บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

ในการศึกษาวิธีการตรวจสอบและปรับแก้ค่าสังเกตที่ผิดปกติในตัวแบบอนุกรมเวลาคงที่ในการวิจัยครั้งนี้ เป็นการศึกษาเปรียบเทียบสัดส่วนของความผิดพลาดทั้งหมดของการตรวจสอบ อำนาจการทดสอบ ร้อยละเฉลี่ยความแตกต่างระหว่างค่าสังเกตที่ผิดปกติ ค่าสังเกตที่ผิดปกติเมื่อมีการปรับแก้แล้ว ค่าสังเกตที่ปกติ และค่าร้อยละเฉลี่ยสัมบูรณ์ของการพยากรณ์อนุกรมเวลาล่วงหน้าเมื่อมีการปรับแก้ค่าสังเกตที่ผิดปกติแล้ว ของตัวสถิติทดสอบ 2 วิธี คือ วิธีการแบบ วี และวิธีการแบบ เอ็ม โดยศึกษา 2 กรณี คือ กรณีที่มีค่าสังเกตที่ผิดปกติ 1 และ 2 ค่า และกรณีที่มีค่าสังเกตที่ผิดปกติเป็น 5% 15% และ 25% ของข้อมูลทั้งหมด ซึ่งจะกล่าวในรายละเอียดต่อไปนี้

2.1 <u>อนุกรมเวลาคงที่</u>

2.1.1 รูปแบบอัตตสัมพันธ์อันดับที่ p (Autoregressive Model of Order p : AR(p)) ซึ่ง p คือ อันดับที่ของรูปแบบอัตสสัมพันธ์ มีรูปแบบเป็น

$$Z_t = \mu + \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \ldots + \Phi_p Z_{t-p} + a_t \qquad (2.1)$$

โดยที่ μ คือ ค่าคงที่

$\Phi_1, \ldots, \Phi_p$ คือ พารามิเตอร์ของอัตตสัมพันธ์ (Autoregressive)

รูปแบบที่นิยมใช้ได้แก่

2.1.1.1 รูปแบบ AR(1) มีรูปแบบเป็น

$$Z_t = \mu + \Phi_1 Z_{t-1} + a_t \quad , |\Phi_1| < 1 \qquad (2.2)$$

ซึ่ง $|\Phi_1| < 1$ เป็นเงื่อนไขที่ทำให้อนุกรมเวลามีคุณสมบัติคงที่

2.1.1.2. รูปแบบ AR(2) มีรูปแบบเป็น

$$Z_t = \beta + \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + a_t \qquad (2.3)$$

$$\Phi_1 + \Phi_2 < 1$$

$$\Phi_2 - \Phi_1 < 1$$

$$|\Phi_2| < 1$$

ซึ่ง $\Phi_1 + \Phi_2 < 1$,$\Phi_2 - \Phi_1 < 1$ และ $|\Phi_2| < 1$ เป็นเงื่อนไขที่ทำให้อนุกรมเวลามีคุณสมบัติคงที่

2.1.2 รูปแบบค่าเฉลี่ยเคลื่อนที่ที่มีอันดับ q (Moving Average Model of Order q : MA(q)) ซึ่ง q เป็นอันดับที่ของรูปแบบค่าเฉลี่ยเคลื่อนที่ มีรูปแบบ

$$Z_t = \mu + a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} - \dots - \Theta_q a_{t-q} \qquad (2.4)$$

โดยที่ Z_t คือ ค่าสังเกตของอนุกรมเวลา ณ คาบเวลา t

μ คือ ค่าคงที่

a_t คือ ค่าความคลาดเคลื่อน ณ คาบเวลา t มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น σ_a^2 และเป็นอิสระกัน

$\Theta_1,\dots,\Theta_q$ คือ พารามิเตอร์ของค่าเฉลี่ยเคลื่อนที่ (Moving Average)

2.1.2.1 รูปแบบ MA(1) มีรูปแบบเป็น

$$Z_t = \mu + a_t - \theta_1 a_{t-1} \quad , \quad |\theta_1| < 1 \qquad (2.5)$$

ซึ่ง $|\theta_1| < 1$ เป็นเงื่อนไขที่ทำให้อนุกรมเวลามีคุณสมบัติผกผันกลับ (inverstible)

2.1.2.2 รูปแบบ MA(2) มีรูปแบบเป็น

$$Z_t = \mu + a_t - a\theta_{1\,t-1} - a\theta_{2\,t-2} \qquad (2.6)$$

$$\theta_1 + \theta_2 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$|\theta_2| < 1$$

ซึ่ง $\theta_1 + \theta_2 < 1$, $\theta_2 - \theta_1 < 1$ และ $|\theta_2| < 1$ เป็นเงื่อนไขที่ทำให้อนุกรมเวลามีคุณสมบัติผกผันกลับ

2.1.3 รูปแบบผสมค่าเฉลี่ยเคลื่อนที่อัตตสหสัมพันธ์ที่มีอันดับ p และ เฉลี่ยเคลื่อนที่อันดับที่ q (Mixed Autoregressive-Moving-Average Model of Order p and q) : ARMA (p,q)) มีรูปแบบเป็น

$$Z_t = \mu + \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \dots + \Phi_p Z_{t-p}$$

$$- \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + a_t \qquad (2.7)$$

2.13.1 รูปแบบ ARMA (1,1) มีรูปแบบ

$$Z_t = \mu + \Phi_1 Z_{t-1} + a_t - \Theta_1 a_{t-1} \quad (2.8)$$

$$, \ |\Phi_1| < 1$$

$$|\Theta_1| < 1$$

ซึ่ง $|\Phi_1| < 1$ เป็นเงื่อนไขที่ทำให้อนุกรมเวลามีคุณสมบัติคงที่ $\Theta_1 < 1$ เป็นเงื่อนไขที่ทำให้อนุกรมเวลามีคุณสมบัติผกผันกลับ

2.2 <u>อนุกรมเวลาที่ไม่คงที่</u>

อนุกรมเวลาส่วนใหญ่มีคุณสมบัติเป็นอนุกรมเวลาที่ไม่คงที่ ดังนั้นการที่จะหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา จึงต้องแปลงอนุกรมเวลาที่มีคุณสมบัติเป็นอนุกรมเวลาไม่คงที่ให้เป็นอนุกรมเวลาที่มีคุณสมบัติคงที่ก่อน จึงจะทำการหารูปแบบให้กับอนุกรมเวลาได้ ซึ่งอาจจะทำได้โดยการหาผลต่าง (Differencing) ของอนุกรมเวลาเดิม ถ้าผลต่างครั้งที่ 1 ของอนุกรมเวลามีคุณสมบัติคงที่แล้ว ก็จะนำอนุกรมนั้นไปหารูปแบบที่เหมาะสมต่อไป ถ้าผลต่างครั้งที่ 1 ของอนุกรมเวลายังไม่มีคุณสมบัติคงที่ จึงจะหาผลต่างครั้งที่ 2 ของอนุกรมเวลา ถ้าผลต่างครั้งที่ 2 ของอนุกรมเวลามีคุณสมบัติคงที่แล้ว ก็จะนำอนุกรมเวลานี้ไปหารูปแบบที่เหมาะสมต่อไปโดย
ผลต่างครั้งที่ 1 จะได้

$$Z_t^{*} = DZ_t$$

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

ผลต่างครั้งที่ 2 จะได้

$$Z_t^{**} = D^2 Z_t$$

$$= Z_t - 2Z_{t-1} + Z_{t-2}$$

รูปแบบของอนุกรมเวลาที่มีคุณสมบัติเป็นอนุกรมเวลาไม่คงที่ จะทำให้อนุกรมเวลาคงที่ได้โดยการกำหนดอันดับของการหาผลต่าง (d > 1) รูปแบบจะเป็นหนึ่งเดียวระหว่างอัตตสัมพันธ์และเฉลี่ยเคลื่อนที่ (Autoregressive Integrate Moving Average : ARIMA (p,d,q) โดย d เป็นอันดับที่ของผลต่าง รูปแบบที่นิยมได้แก่

