บทที่ 5
BOX - JENKINS

ในบทก่อนหน้านี้ได้อธิบายถึงวิธีการพยากรณ์เชิงปริมาณ ซึ่งใช้กันอย่างแพร่หลายในเชิงธุรกิจ เทคนิควิธี Box-Jenkins นี้เป็นตัวแทนซึ่งค้นพบว่าได้ประโยชน์ที่สุดและประหยัดค่าใช้จ่ายที่สุดของบริษัทต่าง ๆ ในสถานการณ์ปัจจุบัน มีเทคนิคการพยากรณ์เชิงปริมาณจำนวนมากที่ผู้บริหารต้องทำความคุ้นเคยกับเทคนิคทั้งหลายเหล่านั้น จุดประสงค์ของเทคนิควิธี Box-Jenkins และ Economic Forecasting จะเกี่ยวข้องกับการเพิ่มวิธีการประยุกต์มากกว่า ซึ่งผู้บริหารอาจจะปล่อยความสมบูรณ์ของเทคนิคเหล่านั้นไว้สำหรับคำแนะนำ ต่าง ๆ โดยผู้เชี่ยวชาญทางด้านการพยากรณ์ หรือโดยผู้อื่นในบริษัทของเขา หรืออุตสาหกรรมของเขา เทคนิคนี้ค้นพบได้ว่ามีบทบาทอย่างมากในหลาย ๆ สถานการณ์ อย่างไรก็ตาม ถ้ามีราคาค่าใช้จ่ายในการพยากรณ์ที่สูงกว่า วิธีอื่น ๆ ที่ได้กล่าวมาแล้ว ความจำเป็นของวิธีนี้จะมีค่ามากกว่าสำหรับการนำไปใช้ในการพยากรณ์ ก่อนที่จะทำให้มันสมบูรณ์ ในทางปฏิบัติยังต้องการกระบวนการจัดการที่สอดคล้องในแต่ละขั้นตอนของโครงการด้วย

5.1 การวิเคราะห์สหสัมพันธ์ในตัวเอง (Autocorrelation Analysis)

สหสัมพันธ์ในตัวเอง (autocorrelation) หมายถึง ความสัมพันธ์ของตัวแปรสุ่มชุดเดียวกันแต่ต่างวาระกัน (time lag) เช่น autocorrelation ระหว่าง X_t กับ X_{t-1} ($r_{X(t)X(t-1)}$) หรือ X_t กับ X_{t-2} หรือระหว่าง X_t กับ X_{t-k} (ในกรณีที่ lag ไป k วาระ) ค่าของ autocorrelation เป็นตัวชี้ให้เห็นว่า กาลเวลาแปรเปลี่ยนไปจะมีผลต่อการเปลี่ยนแปลงค่าของตัวแปรหรือไม่ ให้พิจารณาจาก

1. ค่าของ $r_{X(t)X(t-s)}$ มีค่าเข้าใกล้ศูนย์ เมื่อ $s = 1,2,\ldots,k$ หมายถึง ไม่ว่าเวลาจะ lag ไปในอดีตหรืออนาคต ค่าของข้อมูล X ในกาล t และ t - s ก็ไม่เกี่ยวข้องกัน ซึ่งจะแสดงว่าข้อมูลอนุกรมเวลาชุดนั้นไม่มีรูปแบบ (pattern) ใด ๆ เลย ดังนั้น ข้อมูล X_t มีลักษณะเป็น stationary หรืออีกกรณีหนึ่งที่เปลี่ยนค่า s ไปเพียง 2 - 3 ค่า โดยที่ $s = 1,2,3$ แล้วค่า autocorrelation ค่อย ๆ ลดลง แต่ถ้า $s = 4,5,6,\ldots,k$ ค่าของ $r_{X(t)X(t-s)} = 0$ แสดงว่าค่า X_t กับ X_{t-s} ใกล้เคียงกัน เช่นนี้ค่า $X_i \cong \overline{X}$ แล้วข้อมูลอนุกรมเวลาชุดนั้น ๆ มีลักษณะเป็น stationary

$$r_{X(t)\,X(t-s)} = \frac{\sum_{i=1}^{n-s} (X_{(t,i)} - \overline{X}_t)(X_{(t-s,i)} - \overline{X}_{t-s})}{\sqrt{\sum_{i=1}^{n} (X_{(t,i)} - \overline{X}_t)^2 \sum_{i=1}^{n-s} (X_{(t-s,i)} - \overline{X}_{t-s})^2}}$$

