บทที่ 3

การวิเคราะห์อนุกรมเวลา (Time series Analysis)

3.1 การวิเคราะห์แนวโน้ม

ข้อมูลที่เก็บบันทึกตามเวลา (chronological data) ซึ่งถือเป็นตัวแปรเชิงสุ่ม (Random variable) ที่สุ่มในเวลาที่แบ่งไว้เท่า ๆ กัน อันเป็นลักษณะหนึ่งของอนุกรมเวลา ข้อมูลอนุกรมเวลานี้สามารถนำมาวิเคราะห์แนวโน้ม (Long term trend) ของตัวแปรสุ่มว่าในระยะยาวจะมีลักษณะใดคือมีแนวโน้มที่จะเพิ่มขึ้นหรือลดลงในอัตราใด การวิเคราะห์อนุกรมเวลาตามระเบียบวิธีสถิติกระทำโดยพยายามหาสมการคณิตศาสตร์ (Mathematical model) อธิบายข้อมูลอนุกรมเวลาที่มีอยู่ เพื่ออธิบายแนวโน้มด้วยการแสดงความสัมพันธ์ระหว่างตัวแปรเชิงสุ่มกับเวลาวัตถุประสงค์ที่สำคัญในการใช้สมการคณิตศาสตร์คือ ศึกษาความเปลี่ยนแปลง (Deviation) ที่ต่างไปจากแนวโน้ม, เป็นภาพจำลองเหตุการณ์ในอดีตและใช้ในการประมาณค่าตัวแปรในเวลาที่ต้องการตลอดจนการพยากรณ์ในอนาคต สมการคณิตศาสตร์ที่สำคัญที่มักนำมาใช้กับการสร้างให้กับแนวโน้มระยะยาวคือ

3.1.1 สมการเส้นตรง (Linear regression equation) ที่มีรูปสมการเป็น $Z_t = a + bt$

Z_t เป็น dependent variable ที่ขึ้นกับ t

t เป็น independent variable ในที่นี้หมายถึงช่วงเวลา

a,b เป็นพารามิเตอร์ที่ต้องประมาณค่าจากข้อมูล โดย a หมายถึง intercept ของเส้นตรง ส่วน b หมายถึง slope ของเส้นตรง

วิธีที่ดีที่สุดในการหาพารามิเตอร์ คือ Least square method ซึ่งหา a,b จากการแก้สมการ Normal equations

$$\Sigma Z = Na + b\Sigma t$$

$$\Sigma t Z = a\Sigma t \quad b\Sigma t^2$$

ถ้านำสมการเส้นตรงมาใช้ในหน่วยของ logarithm จะมีสมการเป็น

$$\log Z_t = \log a + t\log b$$

สมการ Normal equations ที่ใช้หาค่า loga, logb ซึ่งเป็นพารามิเตอร์ของสมการจะเปลี่ยนเป็น

$$\Sigma \log Z = N\log a + \log b \Sigma t$$

$$\Sigma t\log Z = \log a \Sigma t + \log b \Sigma t^2$$

๓.๑.๒ สมการ Nonlinear เป็นสมการเส้นโค้งต่างๆ เช่น สมการ Polynomials ที่ dependent variable ขึ้นกับ independent variable ที่มีกำลังมากกว่า ๑, Modified-exponential curve, Logistic curve

จะเห็นได้ว่า มีสมการคณิตศาสตร์หลายชนิดที่สามารถนำมาแสดงความสัมพันธ์ระหว่างตัวแปรเชิงสุ่มกับเวลา เพื่อนำมาวิเคราะห์แนวโน้มระยะยาวของข้อมูลอนุกรมเวลา

การเลือกชนิดของสมการคณิตศาสตร์นั้นขึ้นอยู่กับวัตถุประสงค์ในการนำมาใช้ ถ้านำมาใช้แสดงภาพจำลองของเหตุการณ์ในอดีตและวิเคราะห์ความเบี่ยงเบน (deviation) ที่เบนไปจากแนวโน้ม หรือเส้นที่แสดงแนวโน้มของข้อมูล หรือ การสมการคณิตศาสตร์ของเส้น trend ควรเป็นรูปง่ายพอควรและแสดงการประมาณที่ทำให้นำความแตกต่างมาวิเคราะห์ได้ชัดเจน

อย่างไรก็ดี วิธีตรวจสอบเบื้องต้นอาจทำได้โดยการพิจารณา Scatter diagram และค่าความแตกต่าง (difference)

เช่น สร้างสมการเส้นตรง $Z = a+bt$ ถ้าหากว่าค่าความแตกต่างที่ ๑ (first difference) หรือค่าแตกต่างระหว่างค่าตัวแปรในปีปัจจุบันกับปีที่แล้วมีค่าคงที่ หรือ scatter diagram บนกระดาษกราฟที่มีหน่วยแบบเลขคณิต (Arithmetic paper) มีลักษณะแบบเส้นตรง[1]

สำหรับสมการเส้นตรงในหน่วย log ที่มีรูปเป็น $\log Z_t = \log a + t\log b$ ค่าแตกต่างที่ ๑ ของตัวแปรมีแนวโน้มที่จะลดลงหรือเพิ่มขึ้นในเปอร์เซนต์ที่คงที่

1Croxton,and others,Applied General Statistics(New Delhi 1969) p 282-283

หรือค่าแตกต่างที่ ๑ของค่า log ของตัวแปรคงที่ ถ้าดูจาก scatter diagram บน Semi-logarithmic paper จะมีลักษณะแนวโน้มเป็นเส้นตรง[1]

๓.๒ การวิเคราะห์ส่วนเบี่ยงเบน

เมื่อหาสมการคณิตศาสตร์ที่เป็นตัวแทนแนวโน้มหรือเป็นภาพจำลองในอดีตแล้ว มักจะวิเคราะห์ส่วนที่แตกต่างไปจากแนวโน้ม การสร้างแบบจำลองหรือสมการคณิตศาสตร์สำหรับค่าแตกต่างนั้นจะอาศัยทฤษฎีของ stationary time series ซึ่งกล่าวว่า "ในกรณีที่ข้อมูลแสดงแนวโน้มในระยะยาวที่สามารถอธิบายได้ด้วยความสัมพันธ์ทางคณิตศาสตร์แบบง่ายๆแล้ว ส่วนที่เบนไปจากแนวโน้มให้ถือว่ามีprocessเป็น Stationary "[2]แบบจำลองที่ได้จึงเป็นLinear Stationary Modelซึ่งเป็นแบบจำลองประเภทหนึ่งของ Stochastic Model หรือ Stochastic process

ดังนั้น Stationary Model จึงเป็น Model ที่สร้างให้กับอนุกรมส่วนที่แตกต่างหรือเบี่ยงเบนไปจากระดับหรือค่าคงที่ค่าหนึ่งของอนุกรมเวลานั้นหรือเป็นการสร้างแบบจำลองที่ต้องการวิเคราะห์เกี่ยวกับระดับของอนุกรมเวลา