2.2.1 ARIMA (0,1,1) หรือ IMA (1,1) มีรูปแบบ

$$Z_t - Z_{t-1} = \mu + a_t - \theta_1 a_{t-1} \quad (2.9)$$

$$|\theta_1| < 1$$

2.2.2 ARIMA (0,1,2) หรือ IMA (1,2) มีรูปแบบ

$$Z_t - Z_{t-1} = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (2.10)$$

$$, \theta_1 + \theta_2 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$|\theta_2| < 1$$

2.2.3 ARIMA (1,1,0) หรือ ARI (1,1) มีรูปแบบ

$$(Z_t - Z_{t-1}) - \Phi_1 (Z_{t-1} - Z_{t-2}) = \mu + a_t \quad (2.11)$$

$$, |\Phi_1| > 1$$

2.2.4 ARIMA (2,1,0) หรือ ARI (2,1) มีรูปแบบ

$$(Z_t - Z_{t-1}) - \Phi_1 (Z_{t-1} - Z_{t-2}) - \Phi_2 (Z_{t-2} - Z_{t-3}) = \mu + a_t \quad (2.12)$$

$$\phi_1 + \phi_2 < 1$$

$$\phi_2 - \phi_1 < 1$$

$$|\phi_2| < 1$$

2.2.5 ARIMA (1,1,1) มีรูปแบบ

$$(Z_t - Z_{t-1}) - \phi_1(Z_{t-1} - Z_{t-2}) = \mu + a_t - \theta_1 a_{t-1} \tag{2.13}$$

$$|\phi_1| < 1$$

$$|\theta_1| < 1$$

2.2.6 ARIMA (0,1,0) มีรูปแบบ

$$Z_t - Z_{t-1} = a_t \tag{2.14}$$

ถ้าอนุกรมเวลาที่พิจารณามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง และอนุกรมเวลาที่มีทั้งแนวโน้มและอิทธิพลของฤดูกาล จะหารูปแบบได้จาก SARIMA (p,d,q) ซึ่งแสดงความสัมพันธ์ของอนุกรมเวลาที่อยู่ในช่วงเวลาที่ติดต่อกัน อาจเป็นเดือน เป็นปี ฯลฯ ถ้าสมมติว่าอนุกรมเวลามีคุณสมบัติเป็น SARIMA (0,1,1) และอนุกรมเวลามีความสัมพันธ์ระหว่างช่วงห่าง 12 เดือน จะมีรูปแบบ คือ

$$Z_t - Z_{t-12} = a_t - \theta^* a_{t-12} \tag{2.15}$$

$$, \ |\theta^*| < 1$$

โดย $Z_t - Z_{t-12}$ คือ ค่าสังเกตที่อยู่ห่างกัน 12 เดือน

θ^* คือ พารามิเตอร์ในรูปแบบค่าเฉลี่ยที่ฤดูกาล (Seasonal Moving Average Model)

การกำหนดรูปแบบ การกำหนดรูปแบบของอนุกรมเวลา จะวิเคราะห์จากค่าสังเกตในอดีตซึ่งควรจะมีจำนวนค่าสังเกตอย่างน้อย 30 ค่า หรือบางครั้งอาจถึง 100 ค่า ค่าสถิติที่สำคัญที่ใช้ในการพยากรณ์ ได้แก่ ฟังก์ชั่นสหสัมพันธ์ระหว่างค่าสังเกตที่อยู่ห่างกัน k ช่วงเวลา คือ ρ_k (Autocorrelation Function at lag k) และฟังชั่นสหสัมพันธ์บางส่วนระหว่างค่าสังเกตที่อยู่ห่างกัน k ช่วงเวลา คือ ϕ_{kk} (Partial Autocoorelation Function at lag k)

โดย ρ_k มีคุณสมบัติคือ $-1 < \rho_k < 1$ และ $\rho_k = \rho_{-k}$ ซึ่ง

$$\rho_k = \Sigma(Z_t - \text{Zbar})(Z_{t+k} - \text{Zbar})/C_o \qquad (2.16)$$

$$, t = 1,\ldots,n-k$$

โดยที่

$$C_o = \Sigma(Z_t - \text{Zbar})^2 \qquad , t = 1,\ldots,n$$

$$\text{Zbar} = \Sigma Z_t /n \qquad , t = 1,\ldots,n$$

ส่วน Φ_{kk} เป็นค่าวัดสหสัมพันธ์บางส่วนระหว่างค่าสังเกตที่อยู่ห่างกัน k ช่วงเวลา เมื่อกำหนดให้อิทธิพลเนื่องจากตัวแปรอื่นคงที่ ซึ่ง

$$\Phi_{kk} = \begin{cases} \rho_1 & \text{สำหรับ } k = 1 \\ (\rho_k - R)/(1 - R) & \text{สำหรับ } k = 2,3,\ldots \end{cases} \qquad (2.17)$$

โดยที่ $R = \Sigma\rho_{k-1}\, \rho_{k-j}$ สำหรับ $j = 1,2,\ldots,k-1$

เมื่อ $\rho_{k-j} = \rho_{k-1,j} - \Phi_{kk}\,\rho_{k-1,k-j}$ สำหรับ $j = 1,2,\ldots,k-1$

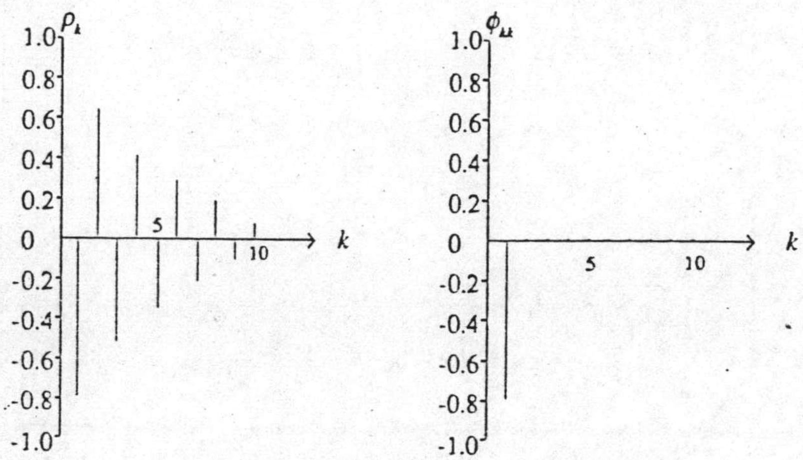
การพิจารณารูปแบบจะได้จากการสร้างกราฟความสัมพันธ์ (Correlogram) โดยนำ ρ_k และ Φ_{kk} ไปพล็อตกับ k และนำไปเปรียบเทียบกับรูปแบบมาตรฐานที่แสดงความสัมพันธ์ของ k