หรือ $$r_{X(t)\,X(t+s)} = \frac{\sum_{i=1}^{n-s} (X_{(t,i)} - \overline{X}_t)(X_{(t+s,i)} - \overline{X}_{t+s})}{\sqrt{\sum_{i=1}^{n} (X_{(t,i)} - \overline{X}_t)^2 \sum_{i=1}^{n-s} (X_{(t+s,i)} - \overline{X}_{t+s})^2}}$$

2. ค่า $r_{X(t)\,X(t-s)}$ ค่อย ๆ ลดลง เมื่อ $s = 1,2,...,k$ แล้วนำค่า autocorrelation ที่ได้ไป plot บนกราฟแกน (r , s) โดย r เป็นแกนนอนมีค่าตั้งแต่ -1 ถึง 1 , s เป็นแกนตั้งเป็นค่าของ time lag ถ้ากราฟเป็นแนวเส้นตรงค่อย ๆ ลดลงลาดไปทางขวา หรือ ลาดจากขวาไปซ้าย หรือ จุด (r , s) กระจายเป็นแนวเส้นตรงดังกล่าว บ่งบอกลักษณะว่าข้อมูลมีรูปแบบแนวโน้มอยู่

3. ค่า $r_{X(t)\,X(t-s)}$ plot แล้วเป็นรูปกราฟขึ้นลงซ้ำรูปแบบเดียวกันเสมอ ในทุกวาระที่คงที่ อาจจะเป็นทุก ๆ สี่เดือน หรือทุก ๆ หกเดือน แสดงว่า ข้อมูลมีรูปแบบของปัจจัยฤดูกาลอยู่

5.1.1 เกณฑ์ในการพิจารณาการวิเคราะห์สหสัมพันธ์ในตัวเอง

1. ให้พิจารณารูปแบบของกราฟ X_t ของ autocorrelation ในลักษณะ 1 - 4 ที่กล่าวมาแล้ว

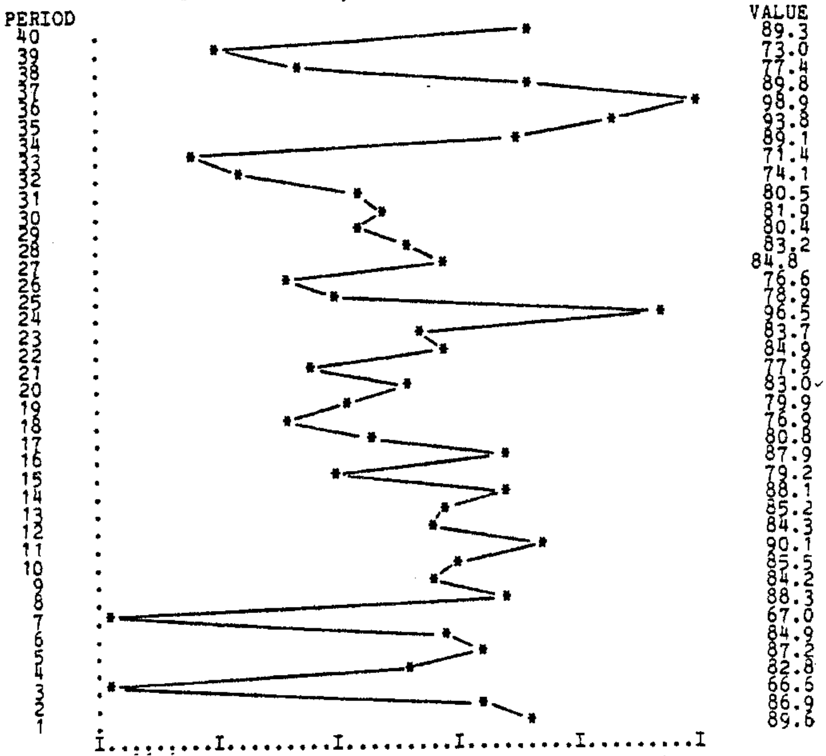
2. การทดสอบนัยสำคัญโดยพิจารณาค่าของ autocorrelation ว่ามีนัยสำคัญหรือไม่ โดยเมื่อเขียนกราฟของ autocorrelation แล้วจะมีเส้นประขนานกัน 1 คู่ เป็น 95 % ช่วงความเชื่อมั่น (confident band) เมื่อคู่ลำดับ (r , s) ตกอยู่ภายในแถบดังกล่าว แสดงว่า $r = 0$ แต่ถ้าตกนอกแถบนี้ จะได้ว่า $r \neq 0$