Linear Stationary Model ที่มักนำมาวิเคราะห์ในทางปฏิบัติมากที่สุดคือ

1. AUTOREGRESSIVE MODEL
2. MOVING AVERAGE MODEL
3. MIXED AUTOREGRESSIVE-MOVING AVERAGE MODEL

แบบจำลองทั้ง๓นี้ล้วนแต่แสดงค่าแต่ละเทอมในอนุกรมในรูปของค่าอนุกรมในหน่วยเวลาที่ล่วงมาแล้วกับshockที่เป็นอิสระ ซึ่งโดยปกติค่าสังเกตแต่ละค่าในอนุกรมเวลามักไม่เป็นอิสระ ค่าสังเกตในอดีตจึงควรแสดงความสัมพันธ์กับค่าสังเกตในปัจจุบัน ความไม่เป็นอิสระนี้เองที่ขัดกับทฤษฎีรีเกรสชั่นจึงต้องแปลงความสัมพันธ์นั้นให้อยู่ในรูปของ shocks ที่เป็นอิสระซึ่งเป็นอนุกรมเวลาของตัวแปรที่อิสระนั่นเอง

๑. AUTOREGRESSIVE MODEL เป็นแบบจำลองที่ใช้ในทางปฏิบัติมากที่สุด เพราะแบบจำลองนี้ค่าของอนุกรมในปัจจุบันจะแสดงความสัมพันธ์กับค่าของอนุกรม

1Croxton and others, Applied General Statistics, (New Delhi, 1969) p282-283

2P.Holgate, "Time Series Analysis Applied to Wildfowl Counts", Journal of the Royal Statistical Society XV No1(1966),p15

ในอดีตในรูปของ

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \qquad (๑)$$

โดย $\tilde{Z}_t$ เป็นค่าเบี่ยงเบนจากแนวโน้มในเวลาที่ t

a_t คือshock ในเวลาที่ t หรือค่าผิดพลาด (error term)

p คือจำนวนเทอมของอนุกรม Z ในอดีตที่นำมาใช้ในการสร้างสมการ

$\phi_1, \phi_2, \dots \phi_p$ เป็นพารามิเตอร์ที่ต้องประมาณจากข้อมูล

วิธีหนึ่งของการประมาณพารามิเตอร์นี้คือ วิธี Yule-Walker estimate ซึ่งจะกล่าวโดยละเอียดต่อไปในหัวข้อ การประมาณค่าพารามิเตอร์ของสมการ AUTOREGRESSIVE และโดยทั่วไปแล้วเรียกสมการ (๑)ว่า AUTOREGRESSIVE ที่มีORDER p ให้สัญญลักษณ์เป็น AR(p)

๒. MOVING AVERAGE MODEL เป็นแบบจำลองที่แสดงค่าอนุกรมเวลาของตัวแปร Z_t หรือส่วนเบี่ยงเบนในรูปของค่าผิดพลาดหรือ random shock a_t ในอดีต q ค่า

รูปของMOVING AVERAGE ORDERq หรือ MA(q) คือ

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

โดย $\theta_1, \theta_2, \dots \theta_q$ เป็นพารามิเตอร์ที่ต้องประมาณจากข้อมูล

๓. MIXED AUTOREGRESSIVE-MOVING AVERAGE MODEL เป็นแบบจำลองที่สร้างให้กับข้อมูลที่มีลักษณะผสมระหว่าง AUTOREGRESSIVE กับ MOVING AVERAGE MODEL จึงมีรูปสมการเป็น

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

ดังนั้น $\phi_1, \phi_2, \dots \phi_p, \theta_1, \theta_2, \dots \theta_q$ ต้องประมาณจากข้อมูล แบบจำลองนี้เรียกสั้นๆว่า ARMA(p,q)

ในทางปฏิบัติแล้ว Stationary time seriesที่ได้จาก AUTOREGRESSIVE MODEL หรือMOVING AVERAGE MODEL หรือMIXED AUTOREGRESSIVE-MOVING AVERAGE MODELมักจะมี ORDERp,q ไม่เกิน ๒

ดังได้กล่าวแล้วว่า จุดประสงค์ที่สำคัญในการหาแบบจำลองของข้อมูลที่มีอยู่ คือ เพื่อใช้แบบจำลองอธิบายความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรไม่อิสระ

ในรูปที่เป็นสูตรที่สามารถคำนวนเป็นตัวเลขได้ แบบจำลองที่ดีจึงต้องเป็นตัวแทนของข้อมูล และมีความใกล้เคียงความจริง เพื่อใช้ประมาณค่าและพยากรณ์ในเวลาที่ต้องการ การเลือกแบบจำลองในการวิเคราะห์อนุกรมเวลาจึงเป็นสิ่งสำคัญ ด้วยเหตุนี้การทดสอบข้อมูลก่อนใช้แบบจำลองจริงเพื่อช่วยให้เห็นลักษณะของแบบจำลองที่เหมาะสมจึงเป็นสิ่งจำเป็นอย่างยิ่งในการเลือกแบบจำลอง

สำหรับการสร้าง Linear Stationary Model นั้นมีวิธีการทดสอบเบื้องต้นที่เป็นแนวทางในการเลือกแบบจำลองเป็นขั้น ๆ ดังนี้

1. ทดสอบความเป็นอิสระระหว่างค่าสังเกตในอนุกรมเวลากับค่าข้างเคียง หรือค่าสังเกตที่เกิดขึ้นต่างเวลากันโดยใช้สัมประสิทธิ์ของสหสัมพันธ์แบบ serial (Serial correlation coefficient) r_k^* เป็นค่าสถิติที่นำมาทดสอบความเป็นอิสระอย่างมีนัยสำคัญ โดยเทียบค่า r_k^* ที่คำนวนได้กับค่าในตารางสหสัมพันธ์แบบ serial ของ R.L. Anderson Table ด้วย degree of freedom N-k

เมื่อ N = จำนวนค่าสังเกตทั้งหมดที่นำมาสร้างสมการ

k = จำนวนเวลาที่ต่างกันหรือจำนวนเวลาที่ล้าหลัง

ถ้าค่า r_k^* ที่คำนวนได้เกินกว่าค่าใน R.L.Anderson Table แสดงว่าปฏิเสธข้อสมมติที่ว่า อนุกรมเวลามีความเป็นอิสระ หรืออนุกรมเวลานั้นไม่เป็น random series[1] นั่นคือ แต่ละเทอมในอนุกรมเวลามีความไม่เป็นอิสระซึ่งกันและกัน ค่าของอนุกรมเวลาที่ล่วงมาแล้วสามารถนำมาใช้ประมาณค่าอนุกรมเวลาในปัจจุบันได้

ค่า r_k^* ดังกล่าวจะหาได้จากสูตร

$$r_k^* = \frac{\sum_{t=1}^{N-k} z_t z_{t+k} - \left(\sum_{t=1}^{N-k} z_t\right)\left(\sum_{t=k+1}^{N} z_t\right)/(N-k)}{\sqrt{\left[\sum_{t=1}^{N-k} z_t^2 - \left(\sum_{t=1}^{N-k} z_t\right)^2/(N-k)\right]\left[\sum_{t=k+1}^{N} z_t^2 - \left(\sum_{t=k+1}^{N} z_t\right)^2/(N-k)\right]}}$$

โดย k เป็นจำนวนเวลาที่ต่างกันหรือจำนวนเวลาที่ล้าหลัง r_k^* จะมีค่าระหว่าง -1 กับ 1 ถ้า r_k^* มีค่าใกล้ 1 แสดงว่าองศาความไม่เป็นอิสระของเทอมต่าง ๆ ในอนุกรมเวลามีมากและแปรผันในทางบวก ตรงกันข้ามถ้า r_k^* มีค่าใกล้ -1 แสดงว่า