กับ ρ_k และ k กับ ϕ_{kk} ซึ่ง $\hat{\rho}_k$ เป็นตัวประมาณของ ρ_k และ $\hat{\phi}_{kk}$ เป็นตัวประมาณของ ϕ_{kk} สำหรับรูปแบบแต่ละรูปแบบ จะทำให้สามารถเลือกรูปแบบได้อย่างเหมาะสม ซึ่งลักษณะของ ρ_k และ ϕ_{kk} สำหรับรูปแบบ ARMA ต่าง ๆ แสดงในตารางที่ 2.1 และสามารถแสดงกราฟความสัมพันธ์ของ k กับ ρ_k และ k กับ ϕ_{kk} สำหรับรูปแบบ ARMA ต่าง ๆ ดังรูปที่ 2.1

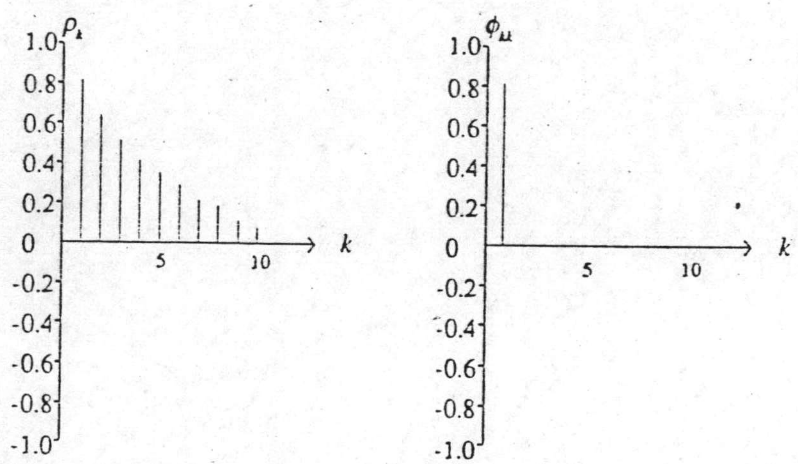
ตารางที่ 2.1 ลักษณะของ ρ_k และ ϕ_{kk} สำหรับรูปแบบของ ARMA ต่าง ๆ

รูปแบบ	ρ_k	ϕ_{kk}
AR (1)	ρ_k มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น	$\phi_{kk} = \begin{cases} 1, & \text{สำหรับ } k = 1 \\ 0, & \text{สำหรับ } k = 2,3,\ldots \end{cases}$
AR (2)	ρ_k มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น	$\phi_{kk} = \begin{cases} 1, & \text{สำหรับ } k = 1,2 \\ 0, & \text{สำหรับ } k = 3,4,\ldots \end{cases}$
MA (1)	$\rho_k = \begin{cases} 1, & \text{สำหรับ } k = 1 \\ 0, & \text{สำหรับ } k = 2,3,\ldots \end{cases}$	ϕ_{kk} มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น
MA(2)	$\rho_k = \begin{cases} 1, & \text{สำหรับ } k = 1,2 \\ 0, & \text{สำหรับ } k = 3,4,\ldots \end{cases}$	ϕ_{kk} มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น
ARMA(0,0)	$\rho_k = 0$, สำหรับ $k = 1,2,..$	$\phi_{kk} = 0$, สำหรับ $k = 1,2,\ldots$
ARMA(1,1)	ρ_k มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น	ϕ_{kk} มีค่าเข้าใกล้ 0 เมื่อ k มีค่ามากขึ้น

รูปแบบ AR(1)



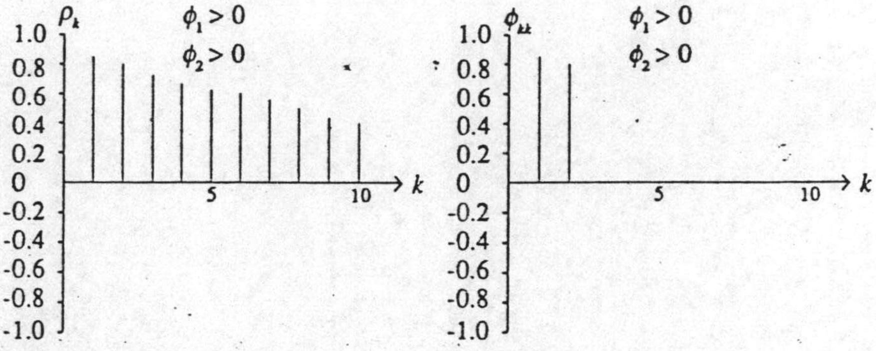
หรือ



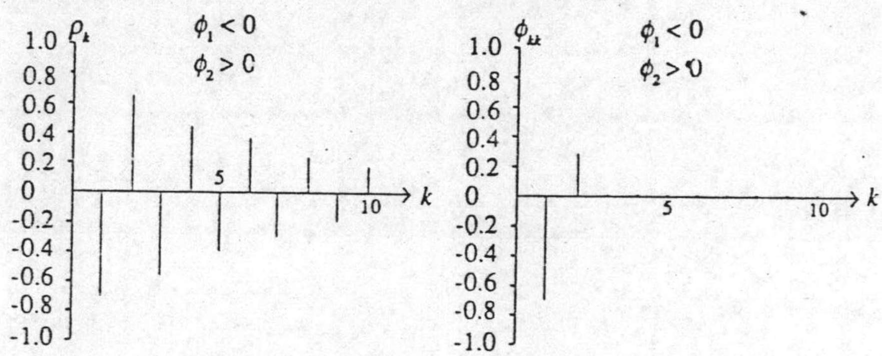
ภาพที่ 2.1 กราฟแสดงความสัมพันธ์ของ k กับ ρ_k และ k กับ ϕ_{kk} สำหรับ

รูปแบบของ ARMA ต่าง ๆ

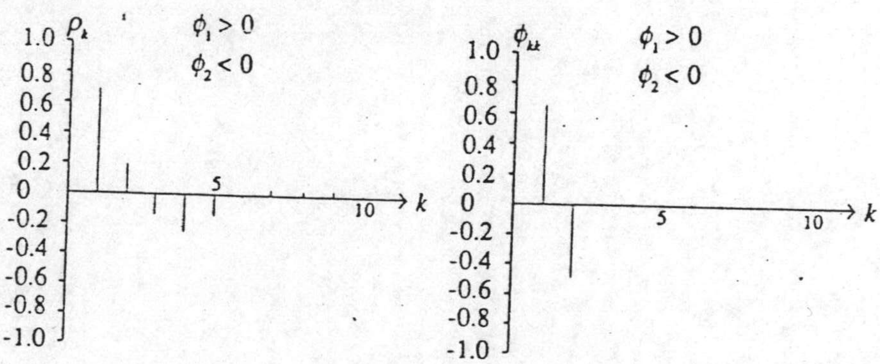
รูปแบบ AR (2)