3. การทดสอบนัยสำคัญ โดยใช้ χ^2 - test เมื่อตั้งสมมติฐานว่า $H_0 : \rho = 0$ $H_1 : \rho \neq 0$

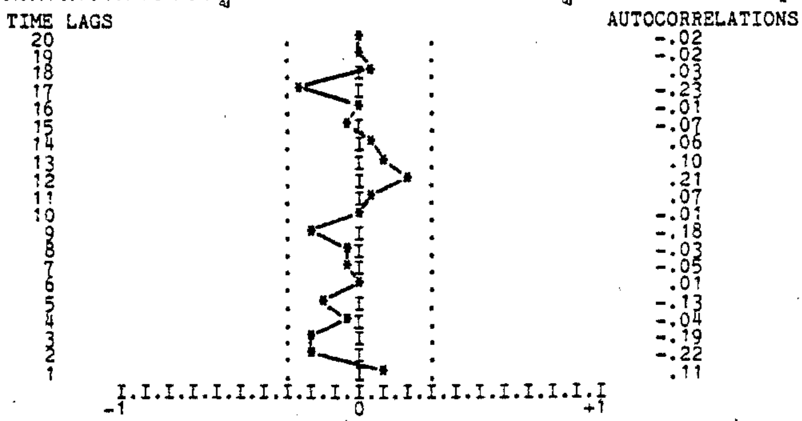
หรือ H_0 : ข้อมูลไม่มี pattern H_1 : ข้อมูลมี pattern โดยปฏิเสธ H_0 เมื่อ $\chi^2_c > \chi^2_{(k-1),(1-\alpha)}$

เมื่อ $\chi^2_c = n \sum_{s=1}^{k} r^2_{X(t)X(t-s)}$ ผลสรุปถ้าเรายอมรับ H_0 นั่นคือ ข้อมูลอนุกรมเวลาเป็น random

ตัวอย่างที่ 5.1 กราฟของข้อมูลเดิมใน 40 period จงวิเคราะห์กราฟของ autocorrelation $r_{X(t)X(t-s)}$ และพิจารณาว่ามี pattern ในข้อมูลหรือไม่



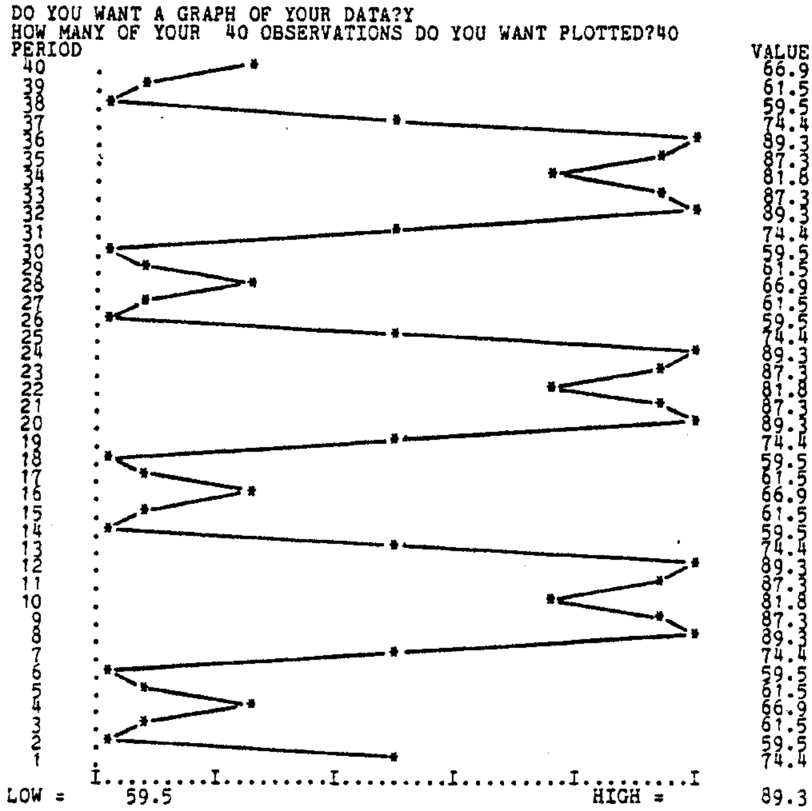
รูปที่ 5.1 แสดงกราฟของข้อมูลเดิมโดยแกนนอนเป็นค่าข้อมูลเดิม แกนตั้งเป็น period