1 Yamane, Statistics; An Introductory Analysis, (2nd ed.1970), p867

องค่าความไม่เป็นอิสระของเทอมต่าง ๆ ในอนุกรมเวลามีมากแต่เป็นไปในทางตรงกันข้ามหรือกลับกัน

๔. ดูลักษณะความสัมพันธ์ของอนุกรมเวลาที่เกิดขึ้นต่างเวลาว่าเป็นไปในแบบเส้นตรงพอที่จะสร้างด้วย Linear Stationary model หรือไม่ การดูความสัมพันธ์ดังกล่าวอาศัยจาก scatter diagram ของ joint probability distribution ของ Z_t กับ Z_{t+k} โดยนำค่าแต่ละคู่ที่เกิดขึ้นต่างเวลากัน k หน่วยเวลามาเขียนลงในแผนภาพ

เมื่อพล็อตแล้วแสดงการแจกแจงร่วมในแนวเส้นตรงลงมา แสดงว่าในช่วงเวลาที่ห่างกัน k หน่วยเวลา ค่าที่อยู่ใกล้กัน k หน่วยเวลามีความสัมพันธ์แบบเส้นตรงที่แปรในทางกลับกัน ดังรูปที่ 3.1.1 เมื่อ k= 1 Z_t กับ Z_{t+k} มีความสัมพันธ์แบบเส้นตรงเป็นลบ ถ้าเส้นตรงชี้ขึ้นแสดงว่า ค่าข้างเคียง k หน่วยเวลานั้นมีความสัมพันธ์แบบเส้นตรงแต่แปรในทางบวก ดังรูปที่ 3.1.2 แสดงความสัมพันธ์แบบเส้นตรงเป็นบวก เมื่อ k = 2

๓. ดูลักษณะของ AUTOREGRESSIVE MODEL, MOVING AVERAGE MODEL และ MIXED AUTOREGRESSIVE-MOVING AVERAGE MODEL จากแผนภาพ autocorrelation function ที่แสดงความสัมพันธ์ระหว่าง lag k กับ autocorrelation coefficient

ค่าของ autocorrelation coefficient จะประมาณจากสูตร

$$r_k = \frac{c_k}{c_0} = \hat{\rho}_k$$

โดย

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

ให้ r_k เป็นค่าประมาณของ autocorrelation coefficient ที่ lagk หรือ $\hat{\rho}_k$

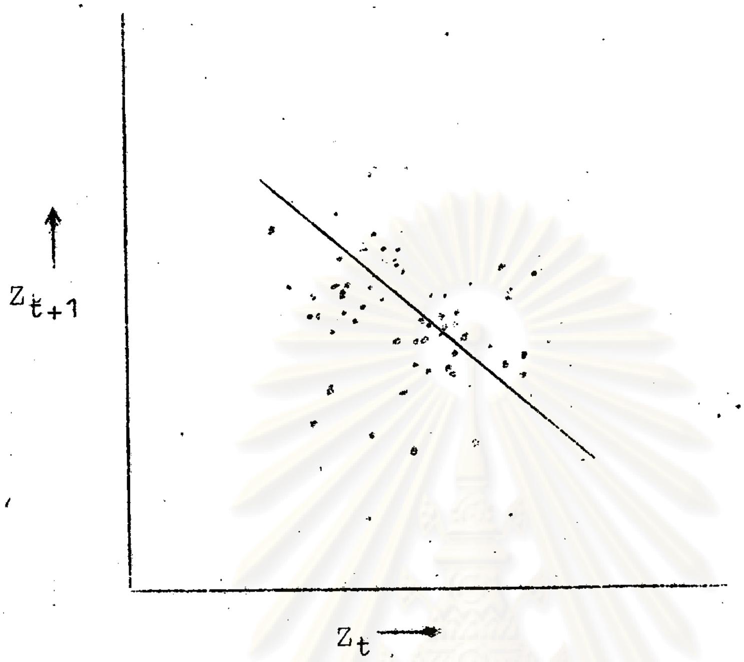
c_k เป็นค่าประมาณของ autocovariance หรือ $\hat{\gamma}_k$

$\bar{Z}$ เป็น mean ของอนุกรมเวลาของค่าสังเกตทั้งหมด

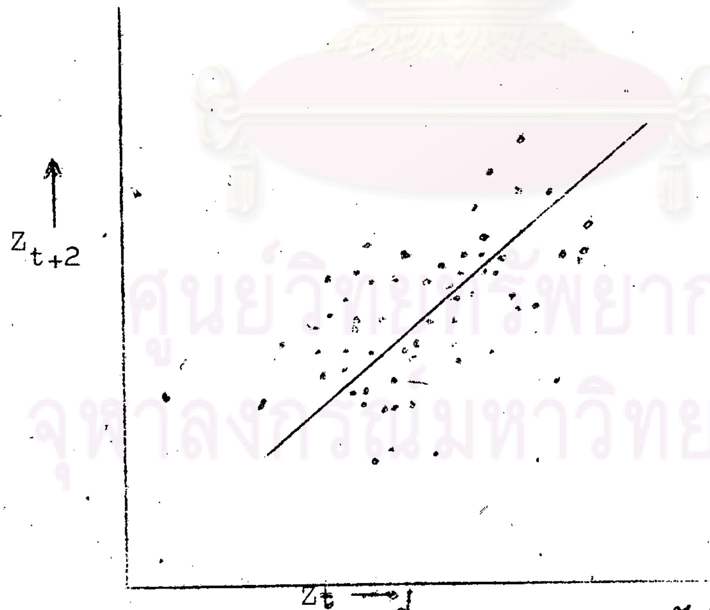
K เป็นค่าที่มากที่สุดของ k ในทางปฏิบัติค่า k เป็น 20 ก็เพียงพอที่จะเห็นลักษณะ autocorrelation function ที่จะเป็นแนวทางในการสร้างสมการ[1]

1Box&Jenkins, Time Series Analysis forecasting and control, ("Holden-day),p 175

ภาพที่ ๓.๑



ภาพที่ ๓.๑.๑ scatter diagram ที่ lag 1 แสดงความสัมพันธ์ระหว่าง Z_t กับ Z_{t+1} เป็นลบ



ภาพที่ ๓.๑.๒ scatter diagram ที่lag 2 แสดงความสัมพันธ์ระหว่าง Z_t กับ Z_{t+2} เป็นบวก

ค่าของ r_k มีค่าไม่เกิน ๑ เสมอ

(แผนภาพ) ถ้ากราฟ autocorrelation function มีลักษณะ damped exponentials หรือ damped sine wave หรือทั้ง damped exponentials และ damed sine wave รวมกัน แสดงถึงความไม่สิ้นสุด (Infinity) ของ autocorrelation คือไม่ว่า lag k มีค่ามากเท่าใด ค่าของ r_k ยังมีค่าไม่ใกล้ ๐ หรือ tails off คือค่า r_k สูงมากเมื่อ k มีค่าน้อยและค่าลดลงอย่างช้าๆเกือบถึง ๐ เมื่อค่า k มีค่าเพิ่มขึ้นมากๆ แบบจำลองที่เหมาะกับข้อมูลที่มีลักษณะ autocorrelation ดังกล่าวคือ AUTOREGRESSIVE MODEL แต่ถ้า autocorrelation function แสดงลักษณะตรงข้ามกับลักษณะแรกคือมีลักษณะ finite หรือ cuts off ซึ่งค่า r_k ใน lag k ต้นๆจะมีค่าใกล้ ๐ นั่นคือ r_k มีค่าลดลงอย่างรวดเร็ว MOVING AVERAGE MODEL จะเหมาะกับข้อมูลที่มีลักษณะดังกล่าวนี้

