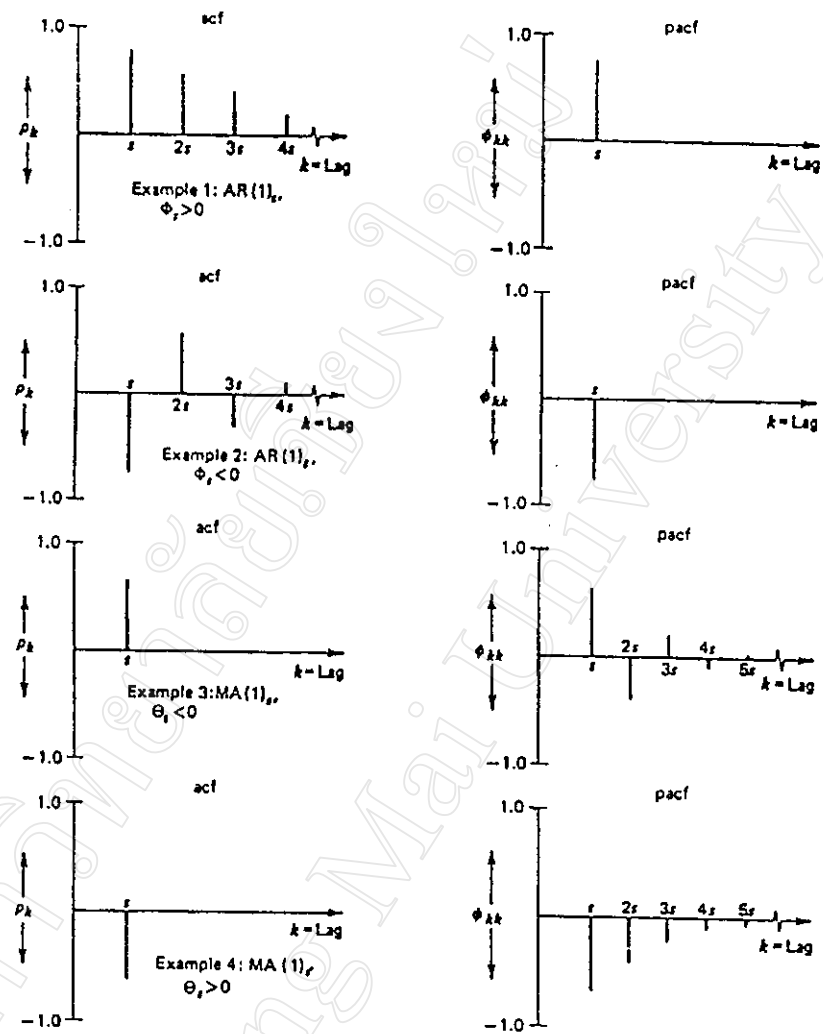
โดยทั่วไปข้อมูลอนุกรมเวลาอาจอยู่ในรูปของอนุกรมเวลาที่มีและไม่มีฤดูกาลก็ได้ ดังนั้น Box-Jenkins จึงได้เสนอรูปแบบทั่วไปของอนุกรมเวลาเป็น ARIMA(p,d,q) $\times$ (P,D,Q)_S ดังนี้

$$\phi_p(B)\Phi_p(B^S)(1-B)^d(1-B^S)^D \tilde{Z}_t = \theta_q(B)\Theta_q(B^S)a_t \quad (2.34)$$

โดยที่

$$\nabla_S^D \tilde{Z}_t = (1-B^S)^D \tilde{Z}_t$$

การคำนวณค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) และค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของอนุกรมเวลาที่มีฤดูกาลจะเหมือนกับอนุกรมเวลาที่ไม่มีฤดูกาล แต่การลดลงของฟังก์ชันสหสัมพันธ์ในตัวเองและฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วนของอนุกรมเวลาที่มีฤดูกาลจะลดลงเป็นช่วง $S, 2S, 3S, \dots$ และในการบ่งชี้รูปแบบของอนุกรมเวลาที่มีฤดูกาลก็จะมีช่วงห่างเป็นฤดูกาลเช่นเดียวกัน รูปที่ 2.8 แสดงลักษณะ ACF และ PACF ของรูปแบบอนุกรมเวลาที่มีฤดูกาล โดยที่รูปแบบ AR(1)_S ค่า ρ_k จะมีค่าลดลงแบบเอ็กซ์โปเนนเชียลที่ช่วง $S, 2S, 3S, \dots$ และมีเพียง $\phi_{kk} \neq 0$ ที่ช่วง $k=S$ สำหรับรูปแบบ MA(1)_S มีเพียงค่า $\rho_k \neq 0$ ที่ช่วง $k=S$ และ ϕ_{kk} จะมีค่าลดลงแบบเอ็กซ์โปเนนเชียลที่ช่วง $S, 2S, 3S, \dots$