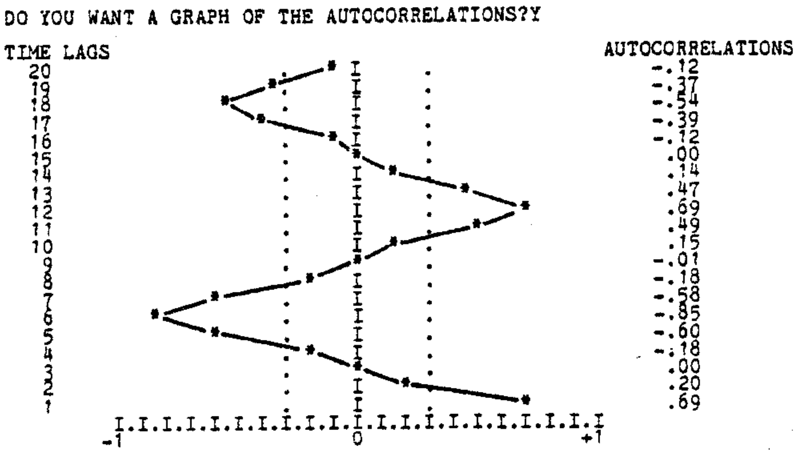
รูปที่ 5.2 แสดงกราฟของ autocorrelation โดย lag ไป 20 วาระ แกนนอนเป็น r แกนตั้งเป็น time lag (s)

ผลสรุป จากรูปที่ 5.2 จะได้ว่าคู่ลำดับ (r , s) ทุกจุดตกอยู่ภายใน confident band ค่า r มีค่าเข้าใกล้ 0 เกือบทุกค่า ข้อมูลอนุกรมเวลาชุดนี้ไม่มี pattern จึงเป็นประเภท randomness

ตัวอย่างที่ 5.2 จากกราฟเป็นค่าของข้อมูลเดิมใน 40 period จงวิเคราะห์ว่าข้อมูลเป็น randomness หรือไม่



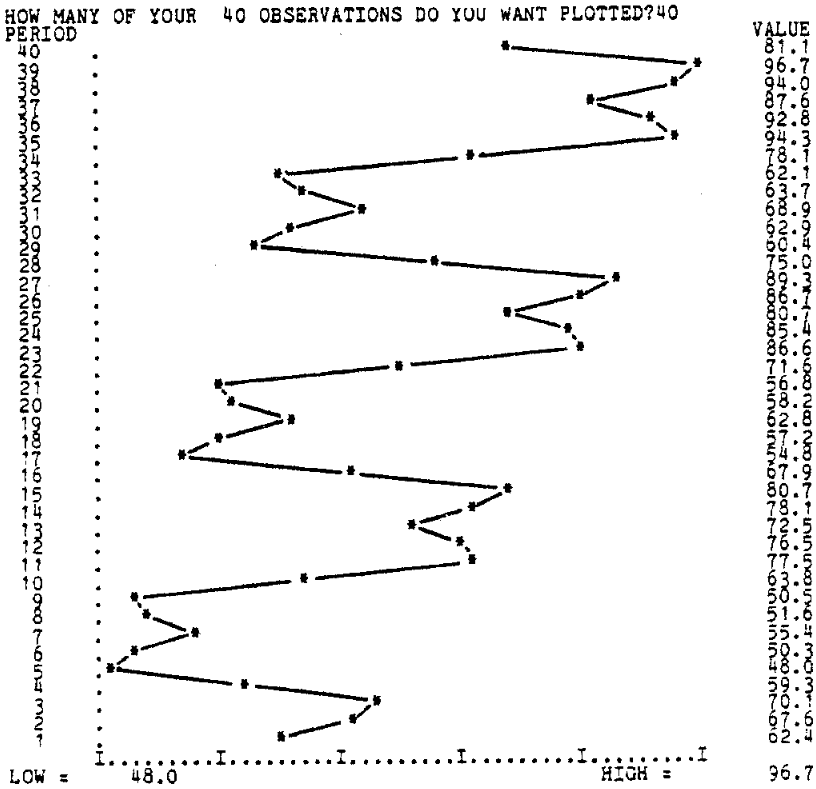
รูปที่ 5.3 แสดงกราฟของข้อมูลเดิม 40 period