สำหรับ autocorrelation function ของ MIXED AUTOREGRESSIVE-MOVING AVERAGE MODEL นั้นมีลักษณะคล้าย autocorrelation function ของ AUTOREGRESSIVE MODEL คือ infinite หรือ tails off ที่มีลักษณะ damped exponentials หรือ damped sine wave หรือเป็นทั้ง damped exponentials กับ damped sine wave หลัง q-p lags แล้วเป็นต้นไป

๔. ดูลักษณะของ partial autocorrelation function เพื่อช่วยในการหา order ของแบบจำลองทั้ง ๓ คือ หา p ถ้าลักษณะ r_k เป็นแบบ AR หา q ถ้าลักษณะ r_k เป็นแบบ MA และหาทั้ง p และ q ถ้าลักษณะ r_k เป็น ARMA

ลักษณะของ partial autocorrelation function จะดูได้จากแผนภาพแสดงความสัมพันธ์ระหว่าง lag kk กับ partial autocorrelation function ϕ_{kk} ซึ่ง ϕ_{kk} นี้ประมาณจากสมการ

$$r_j = \hat{\phi}_{k1} r_{j-1} + \dots + \hat{\phi}_{k(k-1)} r_{j-k+1} + \hat{\phi}_{kk} r_{j-k} , \quad j=1,2,\dots k$$

หรือ
$$\begin{bmatrix} 1 & r_1 & r_2 & \cdots\cdots r_{k-1} \\ r_1 & 1 & r_1 & \cdots\cdots r_{k-2} \\ \vdots & \vdots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \cdots\cdots 1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \vdots \\ \hat{\phi}_{kk} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}$$

หรือจากสูตร
$$\hat{\phi}_{kk} = \begin{cases} r_1 & ,k=1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j} & ,k=2,3,\ldots,K \end{cases}$$

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

ลักษณะของ partial autocorrelation function ของแบบจำลองแต่ละชนิด มักมีลักษณะตรงข้ามกับลักษณะของ autocorrelation function ของแบบจำลองชนิดนั้นยกเว้นแบบจำลอง MIXED AUTOREGRESSIVE-MOVING AVERAGE เช่น แบบจำลอง AUTOREGRESSIVE ซึ่งมีลักษณะ autocorrelation function แบบ infinite หรือ damped exponentials หมายถึงค่า autocorrelation function จะลดลงอย่างช้า ๆ จนเกือบถึง 0 เมื่อ k มีค่าสูงมาก ๆ แต่ partial autocorrelation function ของแบบจำลอง JTOREGRESSIVE จะมีค่าลดลงอย่างรวดเร็วใน lag k น้อย ๆ เรียกว่ามีลักษณะ cuts off หรือ finite และมีค่าเป็น 0 หรือเกือบถึง 0 หลักจาก lag ใด lag หนึ่ง ถ้าหลัง lag p ไปค่า partial autocorrelation function เป็น 0 แบบจำลอง AUTOREGRESSIVE นั้งควรมี order เป็น p

สรุปได้ว่า ถ้าลักษณะ partial autocorrelation function ของแบบจำลอง AUTOREGRESSIVE ไม่เป็น 0 สำหรับ k ที่มีค่าน้อยกว่าหรือเท่ากับ p นอกนั้นเป็น 0

คือ $\hat{\phi}_{kk} \neq 0 \quad , k \leq p$

และ $\hat{\phi}_{kk} = 0 \quad , k > p$

แบบจำลอง AUTOREGRESSIVE นั้นมี order เป็น p
หรือ AR(p) model
ในกรณีที่ partial autocorrelation function มีลักษณะเป็น infinite แบบ damped exponentialsหรือ damped sine wave หรือเป็นทั้ง damped exponentialsกับ damped sine wave รวมกัน อันเป็นลักษณะ tails off ที่ตรงกันข้ามกับลักษณะ autocorrelation functionของ MOVING AVERAGE MODEL ซึ่งเป็น finiteหลังจาก lag q แล้วมีค่าเป็น ๐ q จึงควรเป็น orderของ MOVING AVERAGE MODEL

ส่วนลักษณะของ partial autocorrelation functionของแบบจำลอง MIXED AUTOCORRELATION-MOVING AVERAGE นั้นจะมีลักษณะคล้ายลักษณะ autocorrelation function ของ model นี้อีก คือเป็น infinite แบบdamped exponentials หรือ damped sine wave หรือเป็นทั้ง damped exponentials กับdamped sine waveรวมกัน หลังp-q lags ไป คือยังคงเป็น tails off

ในการดูลักษณะของ autocorrelation function และ partial autocorrelation functionเพื่อช่วยในการเลือก model ค่า k ที่สูงที่สุด คือ 20 หรือ K= 20

ภาพที่๓.๒ แสดงลักษณะdamped exponentialsต่างๆ

ดังกล่าวว่าในทางปฏิบัติค่า p,q มีค่าไม่เกิน ๒ ดังนั้นจึงควรสนใจลักษณะและคุณสมบัติของ AUTOREGRESSIVE MODEL, MOVING AVERAGE MODEL และMIXED AUTOREGRESSIVE MODEL ที่มี ORDER ไม่เกิน ๒ โดยศึกษาจาก autocorrelation function, partial autocorrelation function และเงื่อนไขที่ทำให้เป็น stationary model แบบจำลองที่น่าสนใจศึกษาคือ

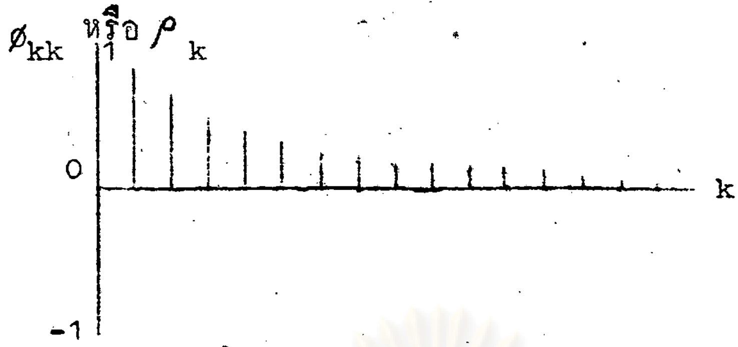
๑. AUTOREGRESSIVE MODEL ที่มี ORDER เป็น 1