หรือ

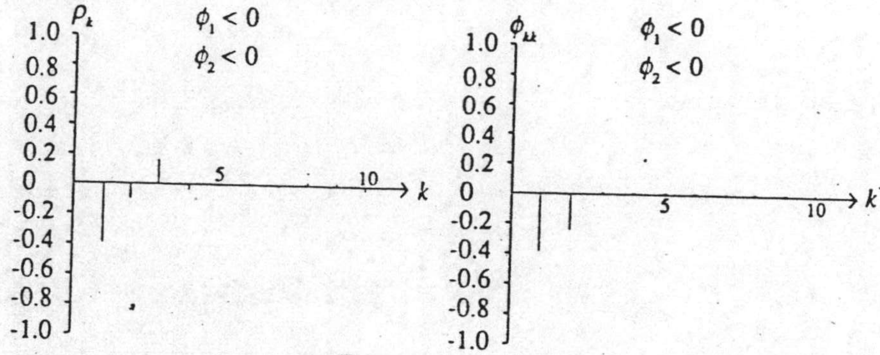


หรือ

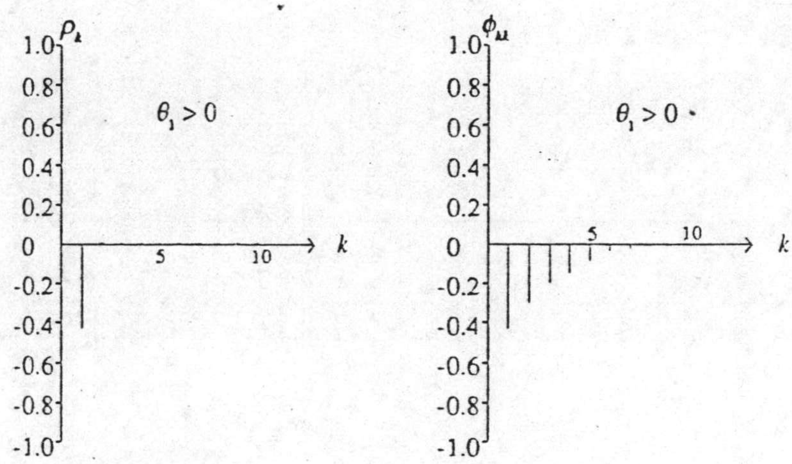


ภาพที่ 2.1 (ต่อ)

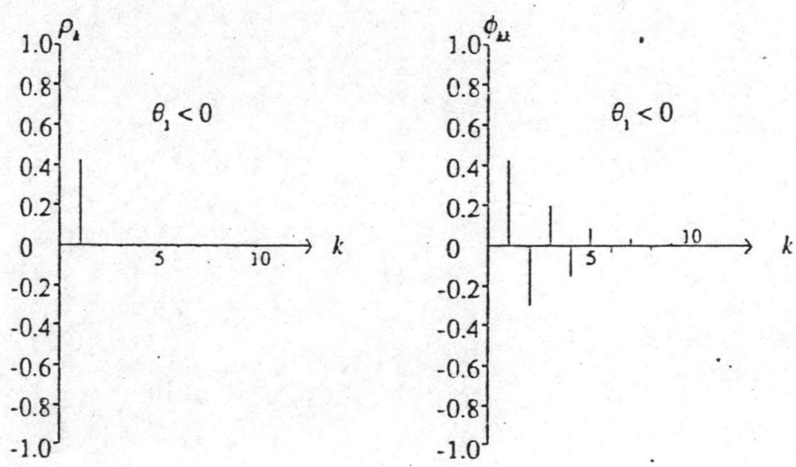
หรือ



รูปแบบ MA (1)

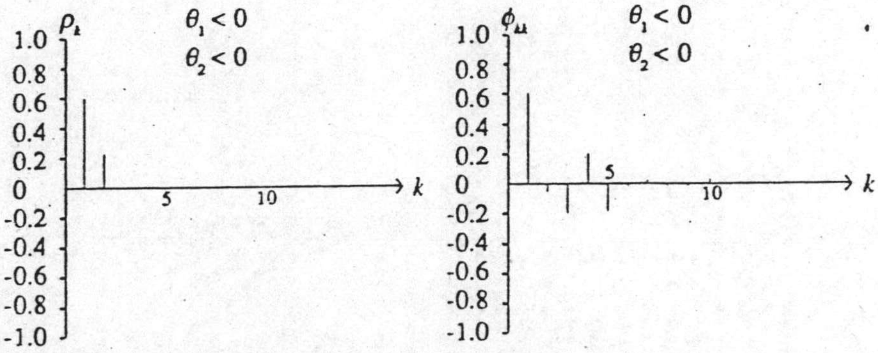


หรือ

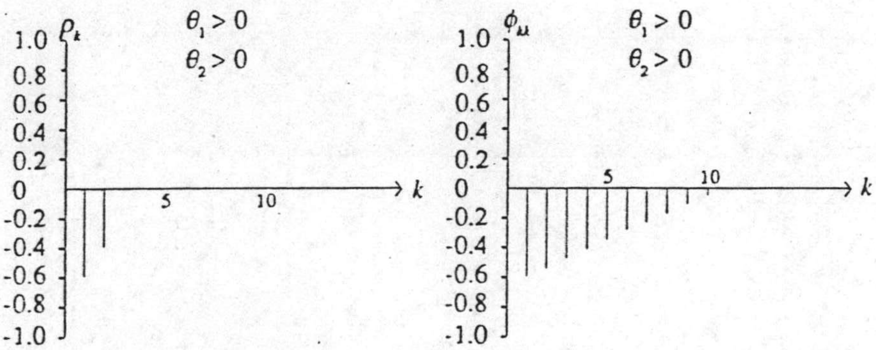


รูปที่ 2.1 (ต่อ)

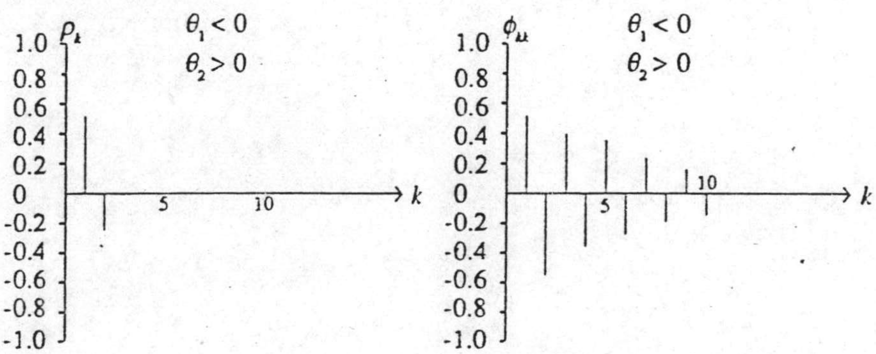
รูปแบบ MA (2)