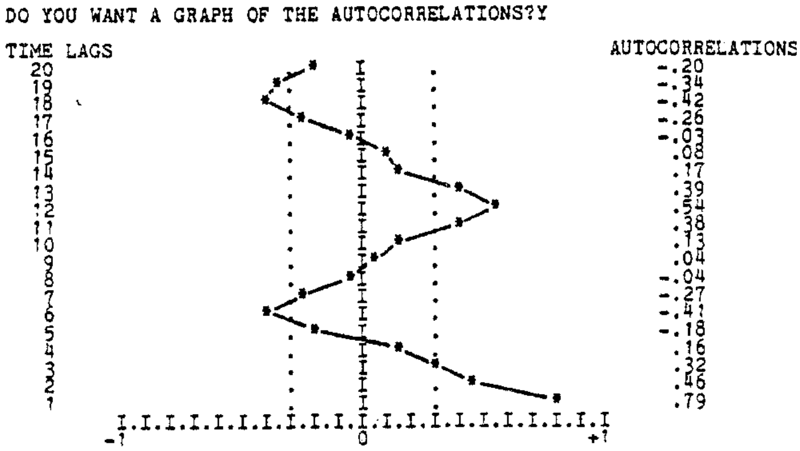
รูปที่ 5.4 แสดงกราฟของ autocorrelation โดย lag ไป 20 วาระ

ข้อสรุป จากรูปที่ 5.3 ไม่สามารถวิเคราะห์ออกมาได้ว่าข้อมูลมี pattern ใด แต่เมื่อ plot กราฟของ autocorrelation จากรูปที่ 5.4 จะเห็นว่า ค่า r ตกนอก confident band หลายจุดและค่า r ส่วนใหญ่ไม่เข้าใกล้ศูนย์ ดังนั้นจึงต้องพิจารณาว่าข้อมูลมีรูปแบบใด เมื่อพิจารณาค่า r จะมีค่าเพิ่มขึ้นทุก ๆ 6 วาระ ข้อมูลจึงเป็นลักษณะของ seasonal มีช่วงกว้างของ seasonal ยาว 6 วาระ

ตัวอย่างที่ 5.3 จากกราฟของข้อมูลเดิม 40 period จงวิเคราะห์ข้อมูลว่ามี pattern แบบใด