๒. AUTOREGRESSIVE MODEL ที่มี ORDER เป็น 2

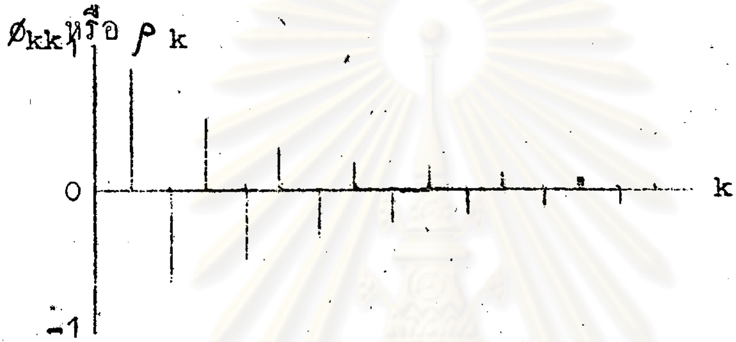
๓. MOVING AVERAGE MODEL ORDER 1

๔. MOVING AVERAGE MODEL ORDER 2

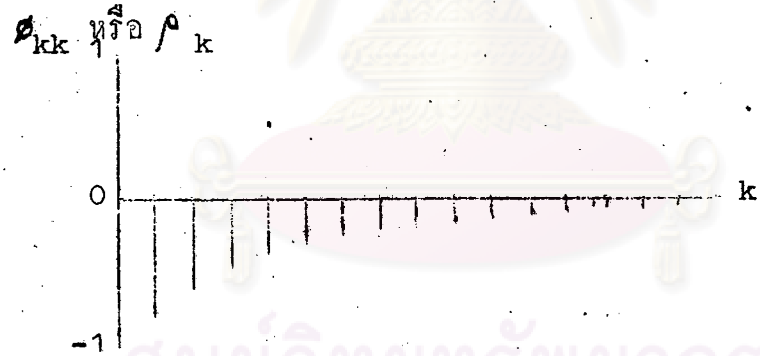
ภาพที่ ๓.๒ ลักษณะที่สำคัญของ autocorrelation function ρ_k และ partial autocorrelation ~~function~~ ϕ_{kk}



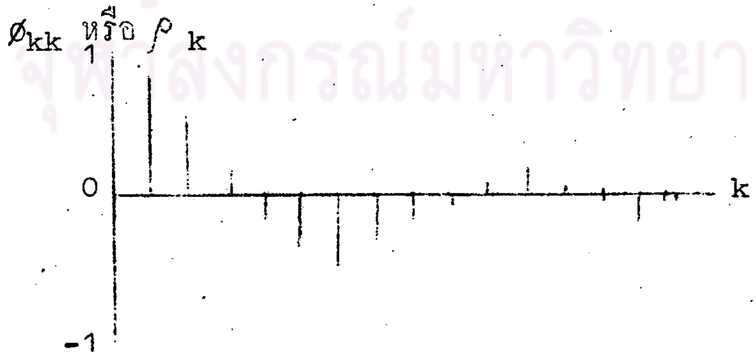
ลักษณะdamped exponential to zeroหรือ smooth ทางบวก



ลักษณะdamped exponention in sign หรือ oscillate in sign



ลักษณะ damped exponential ทางด้านลบ



ลักษณะ damped exponential แบบ sine waves

ลักษณะดังกล่าวรวมเรียกว่า infiniteหรือ tails off

5. THE FIRST-ORDER AUTOREGRESSIVE–FIRST-ORDER MOVING AVERAGE MODEL

1 AUTOREGRESSIVE MODEL ที่มี ORDER 1

รูปสมการคือ $\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t$

ถ้า ϕ_1 ของ process มีค่าระหว่าง -1 กับ 1 process นั้นเป็น Stationary

ลักษณะของ autocorrelation function

เนื่องจาก ρ_k ต้อง satisfy สมการ

$$\rho_k = \phi_1 \rho_{k-1} \quad , k > 0$$

ซึ่งมีค่าตัว unknown เป็น $\rho_k = \phi_1^k \quad , k \geq 0$

$$\therefore \rho_0 = 1$$

เมื่อ k มีค่ามาก ค่า autocorrelation function ยิ่งมีค่าลดลงตามลำดับ มีลักษณะเป็น exponential และสรุปได้ว่า

เมื่อ ϕ_1 มีค่าเป็นบวก autocorrelation function จะมีลักษณะ exponential จนถึงค่าต่ำที่สุดคือศูนย์ (exponentially to zero)

เมื่อ ϕ_1 มีค่าเป็นลบ autocorrelation function จะมีลักษณะเป็น exponentially to zero เหมือนกรณีที่ ϕ_1 มีค่าเป็นบวก แต่ค่าของ ρ_k จะสลับระหว่างค่าบวกและค่าลบตามลำดับของ k (oscillate in sign)

ลักษณะดังกล่าวจะเห็นได้จากรูปที่ 3.3

2. AUTOREGRESSIVE MODEL ที่มี ORDER 2

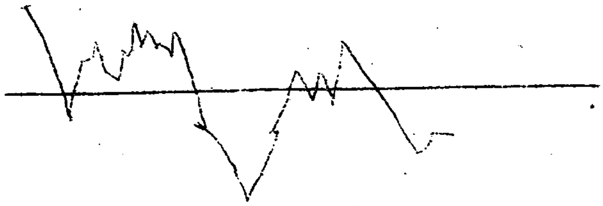
รูปสมการคือ $\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t$

เงื่อนไขที่ทำให้ process เป็น Stationary (stationary condition) คือ รากของสมการ characteristic $\phi(B) = 1-\phi_1 B-\phi_2 B^2 = 0$ ต้องมีค่าอยู่นอก unit circle นั่นคือ ค่าพารามิเตอร์ ϕ_1 และ ϕ_2 ต้องตกอยู่ภายในบริเวณเนื้อที่รูปสามเหลี่ยม (triangular region) ในภาพ 3.4.1

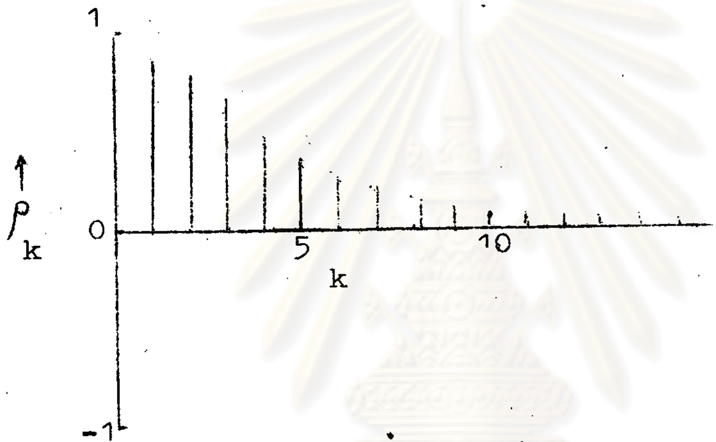
ลักษณะของ autocorrelation, partial autocorrelation function

ลักษณะของ autocorrelation function ของ AR (2) ไม่ว่ารากของสมการ characteristic จะเป็นค่าจริงหรือ complex ก็ยังคงมี

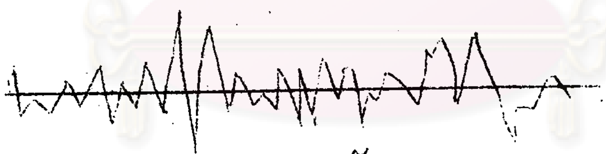
ภาพที่ ๓.๓ ลักษณะ autocorrelation function จาก AUTOREGRESSIVE PROCESS ที่มี ORDER เป็น ๑