หรือ

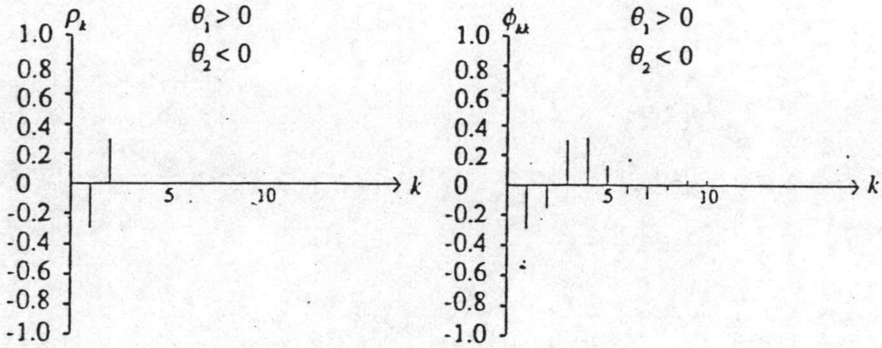


หรือ

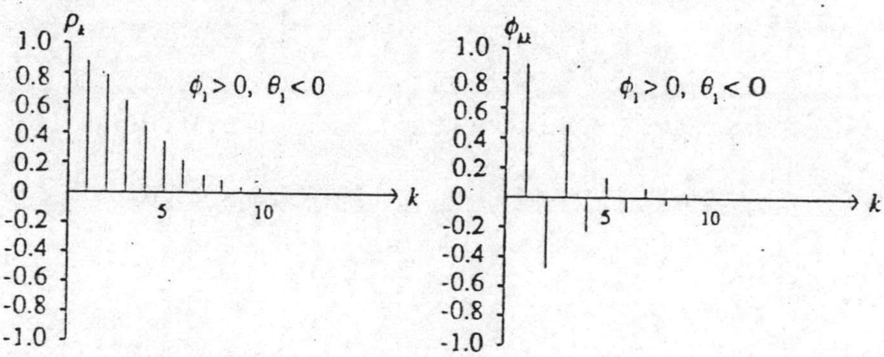


รูปที่ 2.1 (ต่อ)

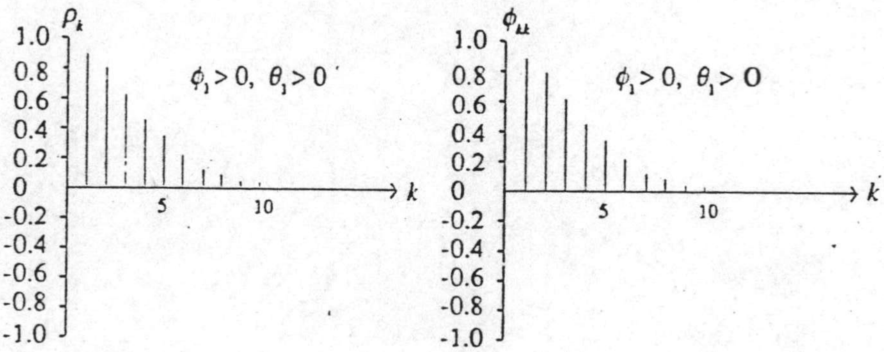
หรือ



รูปแบบ ARMA (1,1)

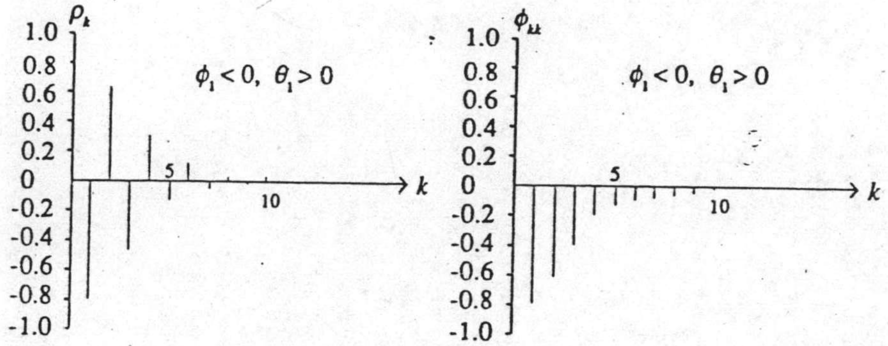


หรือ

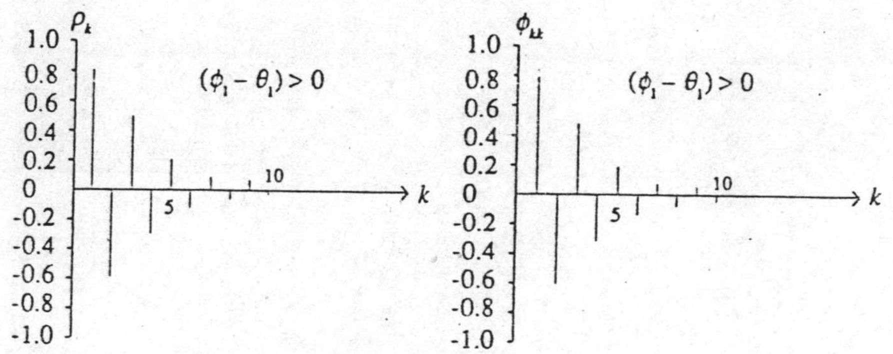


ภาพที่ 2.1 (ต่อ)

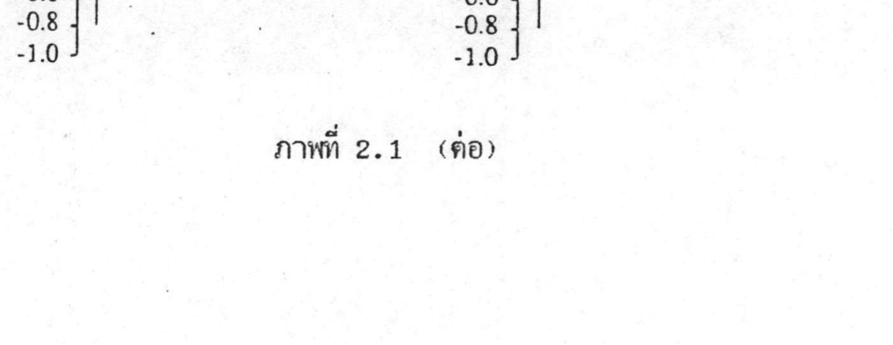
หรือ



หรือ

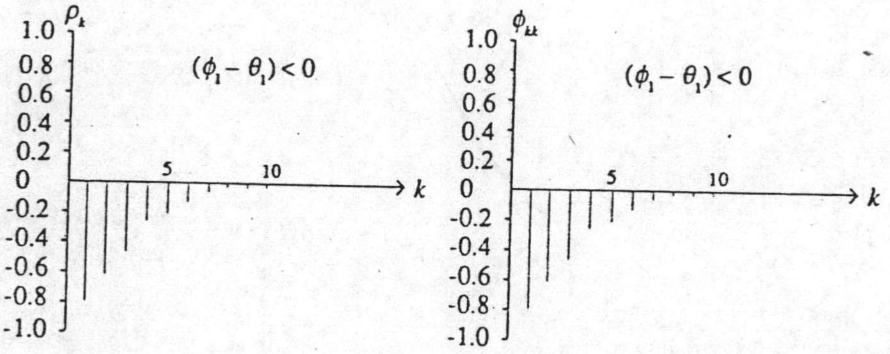


หรือ



ภาพที่ 2.1 (ต่อ)

หรือ



ภาพที่ 2.1 (ต่อ)