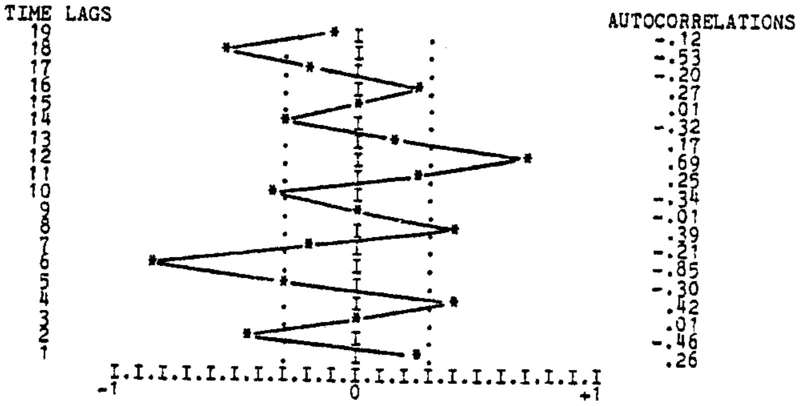


รูปที่ 5.5 แสดงกราฟของข้อมูลเดิม 40 วาระ



รูปที่ 5.6 แสดงกราฟของ autocorrelation lag 20 วาระ

DO YOU WANT A GRAPH OF THE AUTOCORRELATIONS?Y

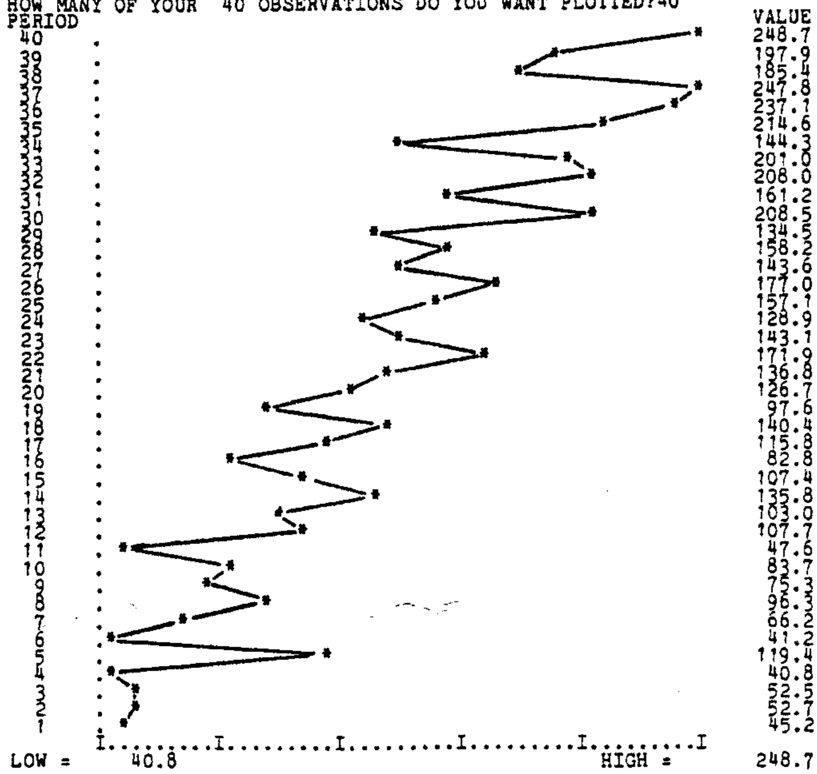


รูปที่ 5.7 แสดงกราฟของ autocorrelation หลังจากทำ first difference แล้ว

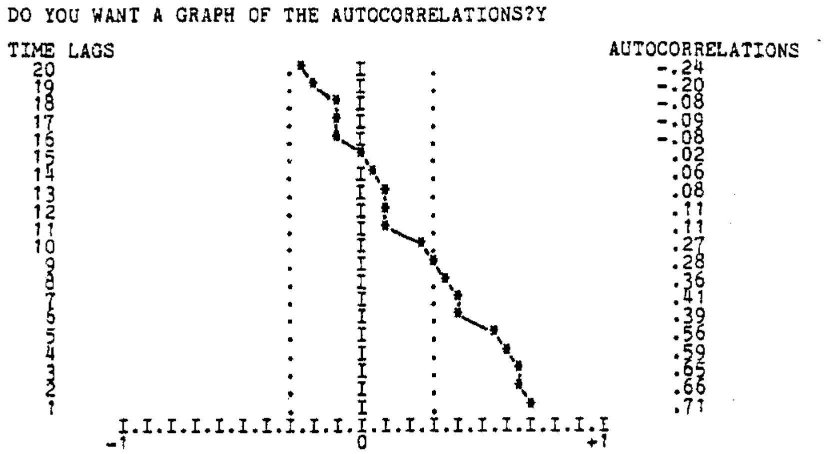
<u>ข้อสรุป</u> จากรูปที่ 5.5 กราฟของข้อมูลเดิมมีลักษณะของ trend ปรากฏอยู่ แต่จะมี pattern อื่นอีกหรือไม่ลองพิจารณาจากรูปที่ 5.6 กราฟของ autocorrelation มีลักษณะของ seasonal อยู่โดยมีช่วงกว้าง 6 วาระ จากนั้นเรากำจัด trend ออกจากข้อมูลเดิมโดยการทำ first difference คือ กราฟของ autocorrelation ตามรูปที่ 5.7 จะเห็นว่าข้อมูลมี seasonal อยู่อย่างแน่นอน ดังนั้นข้อมูลอนุกรมเวลาชุดนี้จึงมีทั้ง trend และ seasonal

ตัวอย่างที่ 5.4 จากกราฟของข้อมูลเดิม 40 period จงวิเคราะห์ข้อมูลว่ามี pattern ใด

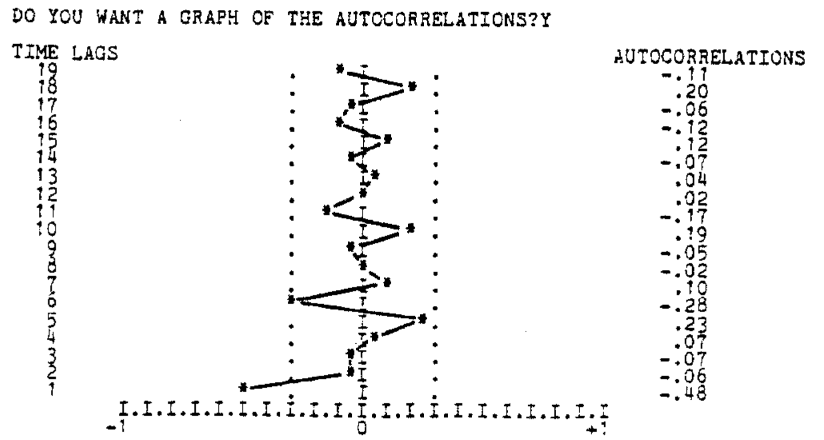
DO YOU WANT A GRAPH OF YOUR DATA?Y
HOW MANY OF YOUR 40 OBSERVATIONS DO YOU WANT PLOTTED?40