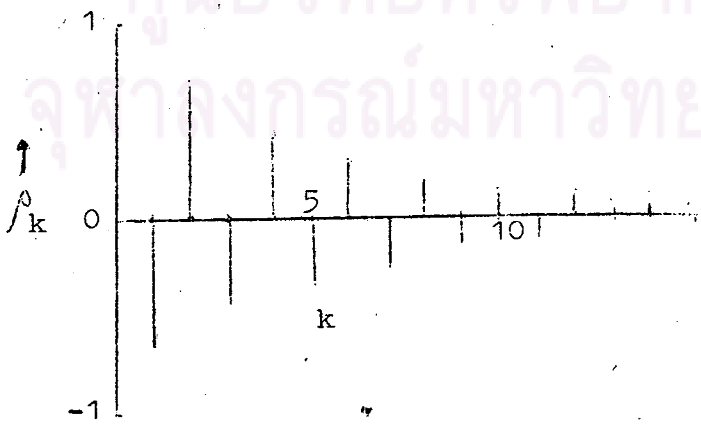
$$\tilde{z}_t = 0.8\,\tilde{z}_{t-1} + a_t$$



ลักษณะ autocorrelation function exponential to zero



$$\tilde{z}_t = -0.8\,\tilde{z}_{t-1} + a_t$$



ลักษณะ autocorrelation function แบบ oscillate in sign

ลักษณะ infinite หรือ tails off

๑.ถ้ารากของสมการ characteristic คือ $\phi(B)= 1- \phi_1 B- \phi_2 B^2 =0$ มีค่าจริง autocorrelation function จะมีลักษณะเป็นลักษณะผสมระหว่าง damped exponentials (Mixture of damped exponentials) หรือเมื่อ $\phi_1^2 +4 \phi_2 < 0$ ค่านี้จะอยู่ในบริเวณเนื้อที่ที่ ๑ และ ๒ ซึ่งอยู่เหนือเขตของพาราโบลา (parabolic boundary) โดยเฉพาะในพื้นที่เขต ๑ autocorrelation function ยังคงเป็นบวก ในขณะที่มีลักษณะเป็น damped out ด้วย ส่วนในพื้นที่ที่ ๒ ค่า autocorrelation function จะสลับระหว่างค่าบวกและลบในขณะที่ damped out

ลักษณะต่างๆดังกล่าวแสดงในรูป๓.๔.๑

๒. ถ้ารากของสมการ characteristic $\phi(B)= 1- \phi_1 B-\phi_2 B^2 =0$ เป็น complex หรือ เมื่อ$\phi_1^2+ 4\phi_2$ มีค่าน้อยกว่า ๐ AUTOREGRESSIVE PROCESS ORDER 2 จะมีลักษณะเป็นแบบ periodic เทียม(pseudo periodic behavior) ลักษณะนี้ทำให้ autocorrelation function มีลักษณะเป็น damped sine wave ซึ่งดูได้จากภาพที่ ๓.๔.๑ในพื้นที่เขต ๓ และ ๔

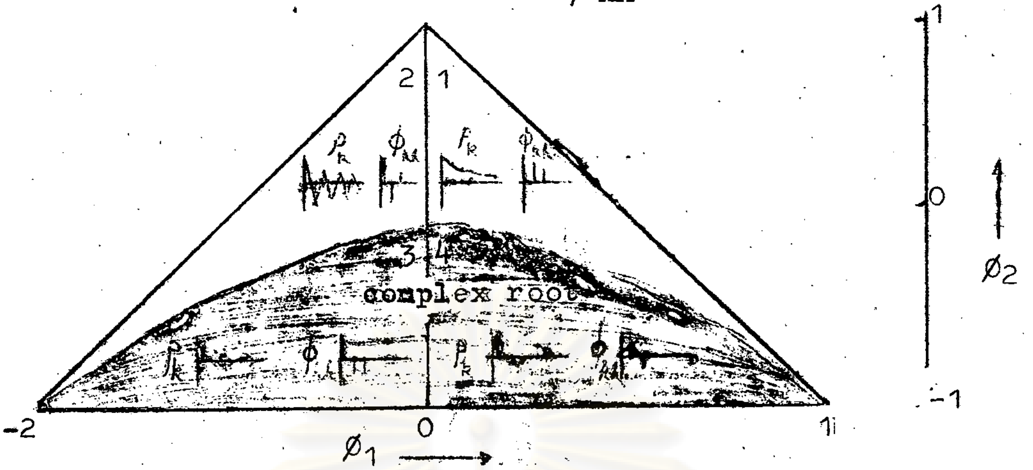
ในพื้นที่เขต ๔ ค่าของ autocorrelation function เริ่มด้วยค่าบวกใน ๒-๓ lag แรก แต่ในพื้นที่เขต ๓ ค่าของ autocorrelation function มักจะมีค่า switches sign จาก lag ๐ ถึง lag 1

สำหรับลักษณะของ partial autocorrelation function ของ AUTOREGRESSIVE MODEL ที่มี ORDER เป็น ๒ จะมีลักษณะ cuts off หลังจาก lag 2 เป็นต้นไป ไม่ว่า รากสมการ characteristic จะมีลักษณะเช่นใด ดังจะเห็นในภาพ ๓.๔.๑

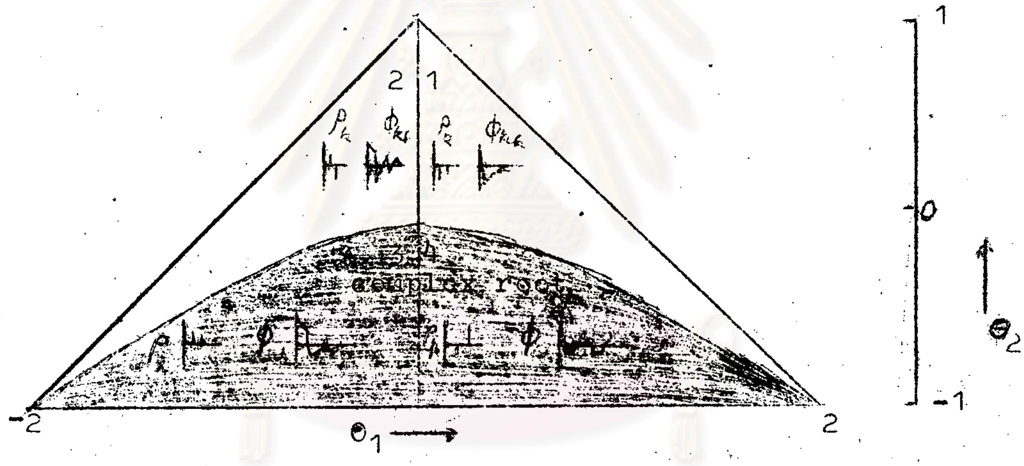
การประมาณค่าพารามิเตอร์ของสมการ AUTOREGRESSIVE

จากสมการ Yule-Walker ซึ่งแสดงความสัมพันธ์ระหว่างพารามิเตอร์กับ autocorrelation function ค่าประมาณของพารามิเตอร์ได้จากการแก้สมการ Yule- Walker เรียกค่าพารามิเตอร์นี้ว่า Yule-Walker estimate

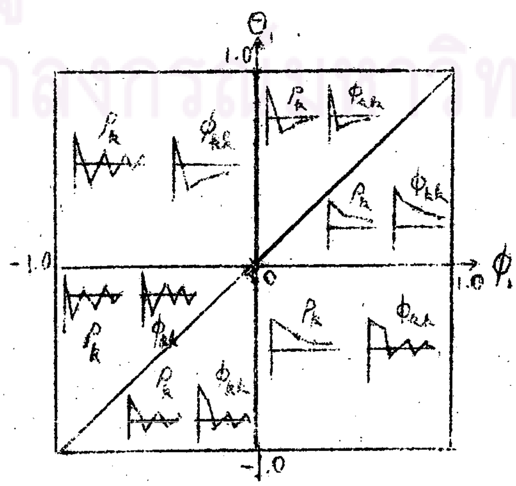
ภาพที่ ๓.๔ ลักษณะ autocorrelation function ρ_k และ partial autocorrelation function ϕ_{kk} ของแบบจำลอง