การพิจารณาว่า ρ_k มีค่าเท่ากับ 0 หรือ ϕ_{kk} จะมีค่าเท่ากับ 0 หรือไม่ มีจุดวิกฤติคือ 2σ ซึ่ง σ เป็นส่วนเบี่ยงเบนมาตรฐานของ ρ_k หรือ ϕ_{kk} หมายความว่า ถ้า ρ_k หรือ ϕ_{kk} มีค่าอยู่ในช่วงหรือเท่ากับ 2σ จะยอมรับว่า ρ_k หรือ ϕ_{kk} มีค่าเท่ากับ 0 ถ้า ρ_k หรือ ϕ_{kk} มีค่าอยู่นอกช่วง 2σ จะไม่ยอมรับว่า ρ_k หรือ ϕ_{kk} มีค่าเท่ากับ 0 โดยค่าประมาณความแปรปรวนของ ρ_k คือ $Var(\rho_k)$ มีค่าประมาณเท่ากับ

$$(1/n) \cdot \{1 + 2\rho_1^2 + \ldots + 2\rho_q^2\} \qquad \text{สำหรับ } k > q \quad .$$

และเมื่อ n มีขนาดใหญ่จะประมาณว่า $Var(\rho_k)$ มีค่าประมาณเท่ากับ n^{-1} ส่วนค่าประมาณความแปรปรวนของ ϕ_{kk} คือ $Var(\phi_{kk})$ มีค่าประมาณเท่ากับ n^{-1} สำหรับ $k > q$

การประมาณค่าของพารามิเตอร์ การประมาณค่าพารามิเตอร์ในรูปของ ARIMA เบื้องต้น จะใช้ความสัมพันธ์ของฟังก์ชั่นสหสัมพันธ์กับพารามิเตอร์ ซึ่งความสัมพันธ์ สำหรับอนุกรมเวลาแต่ละรูปแบบ ดังตารางที่ 2.2

<u>ตารางที่ 2.2</u> แสดงความสัมพันธ์ของ ρ_k กับพารามิเตอร์ในรูปแบบต่าง ๆ

รูปแบบ	พารามิเตอร์	ความสัมพันธ์	ขอบเขตของพารามิเตอร์
AR (1)	ϕ_1	$\rho_1 = \phi_1$	$-1 < \phi < 1$
AR (2)	ϕ_1 , ϕ_2	$\rho_1 = \phi_1 + \phi_2 \rho_1$ $\rho_2 = \phi_1 \rho_1 + \phi_2$	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $-1 < \phi_2 < 1$
MA (1)	θ_1	$\rho_1 = -\theta_1 /(1 + \theta_1^2)$	$-1 < \theta < 1$
MA (2)	θ_1 , θ_2	$\rho_1 = \dfrac{-\theta_1 (1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}$ $\rho_2 = \dfrac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$	$\theta_1 + \theta_2 < 1$ $\theta_2 - \theta_1 < 1$ $-1 < \theta_2 < 1$
ARMA(1,1)	ϕ_1 , θ_1	$\rho_1 = \dfrac{(1 - \theta_1 \phi_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1 \theta_1}$ $\rho_2 = \rho_1 \phi_1$	$1 < \phi_1 < 1$ $-1 < \theta_1 < 1$

จากความสัมพันธ์ดังกล่าว จะประมาณค่าพารามิเตอร์นี้ได้โดยการแทนค่า ρ_k ด้วย ρ_k แล้วแก้สมการหาค่าประมาณของพารามิเตอร์

การตรวจสอบรูปแบบ ในการตรวจสอบรูปแบบเป็นการตรวจสอบว่ารูปแบบที่กำหนดไว้เหมาะสมกับอนุกรมเวลาหรือไม่ ซึ่ง Box และ Pierce (Box และ Pierce 1970) ได้เสนอวิธีการตรวจสอบความเหมาะสมของรูปแบบโดยใช้ตัวสถิติ บ๊อกซ์เพียซไคสแควร์ (Box-Pierce Chi-Square) คือ Q เป็นการตรวจสอบว่าความคลาดเคลื่อน ณ เวลา t คือ a_t , $t = 1,2,...,n$ มีความเป็นอิสระต่อกันหรือไม่ โดยเปรียบเทียบผลรวมของค่าสหสัมพันธ์ของ a_t ณ เวลาต่าง ๆ กับช่วงความเชื่อมั่น

$$Q = (n-d)\ \Sigma\rho_j^2(a_t) \qquad (2.18)$$

เมื่อ $j = 1,2,...,k$

โดย n คือ จำนวนค่าสังเกตในอนุกรมเวลา

d คือ ลำดับผลต่างของอนุกรมเวลาที่ทำให้อนุกรมเวลาเป็นอนุกรมเวลาคงที่

$\rho_j(a_t)$ คือ ค่าสหสัมพันธ์ระหว่างค่าสังเกตที่อยู่ห่างกัน j ช่วงเวลาของอนุกรมเวลาของค่าความคลาดเคลื่อน

Q มีการแจกแจงแบบไคสแควร์โดยประมาณ มีองศาอิสระ (degree of freedom) เท่ากับ $k - n_p$ โดย n_p เป็นจำนวนพารามิเตอร์ที่ประมาณได้ ถ้า Q น้อยกว่าหรือเท่ากับ $\chi^2_\alpha(k - n_p)$ จะได้ว่า a_t เป็นอิสระกัน แสดงว่ารูปแบบที่ใช้เหมาะสมแล้ว ถ้า Q มากกว่า $\chi^2_\alpha(k - n_p)$ จะได้ว่า a_t ไม่เป็นอิสระกัน แสดงว่ารูปแบบที่ใช้ยังไม่เหมาะสมจึงจะกลับไปพิจารณาหารูปแบบที่เหมาะสมใหม่ต่อไป

2.3 วิธีการตรวจหาและปรับแก้ค่าสังเกตที่ผิดปกติ

2.3.1 วิธีการแบบวี เป็นวิธีการที่เสนอโดย Hsu (1977)

ให้อนุกรมเวลา Z_t และ Y_t เป็นอนุกรมเวลาที่มีค่าสังเกตที่ผิดปกติอย่าง

อิสระ (Outlier free) กำหนดให้ว่า $\{Y_t\}$ เป็นอนุกรมเวลาที่มีตัวแบบเป็น

$$\Phi(B)y_t = \Theta(B)a_t \tag{2.19}$$

เมื่อ $\Phi(B)$ เป็นฟังชันพารามิเตอร์ของขบวนการตัวแบบ AR(p)

$\Theta(B)$ เป็นฟังชันพารามิเตอร์ของขบวนการตัวแบบ MA(q)

B เป็นตัวดำเนินการย้อนเวลา (back-shiff operator)

ซึ่ง $\Phi(B) = (1 - \Phi_1 B - \ldots - \Phi_p B^p)$

$\Theta(B) = (1 - \Theta_1 B - \ldots - \Theta_q B^q)$

และ $\{a_t\}$ เป็นความคลาดเคลื่อนที่มีการแจกแจงปกติมีค่าเฉลี่ยเท่ากับ 0 มีความแปรปรวนเท่ากับ σ_a^2

ดังนั้น

$$Z_t = \begin{cases} Y_t & \text{เมื่อ } t \neq T \\ Y_t + w & \text{เมื่อ } t = T \end{cases} \tag{2.20}$$

$$= Y_t + w\, I_t^{(T)} \tag{2.21}$$

$$= (\Theta(B)/\Phi(B))\, a_t + w\, I_t^{(T)}$$

w เป็นพารามิเตอร์สำหรับปรับค่าสังเกตที่ผิดปกติในอนุกรมเวลา

T เป็นตำแหน่งคาบเวลาที่มีค่าสังเกตที่ผิดปกติเกิดขึ้น

$$Z_t = [\Theta(B)/\Phi(B)]\, a_t + w\, I_t^{(T)} \tag{2.22}$$

เมื่อ

$$I_t^{(T)} = \begin{cases} 1 & \text{เมื่อ } t = T \\ 0 & \text{เมื่อ } t \neq T \end{cases}$$