รูปที่ 5.8 แสดงกราฟของข้อมูลเดิม 40 period



รูปที่ 5.9 แสดงกราฟของ autocorrelation lag 20 วาระ



รูปที่ 5.10 แสดงกราฟของ autocorrelation หลังจากทำ first difference แล้ว

ข้อสรุป จากรูปที่ 5.8 กราฟของข้อมูลเดิมพิจารณาแล้วมีลักษณะของ trend อยู่ แต่จะมี pattern อื่น ๆ อีกหรือไม่ จึงเขียนกราฟของ autocorrelation ตามรูปที่ 5.9 ลักษณะกราฟของ autocorrelation ค่อย ๆ ลดลง แสดงว่ามี trend อยู่ในข้อมูล แต่ถ้าต้องการดูว่ามี pattern อื่น ๆ อีกหรือไม่ ให้กำจัด trend ออกจากข้อมูล โดยทำ first difference กราฟของ autocorrelation หลังจากทำ first difference แล้วตามรูปที่ 5.10 ค่า r มีค่าเข้าใกล้ 0 จึงสรุปได้ว่าไม่มี pattern อื่น ๆ อีก ดังนั้นข้อมูลอนุกรมเวลาชุดนี้มี trend เพียงปัจจัยเดียว

การเข้าใจในการวิเคราะห์กราฟของ autocorrelation เป็นสิ่งสำคัญมาก ซึ่งไม่เหมือนกับกราฟของข้อมูลเดิม คือแทนที่จะมีการตรวจสอบกราฟของข้อมูลทั้งหมดเพื่อดู pattern ของข้อมูล เราก็ไปพิจารณา autocorrelation ที่สอดคล้องกับข้อมูลชุดนั้น การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองสำหรับ k time lag สามารถหาได้จากสูตร

หรือ $r_k = \dfrac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$

เมื่อ r_k เป็นสัมประสิทธิ์สหสัมพันธ์ในตัวเอง

k เป็นจำนวน time lag

n เป็นจำนวนค่าข้อมูล

X_t เป็นค่าข้อมูล ณ วาระ t

$\bar{X}$ เป็นค่าเฉลี่ยของข้อมูลทั้งหมด

ตัวอย่างที่ 5.5 จากข้อมูลต่อไปนี้ จงหา r_k

t	1	2	3	4	5	6	7	8	9	10
X_t	13	8	15	4	4	12	11	7	14	12

เมื่อ $k = 1$ $\bar{X} = (13+8+15+\ldots+14+12)/10 = 10$

$r_1 = [(X_1 - \bar{X})(X_2 - \bar{X}) + (X_2 - \bar{X})(X_3 - \bar{X}) + \ldots + (X_9 - \bar{X})(X_{10} - \bar{X})]$ /

$[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_9 - \bar{X})^2 + (X_{10} - \bar{X})^2]$

$r_1 = [(13-10)(8-10) + (8-10)(15-10) + \ldots + (14-10)(12-10)] / [(13-10)^2 + (8-10)^2 + \ldots +$

$(14-10)^2 + (12-10)^2]$

$r_1 = -27 / 144 = -0.188$

เมื่อ $k = 3$ ได้

$r_3 = [(X_1 - \bar{X})(X_4 - \bar{X}) + (X_2 - \bar{X})(X_5 - \bar{X}) + \ldots + (X_7 - \bar{X})(X_{10} - \bar{X})]$ /

$[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_9 - \bar{X})^2 + (X_{10} - \bar{X})^2]$

$r_3 = [(13-10)(4-10) + (8-10)(4-10) + \ldots + (11-10)(12-10)] / [(13-10)^2 + (8-10)^2 + \ldots +$

$(14-10)^2 + (12-10)^2]$

$r_3 = 26 / 144 = 0.18$

5.1.2 กระบวนการทำให้ข้อมูลเป็นเชิงสุ่ม (Stationary Process)