ภาพที่ ๓.๔.๑ ลักษณะ ρ_k และ ϕ_{kk} สำหรับแบบจำลอง AR (2)



ภาพที่ ๓.๔.๒ ลักษณะ ρ_k และ ϕ_{kk} สำหรับแบบจำลอง MA(2)



ภาพที่ ๓.๔.๓ ลักษณะ ρ_k และ ϕ_{kk} สำหรับแบบจำลอง ARMA(1, 1)

Yule-Walker estimate ของ AR process order p

ในรูปmatrix เป็น

$$\hat{\phi} = R^{-1} r$$

$$\text{โดยที่} \quad R = \begin{bmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{bmatrix}$$

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix}$$

เมื่อ r เป็น vector ของ autocorrelation function ของ lag ที่ $k=1,2,\ldots,p$

ดังนั้น พารามิเตอร์ของ AR process order 1 คือ

$\hat{\phi}_1 = r_1$ = ค่าประมาณ autocorrelation function lag $k=1$

สำหรับAR process order2ค่าพารามิเตอร์ได้จากการแก้สมการ

$$r_1 = \hat{\phi}_1 + \hat{\phi}_2 r_1$$

$$r_2 = \hat{\phi}_1 r_1 + \hat{\phi}_2$$

เมื่อแก้สมการทั้ง ๒ จะได้ค่าพารามิเตอร์ในรูปของ

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}$$

$$\hat{\phi}_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

โดย r_1, r_2 = ค่าประมาณของ autocorrelation function ที่ lag $k=1,2$ ตามลำดับ

๓. MOVING AVERAGE MODEL ORDER 1

รูปสมการคือ $\tilde{Z}_t = a_t - \theta_1 a_{t-1}$

process เป็น stationary สำหรับทุกค่าของ θ_1

ลักษณะของ autocorrelation function

เนื่องจาก autocorrelation function มีค่าเป็น

$$\rho_k = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & , k=1 \\ 0 & , k \geq 2 \end{cases}$$

ดังนั้นค่า autocorrelation function จะมีลักษณะ cuts off หลังจาก lag 1

ส่วน partial autocorrelation function ได้จาก

$$\phi_{kk} = -\theta_1^k \{1-\theta_1^2\} / \{1-\theta_1^{2(k+1)}\}$$

ดังนั้นค่า $/\phi_{kk}/$ มีค่าน้อยกว่า θ_1^k

จึงสรุปลักษณะ partial autocorrelation function ของ MOVING AVERAGE ORDER 1 ได้ดังนี้

เมื่อ θ_1 มีค่าเป็นบวก partial autocorrelation function มีลักษณะเป็น damped exponentials ทางด้านลบอย่างเดียว

เมื่อ θ_1 มีค่าลบ partial autocorrelation function มีลักษณะเป็น damped exponentials ที่สลับค่าบวกกับค่าลบตามลำดับของ k

4. MOVING AVERAGE MODEL ORDER 2

รูปสมการ $\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

process เป็น stationary ได้ทุกค่าของ θ_1 และ θ_2

ลักษณะของ autocorrelation function

เนื่องจาก autocorrelation function มีค่าเป็น

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

$$\rho_k = 0 \quad , k \geqslant 3$$

ดังนั้น autocorrelation function จึงมีลักษณะ cuts off หลัง lag 2

ลักษณะของ partial autocorrelation function สรุปได้ดังนี้

ถ้ารากของสมการ characteristic $1-\theta_1 B - \theta_2 B^2 = 0$ เป็นจริง ลักษณะของ partial autocorrelation function จะ damp แบบ exponential ถ้ารากของสมการ characteristic เป็น complex ลักษณะของ

partial autocorrelation function จะเป็น damped sine wave

ลักษณะ partial autocorrelation function ของ MOVING AVERAGE MODEL ORDER 2 จึงเหมือนกับ autocorrelation function ของ AUTOREGRESSIVE MODEL ORDER 2 ทุกประการ จะดูลักษณะของ autocorrelation function และ partial autocorrelation function ของ MA (2) ได้ที่ภาพที่ ๓.๔.๒

5. THE FIRST-ORDER AUTOREGRESSIVE-FIRST-ORDER MOVING AVERAGE MODEL

มีรูปสมการเป็น $Z_t - \phi_1 Z_{t-1} = a_t - \phi_1 a_{t-1}$

process เป็น stationary เมื่อ $-1 < \phi_1 < 1$

autocorrelation function หาได้จาก

$$\rho_1 = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \quad , \quad \rho_2 = \phi_1 \rho_1$$

ลักษณะของ autocorrelation function สรุปได้ดังนี้

ถ้า ϕ_1 เป็นบวก ลักษณะของ autocorrelation function เป็น exponential ที่ค่อนข้างเรียบ (smooth) ตั้งแต่ค่า ρ_1 เป็นต้นมา

ถ้า ϕ_1 เป็นลบ ลักษณะของ autocorrelation function ยังคงเป็น exponential แต่มีค่าสลับระหว่างค่าบวกค่าลบตั้งแต่ค่า ρ_1 เป็นต้นมา

นอกจากนั้นเครื่องหมายของ ρ_1 จะเหมือนกับเครื่องหมายของ $(\phi_1 - \theta_1)$

ลักษณะของ partial autocorrelation function

เนื่องจากค่า partial autocorrelation function มีค่าเท่ากับ ρ_1 ดังนั้นลักษณะของ partial autocorrelation function ของ ARMA(1,1) จึงมีลักษณะแบบ MA(1) คือ damped exponentials และสรุปได้ดังนี้

ถ้า θ_1 เป็นบวก damped exponentials ที่ค่อนข้างเรียบตั้งแต่ค่า ρ_1 เป็นต้นมา ตามเครื่องหมายของ $(\phi_1 - \theta_1)$

ถ้า θ_1 เป็นลบ ต่างกับกรณีแรกคือ exponential ซึ่ง oscillate ตั้งแต่ค่า ρ_1 ตามเครื่องหมายของ $(\phi_1 - \theta_1)$

ลักษณะของ autocorrelation function และ partial autocorrelation function ของ ARMA(1,1) ดูได้จากภาพที่ ๓.๔.๓

จากลักษณะของ autocorrelation function และ partial autocorrelation function ที่สำคัญดังกล่าวเหล่านั้นจึงพอเป็นแนวทางในการเลือกแบบจำลองสำหรับข้อมูลอนุกรมเวลาคือ

สร้างแบบจำลอง AUTOREGRESSIVE ORDER 1 ถ้า autocorrelation function มีลักษณะเป็น exponential to zero ทั้งแบบสลับเครื่องหมายหรือไม่ก็ตามและ partial autocorrelation function มีค่าเป็น ๐ หลังจาก lag 1 แต่ถ้าลักษณะ exponential เป็นลักษณะของ partial autocorrelation function แทน และ autocorrelation function มีค่า cuts off หลัง lag 1 แล้ว ควรสร้างแบบจำลอง MOVING AVERAGE ORDER 1