ขบวนการที่ตรวจหาค่าสังเกตที่ผิดปกติ จะพิจารณาจาก T และ พารามิเตอร์ทุกตัวในสมการ (2.19) และกำหนดให้ว่า

$$b_t = [\Phi(B)/\theta(B)]\ Z_t \tag{2.23}$$

ดังนั้น

$$b_t = \begin{cases} a_t & \text{เมื่อ } t \neq T \\ a_t[1 + wv] & \text{เมื่อ } t = T \end{cases}$$

และอัตราส่วนของความแปรปรวนของ b_t คือ

$$r_T = [(T - 1)\ wb_t]/[(n - T + 1)\ ub_t] \tag{2.24}$$

เมื่อ $wb_t = \Sigma b_t^2 \quad ; t = T,T+1,\ldots,n$

$ub_t = \Sigma b_t^2 \quad ; t = 1,2,\ldots,T-1$

ซึ่ง $(T-1) > 0$ และ $(n-T+1) > 0$ และอัตราส่วน r_T จะเป็นตัวประมาณของ $(1 + wv)^2$

การทดสอบสมมติฐาน

Ho : อนุกรมเวลา Z_t ไม่มีค่าผิดปกติ

H1 : อนุกรมเวลา Z_t มีค่าผิดปกติ

ตัวสถิติที่ใช้ในการทดสอบสมมติฐาน H_1 คู่กับ H_o คือ

$$r_T = [(T - 1)\ wb_t]/[(n - T + 1)\ ub_t] \tag{2.25}$$

ซึ่ง r_T จะมีการแจกแจงแบบ F (F-distribution) ด้วยองศาแห่งความเป็นอิสระ (n-T+1) และ (T-1)

2.3.2 วิธีการแบบเอ็ม เป็นวิธีการที่เสนอโดย Cheng Tiao และ Chen (1988)

ให้อนุกรมเวลา Z_t และ Y_t เป็นอนุกรมเวลาที่มีค่าสังเกตที่ผิดปกติอย่างอิสระ กำหนดให้ว่า Y_t เป็นอนุกรมเวลาที่มีตัวแบบเช่นเดียวกับสมการ (2.19) (2.20) (2.21) (2.22) ขบวนการตรวจหาค่าสังเกตที่ผิดปกติ จะพิจารณาจาก T และพารามิเตอร์ทุกตัวในสมการ (2.19)

$$\pi(B) = [\Phi(B)/\Theta(B)] = (1 - \pi_1 B - \pi_2 B^2 - \dots) \qquad (2.26)$$

$\pi(B)$ เป็นฟังชันอัตราส่วนพารามิเตอร์ของขบวนการในตัวแบบ ARMA(p,q) และกำหนดให้ว่า

$$e_t = \pi(B) Z_t \qquad (2.27)$$

จากสมการที่ (2.20) และ (2.21) ได้ว่า

$$e_t = w\pi(B) I_t^{(T)} + a_t \qquad (2.28)$$

สำหรับค่าสังเกต n ค่า และตัวแบบในสมการที่ (2.28) เขียนเป็นเมตริกซ์ได้ดังนี้

$$\begin{bmatrix} e_1 \\ . \\ e_{T-1} \\ e_T \\ e_{T+1} \\ . \\ e_n \end{bmatrix} = w \begin{bmatrix} 0 \\ . \\ 0 \\ 1 \\ -\pi_1 \\ . \\ -\pi_{n-T} \end{bmatrix} + \begin{bmatrix} a_1 \\ . \\ a_{T-1} \\ a_T \\ a_{T+1} \\ . \\ a_n \end{bmatrix} \tag{2.29}$$

ให้ W_{AT} เป็นตัวประมาณค่าแบบลีสสแควร์ (least square : LS) ของ w สำหรับตัวแบบ เพราะว่า $\{a_t\}$ เป็นความคลาดเคลื่อนจากทฤษฎีลีสสแควร์ ดังนั้นจะได้ว่า

$$W_{AT} = [e_t - \Sigma\pi_j e_{T+j}]/\Upsilon^2 \tag{2.30}$$

เมื่อ $j=1,2,...,n-T$

$$\Upsilon^2 = \Sigma\pi_j^2$$ เมื่อ $j=0,1,...,n-T$

และ $\pi_0 = 1$

$$\mathrm{Var}(W_{AT}) = \sigma_a^2/\Upsilon^2 \tag{2.31}$$

ซึ่ง $\mathrm{Var}(W_{AT})$ น้อยกว่าหรือเท่ากับ σ_a^2

การทดสอบสมมติฐาน

Ho : อนุกรมเวลา Z_t ไม่มีค่าผิดปกติ

H1 : อนุกรมเวลา Z_t มีค่าผิดปกติ

ตัวสถิติที่ใช้ในการทดสอบสมมติฐาน H_1 คู่กับ H_o คือ

$$\lambda_T = \Upsilon\, W_{AT} / \sigma_a \tag{2.32}$$

ซึ่ง λ_T จะมีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็น 0 ความแปรปรวนเป็น 1

2.4 ขั้นตอนการตรวจหาและปรับแก้ค่าสังเกตที่ผิดปกติ

2.4.1 วิธีการแบบวี

2.4.1.1 กำหนดให้อนุกรมเวลา Z_t ไม่มีภาวะการค่าสังเกตที่ผิดปกติทำการคำณวนค่าพารามิเตอร์ ความคลาดเคลื่อนของอนุกรมเวลาและกำหนดค่าระดับนัยสำคัญ C สำหรับการทดสอบสมมติฐาน ดังสมการ (2.23)

$$\hat{b}_t = [\hat{\Phi}(B)/\hat{\theta}(B)]\, Z_t$$

เมื่อ

$$\hat{\Phi}(B) = (1 - \hat{\Phi}_1 B - \ldots - \hat{\Phi}_p B^p)$$

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

2.4.1.2 คำนวณค่าอัตราส่วนความแปรปรวนโดยใช้ค่าความคลาดเคลื่อนที่ได้ในหัวข้อ 2.4.1.1 เพื่อประมาณค่า $\hat{r}_T$ ดังสมการ (2.24)

$$\hat{r}_T = [(T - 1)\, w\hat{b}_t]/[(n - T + 1)\, u\hat{b}_t]$$

เมื่อ

$$w\hat{b}_t = \Sigma\hat{b}_t^2 \qquad ; t = T, T+1, \ldots, n$$

$$u\hat{b}_t = \Sigma\hat{b}_t^2 \qquad ; t = 1, 2, \ldots, T-1$$

ซึ่ง $(T-1) > 0$ และ $(n-T+1) > 0$ ใช้ประมาณตัวแบบ

$$\hat{\lambda}_V = \max_T \{\hat{r}_T\} \qquad (2.33)$$

เมื่อ T แสดงถึงคาบเวลาที่เป็นไปได้สูงสุด ที่คาดว่าจะเกิดค่าสังเกตที่ผิดปกติขึ้นในอนุกรม