รูปที่ 2.8 แสดงลักษณะ ACF และ PACF ของอนุกรมเวลาที่มีฤดูกาลรูปแบบ $AR(1)_s$ และ $MA(1)_s$

2.3 การวิเคราะห์การถดถอยกับอนุกรมเวลา

เนื่องจากการวิเคราะห์อนุกรมเวลาไม่สามารถอธิบายถึงความสัมพันธ์ของตัวแปรต่างๆ จึงใช้การวิเคราะห์การถดถอยเพื่อหาตัวแปรที่มีความสัมพันธ์กันมาอธิบายรูปแบบในการเดินทางและพยากรณ์ปริมาณการเดินทางของผู้โดยสารในอนาคตได้ ซึ่งมีขั้นตอนในการวิเคราะห์ ดังนี้

2.3.1 หาปัจจัยหรือตัวแปรที่มีอิทธิพลต่อการเดินทางของผู้โดยสารและสร้างแบบจำลองโดยวิธีการวิเคราะห์การถดถอย แสดงได้ดังสมการ

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j + e \quad (2.35)$$

เมื่อ

y คือ ตัวแปรตาม ซึ่งในงานวิจัยนี้จะเป็นค่าปริมาณผู้โดยสาร

x_i คือ ตัวแปรอิสระ ; $i = 1, 2, 3, \dots, j$

ในการวิเคราะห์ปริมาณการเดินทางของผู้โดยสารเครื่องบิน รถประจำทาง และรถไฟระหว่างเชียงใหม่และกรุงเทพฯ ตัวแปรอิสระที่มีอิทธิพลต่อการเดินทางนั้นอาจจะเป็นฟังก์ชันของราคาตั๋วโดยสาร ระยะเวลาในการเดินทาง ความถี่ในการบริการ จำนวนประชากร และรายได้ผลิตภัณฑ์จังหวัด เป็นต้น

2.3.2 พิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง ACF ของค่าคลาดเคลื่อนในสมการถดถอย โดยที่ $\hat{\rho}_k = 0$ เมื่อ $|\hat{\rho}_k| \leq 1.96SE(\hat{\rho}_k)$ และ $\hat{\rho}_k \neq 0$ เมื่อ $|\hat{\rho}_k| > 1.96SE(\hat{\rho}_k)$ เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง ACF ของค่าคลาดเคลื่อนมีค่าแตกต่างจากศูนย์ แสดงว่าอนุกรมเวลาของค่าคลาดเคลื่อนไม่เป็นสเตชันนารีจะต้องแปลงข้อมูลอนุกรมเวลาของค่าคลาดเคลื่อนให้เป็นสเตชันนารีเสียก่อน จึงจะหารูปแบบให้กับอนุกรมเวลาของค่าคลาดเคลื่อนได้ ดังนั้นแบบจำลองของอนุกรมเวลาสามารถแทนได้ดังสมการ (2.36) หรือ

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_n x_{tn} + e_t \quad (2.36)$$

โดยที่ค่า e_t จะได้จากการวิเคราะห์อนุกรมเวลาและการหารูปแบบที่เหมาะสม ซึ่งจะอยู่ในรูปแบบทั่วไปของอนุกรมเวลา $ARIMA(p,d,q)x(P,D,Q)_S$

2.3.3 จะได้แบบจำลองพฤติกรรมการเดินทางของผู้โดยสาร ในรูปแบบสมการถดถอยกับ
อนุกรมเวลา ซึ่งสามารถอธิบายตัวแปรหรือปัจจัยที่มีอิทธิพลต่อการเดินทางและสามารถพยากรณ์
ปริมาณการเดินทางของผู้โดยสารในอนาคตได้