สำหรับการสร้างแบบจำลองที่มี ORDER 2 ยังคงพิจารณาทั้งลักษณะ autocorrelation function และ partial autocorrelation function ถ้าลักษณะของ autocorrelation function เป็น damped exponentials แบบใดก็ตาม เช่น แบบ sine wave, แบบ oscillate แบบ smooth และลักษณะ partial autocorrelation function มีค่าเป็น ๐ หลังจาก lag 2 เป็นต้นไป สร้างแบบจำลอง AUTOREGRESSIVE ORDER 2 แต่ถ้าลักษณะของ autocorrelation function เป็น ๐ หลัง lag 2 และลักษณะ partial autocorrelation function เป็น damped exponentials แบบใดก็ได้ แบบจำลอง MOVING AVERAGE ORDER 2 เป็นแบบที่เหมาะกว่าแบบอื่น

ในกรณีที่ autocorrelation และ partial autocorrelation function มีลักษณะนอกเหนือจากลักษณะดังกล่าว ควรใช้แบบจำลอง ARMA(1,1) ซึ่งทั้ง autocorrelation function และ partial autocorrelation function มีลักษณะ damped exponentials แต่เริ่มที่จุด ρ_1 เป็นต้นไป ซึ่ง ρ_1 นี้มีเครื่องหมายตาม $(\phi_1 - \theta_1)$ และลักษณะ damped exponentials ของ autocorrelation function เป็นแบบใดนั้นขึ้นกับเครื่องหมายของ ϕ_1 ขณะเดียวกันเครื่องหมายของ θ_1 ก็กำหนดแบบของ damped exponentials ของ partial autocorrelation function ด้วย

๓.๓ แบบจำลอง NONSTATIONARY

ในการวิเคราะห์อนุกรมเวลา อาจจะมีอนุกรมเวลาที่ไม่ต้องการวิเคราะห์เกี่ยวกับความเบี่ยงเบนไปจากแนวโน้มหรือ fixed mean แต่อนุกรมเวลายังคงแสดงความ HOMOGENEOUS หรือมีลักษณะคล้ายแบบเดิมและเมื่อเวลาเปลี่ยนไประดับหรือค่าเฉลี่ยของความกระเพื่อม (fluctuation) ก็เปลี่ยนไปด้วย ลักษณะดังกล่าวนี้เป็นลักษณะหนึ่งของ Nonstationary time series ซึ่งอนุกรมส่วนแตกต่างที่d (d^{th} difference) หรือส่วนแตกต่างระหว่างค่าในอนุกรม Nonstationary นั้นเองจะมี process เป็น Stationary อนุกรมส่วนแตกต่างที่ d นี้จึงสามารถแทนได้ด้วยแบบจำลอง Linear Stationary ส่วนค่า d นี้ในทางปฏิบัติพบว่ามีค่าไม่เกิน ๒ คือ มักมีค่า ๐, ๑, ๒ เสมอ[1]

ให้ $w_t = \nabla^d z_t$ หรือส่วนแตกต่างที่ d

$$\nabla^d z_t = z_t - z_{t-1} - \ldots z_{t-d}$$

d = 0 แสดงว่าไม่นำความแตกต่างมาเกี่ยวข้องยังคงเป็นอนุกรมเวลาเดิมคือ z_t

$$d = 1, \quad \nabla z_t = z_t - z_{t-1}$$

$$d = 2, \quad \nabla^2 z_t = z_t - z_{t-1} - z_{t-2}$$

การหาแบบจำลอง NONSTATIONARY จึงเริ่มด้วยการสร้างแบบจำลอง Linear Stationary ให้กับอนุกรมส่วนแตกต่างที่ ๐, อนุกรมส่วนแตกต่างที่๑, อนุกรมส่วนแตกต่างที่๒, แล้วเลือกแบบจำลองที่สอดคล้องกันระหว่าง ๓ แบบจำลองนี้ เป็นแบบจำลองสำหรับอนุกรม Nonstationary

ดังนั้นการเลือกแบบจำลองNONSTATIONARY จึงต้องอาศัยการดูลักษณะของautocorrelation function และpartial autocorrelation function ของอนุกรม z_t, ∇z_t และ $\nabla^2 z_t$ แทน z_t เพื่อหาค่า p,q,d ที่เหมาะกับข้อมูล

จะเห็นได้ว่า โดยทั่วไปแล้ว NONSTATIONARY MODEL เป็นแบบจำลองที่วิเคราะห์อนุกรมเวลาของตัวแปรโดยไม่เกี่ยวข้องกับระดับของค่าเฉลี่ยหรือแนวโน้ม

[1]Box&Jenkins, Time Series Analysis forecasting and control, ("Holden-day), p.11

ของอนุกรมเวลาใดๆ ส่วนSTATIONARY MODEL เป็นแบบจำลองที่วิเคราะห์เกี่ยวกับความเบี่ยงเบนไปจากระดับของค่าเฉลี่ยที่คงที่(a constant mean level) หรือค่าที่คงที่ค่าหนึ่งที่เป็นค่าควบคุม(control) ค่าที่แปรไป

ดังนั้นการวิเคราะห์อนุกรมเวลาของอัตราตายของทารกในบทต่อไปเพื่อหาแบบจำลองที่เหมาะกับส่วนเบี่ยงเบนจากแนวโน้ม (Long term trend)ด้วยทฤษฎีบางประการของStationary time series ซึ่งให้แบบจำลอง Linear Stationary นอกจากนี้ยังพิจารณา Linear Nonstationary model ให้กับอนุกรมเวลาอัตราตายของทารกโดยตรงอีกด้วย.

3.4 การทดสอบความเหมาะสมของแบบจำลอง (Lack of fit test)

ความเหมาะสมของแบบจำลองนี้จะพิจารณาจาก autocorrelation function ของ residual ของแบบจำลองคือ a_t เป็นการทดสอบค่าของ autocorrelation function รวมทั้งหมด ด้วย statistic Q

$$Q = n \sum_{k=1}^{K} r_k^2 (\hat{a})$$

การแจกแจงของ Q เป็น approximate Chi-square ที่มี degree of freedom เป็น K-p-q

เมื่อ K เป็นค่า lag ที่มากที่สุด

p เป็น ORDER ของ Autoregressive process

q เป็น ORDER ของ Moving Average process

n = N-d คือ จำนวนค่าสังเกตที่ใช้สร้างแบบจำลอง

d คือ จำนวนช่วงเวลาที่แตกต่าง เช่น d=1 อนุกรมเวลาของตัวแปรจะเปลี่ยนเป็นอนุกรมเวลาของส่วนแตกต่างที่ 1 ของตัวแปรนั้น และมีจำนวนค่าสังเกตลดลง 1 ค่า

การทดสอบจะอาศัยค่า Q ที่คำนวนได้มาเทียบกับตาราง χ^2 ที่มี degree of freedom เป็น K-p-q ถ้า Q มีค่าน้อยกว่า χ^2 ในตาราง แสดงว่าแบบจำลองที่สร้างให้กับอนุกรมเวลาของค่าสังเกตมีความเหมาะสมพอ เพราะถ้ายิ่งแบบจำลองที่สร้างไม่เหมาะกับข้อมูล จะทำให้ค่าเฉลี่ยของ Q สูงมากขึ้น (inflate)