2.4.1.3 เปรียบเทียบ $\hat{\lambda}_V$ กับค่าวิกฤติ C ถ้า $\hat{\lambda}_V$ มากกว่าหรือเท่ากับค่า C แสดงว่าในขณะนั้นเกิดค่าสังเกตที่ผิดปกติ ณ คาบเวลา T จึงทำการปรับปรุงข้อมูลโดยสมการ

$$Z_t = \text{Zbar} - (Z_t - \text{Zbar})(\hat{\lambda}_V)^{-1/2} \qquad (2.34)$$

เมื่อ Zbar เป็นค่าเฉลี่ยของข้อมูล Z_t

2.4.1.4 ทำย้อนกลับไปที่ข้อ 2.4.1.1 ถึง 2.4.1.3 ใหม่ทุก ๆ ภาวะการที่ตรวจพบค่าสังเกตที่ผิดปกติ จนกว่าอนุกรมเวลาที่กำลังพิจารณานั้นไม่พบค่าสังเกตที่ผิดปกติ

2.4.2 วิธีการแบบเอ็ม

2.4.2.1 กำหนดให้อนุกรมเวลา Z_t ไม่มีภาวะการค่าสังเกตที่ผิดปกติทำการคำนวนค่าพารามิเตอร์ ความคลาดเคลื่อนของอนุกรมเวลาและกำหนดค่าระดับนัยสำคัญ C สำหรับการทดสอบสมมติฐาน ดังสมการ (2.27)

$$\begin{aligned} \hat{e}_t &= \hat{\pi}(B)*Z_t \\ &= [\hat{\Phi}(B)/\hat{\Theta}(B)]\, Z_t \end{aligned}$$

$$t = 1,2,\dots,n$$

เมื่อ $\hat{\pi}(B) = (1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \dots)$

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

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

โดยที่ ความแปรปรวน $(\hat{\sigma}_a^2) = (\Sigma\hat{e}_t^2)/n$ เพื่อใช้เป็นค่าเริ่มต้นของการประมาณ σ_a^2

2.4.2.2 คำณวนค่า $\hat{\lambda}_T$ สำรับ $t = 1,2,\ldots,n$

$$\hat{\lambda}_t = \hat{\tau}\,\hat{W}_{AT}/\hat{\sigma}_a \qquad (2.35)$$

ใช้ประมาณตัวแบบ

$$\hat{\lambda}_T = \max_T \{|\hat{\lambda}_t|\}$$

เมื่อ T แสดงถึงคาบเวลาที่เป็นไปได้สูงสุด ที่คาดว่าจะเกิดค่าผิดปกติขึ้นในอนุกรมเวลา ถ้า $\hat{\lambda}_T \geq |\hat{\lambda}_t| > C$ แสดงว่าในขณะนั้นเกิดค่าผิดปกติ ณ เวลา T จึงทำการประมาณพารามิเตอร์ W_{AT} ดังสมการ (2.36)

$$\hat{W}_{AT} = [\hat{e}_t - \Sigma\,\hat{\pi}_j\,\hat{e}_{T+j}]/\hat{\tau}^2 \qquad (2.36)$$

เมื่อ $j = 1,2,\ldots,n-T$

$$\hat{\tau}^2 = \Sigma\,\hat{\pi}_j^2$$ เมื่อ $j=0,1,\ldots,n-T$

และ $\hat{\pi}_0 = 1$

$$\mathrm{Var}(\hat{W}_{AT}) = \hat{\sigma}_a^2/\hat{\tau}^2$$

$$I_t(T) = \begin{cases} 1 & \text{เมื่อ } t = T \\ 0 & \text{เมื่อ } t \neq T \end{cases}$$

และให้ความหมายความคลาดเคลื่อนตัวใหม่

$$\tilde{e}_t = \hat{e}_t - \hat{W}_{AT}\,\hat{\pi}(B)\,I_t^{(T)} \qquad (2.37)$$

ทำการประมาณความค่า σ_a^2 ตัวใหม่ จากการปรับปรุงความคลาดเคลื่อนแล้วทำการปรับปรุงข้อมูลโดยสมการ

$$\tilde{Z}_t = Z_t - \hat{W}_{AT} I_t^{(T)} \qquad (2.38)$$

2.4.2.3 ทำการคำนวนค่า $\hat{\lambda}_t$ ใหม่โดยยึดหลักการปรับปรุงความคลาดเคลื่อน และ σ_a^2 ตามที่กล่าวมาแล้ว ย้อนกลับไปที่ข้อ 2.4.2.2 ใหม่ทุก ๆ ภาวะการที่ตรวจพบค่าสังเกตที่ผิดปกติ

2.4.2.4 ทำในข้อที่ 2.4.2.1 ถึง 2.4.2.3 ซ้ำ ๆ ทั้งหมดจนกว่าอนุกรมเวลาที่กำลังพิจารณานั้นไม่พบค่าสังเกตที่ผิดปกติ

2.4.3 วิธีการประมาณค่า $\hat{\pi}(B)$

2.4.3.1 เมื่ออนุกรมเวลามีตัวแบบเป็น AR(1)

$$\hat{\pi}(B) = \hat{\Phi}(B)/\hat{\Theta}(B) \qquad (2.39)$$

$$= (1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \ldots)$$

$$(1 - \hat{\Phi}_1 B) = 1 - \hat{\pi}_1 B$$

ดังนั้น

$$\hat{\pi}_1 = \hat{\Phi}_1$$

2.4.3.2 เมื่ออนุกรมเวลามีตัวแบบเป็น MA(1)

$$1/(1 - \hat{\Theta}_1) = (1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \ldots) \qquad (2.40)$$

$$1 = (1 - \hat{\Theta}_1 B)(1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \ldots)$$

$$1 = 1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \ldots$$

$$- \hat{\Theta}_1 B + \hat{\pi}_1 \Theta_1 B^2 - \ldots$$

$$\ldots$$

เทียบค่าของ B : $-\hat{\pi}_1 - \hat{\theta}_1 = 0$

$$-\hat{\pi}_1 = \hat{\theta}_1$$

ดังนั้น $$\hat{\pi}_1 = -\hat{\theta}_1$$

2.4.3.3 เมื่ออนุกรมเวลามีตัวแบบเป็น ARMA(1,1)

$$(1 - \hat{\phi}_1 B)/(1 - \hat{\theta}_1 B) = (1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \ldots) \qquad (2.41)$$

$$(1 - \hat{\phi}_1 B) = (1 - \hat{\theta}_1 B)(1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \ldots)$$

จัดรูปแบบใหม่

$$(1 - \hat{\phi}_1 B) = (1 - (\hat{\pi}_1 + \hat{\theta}_1)B - (\hat{\pi}_2 - \hat{\pi}_1)B^2 - \ldots)$$

เทียบสัมประสิทธิ์

$$\hat{\phi}_1 = \hat{\pi}_1 + \hat{\theta}_1$$

$$\hat{\pi}_1 = \hat{\phi}_1 - \hat{\theta}_1$$

2.4.4 เครื่องมือที่ใช้วัดความถูกต้องการพยากรณ์ล่วงหน้าของอนุกรมเวลาคือ ค่าร้อยละเฉลี่ยสัมบูรณ์ (mean absolute percentage error : MAPE)

$$MAPE = \Sigma[|\hat{e}_t / Z_t|]/n \ (100) \qquad \text{เมื่อ } t = 1,2,\ldots,n$$