เมื่อกาลเวลาเปลี่ยนไปค่าเฉลี่ยของข้อมูลมีค่าคงที่ ความผันแปรคงที่ ความผันแปรร่วมไม่เปลี่ยนแปลง ลักษณะเช่นนี้เป็น stationary ไม่มี pattern อยู่ในข้อมูล แต่ถ้าค่าเฉลี่ยและความผันแปรเปลี่ยนแปลงไปตามกาลเวลา ลักษณะเช่นนี้เรียก nonstationary ถ้าข้อมูลเกิด trend สามารถกำจัด trend ออก โดยการทำ first difference หรือ second difference ซึ่งผลจากการทำจะได้ข้อมูลมีระดับ (level) คงที่ ปัจจัย trend ถูกกำจัดออกจากข้อมูลอนุกรมเวลา คือ

t	1^{st} diff ($X'_t = \nabla X_t$)	2^{st} diff ($X''_t = \nabla^2 X_t$)
1	-	-
2	$X'_2 = X_2 - X_1$	-
3	$X'_3 = X_3 - X_2$	$X''_3 = X'_3 - X'_2 = X_3 - 2X_2 + X_1$
4	$X'_4 = X_4 - X_3$	$X''_4 = X'_4 - X'_3 = X_4 - 2X_3 + X_2$
5	$X'_5 = X_5 - X_4$	$X''_5 = X'_5 - X'_4 = X_5 - 2X_4 + X_3$

ตัวอย่างที่ 5.6 จากข้อมูลต่อไปนี้ จงหาค่าข้อมูลหลังจากทำ 1^{st} difference และ 2^{nd} difference

t	X_t	X'_t	X''_t
1	13	-	-
2	8	-5	-
3	15	7	12
4	4	-11	-18
5	4	0	11
6	12	8	8
7	11	-1	-9
8	7	-4	-3
9	14	7	11
10	12	-2	-9

หรือถ้าในกรณีที่มีข้อมูลเป็น simple linear series เช่น 2,4,6,8,10,12,14 สามารถแสดงให้ได้ว่าหลังจากทำ 1^{st} diff แล้ว ข้อมูลจะมี level คงที่ คือ 4 - 2 = 2 , 6 - 4 = 2 , 8 - 6 = 2 , 10 - 8 = 2 , 12 - 10 = 2 , 14 - 12 = 2

5.1.3 การพิจารณาลักษณะของข้อมูลว่ามี pattern หรือไม่ โดยใช้หลาย ๆ วิธีร่วมกัน ดังนี้

1. การใช้ค่าเฉลี่ยของ autocorrelation ของอนุกรมเวลาทั้งหมด ($\bar{r}$) ซึ่งถ้า $\bar{r}$ มีค่าต่ำ แสดงว่าข้อมูลเป็น randomness (หรือ stationary data) แต่ถ้า $\bar{r}$ มีค่าสูงแสดงว่าข้อมูลชุดนั้นมี pattern หรือเป็น nonstationary data โดยที่

$$\bar{r} = (1/k) \sum_{s=1}^{k} r_{X(t)\,X(t-s)}$$

เมื่อ k เป็นจำนวน time lag

$r_{X(t)\,X(t-s)}$ เป็นค่า autocorrelation coefficient ในแต่ละ time lag

2. ค่าเฉลี่ยของ autocorrelation ครึ่งแรกของอนุกรมเวลาทั้งหมด ($\bar{r}_1$) และ ค่าเฉลี่ยของ autocorrelation ครึ่งหลังของอนุกรมเวลาชุดนั้น ($\bar{r}_2$) ซึ่งถ้า $\bar{r}_1$ และ $\bar{r}_2$ มีค่าใกล้เคียงกัน แสดงว่าเป็น stationary data แต่ ถ้า $\bar{r}_1$ และ $\bar{r}_2$ มีค่าแตกต่างกันมาก แสดงว่า ข้อมูลเป็น nonstationary data

3. ส่วนเบี่ยงเบนมาตรฐานของ autocorrelation ($\hat{\sigma}_r$) ถ้ามีค่าน้อยแสดงว่า autocorrelation (r_k) มีค่าใกล้เคียงกัน จึงเป็นแบบ stationary data แต่ถ้า $\hat{\sigma}_r$ มีค่ามาก แสดงว่าข้อมูลเป็น nonstationary data โดยที่ $\hat{\sigma}_r = 1/\sqrt{n}$

4. สัมประสิทธิ์ของความผันแปร (c.v.) ให้นำค่า c.v. ของข้อมูล X_t เปรียบเทียบกับ c.v. ของข้อมูล X'_t (ในกรณีที่ทำ first difference) และอาจเปรียบเทียบกับ c.v. ของ X''_t (ในกรณีที่ทำ second difference) ถ้าค่า c.v. ลดลงแสดงว่าข้อมูลอนุกรมเวลาเป็น nonstationary data แต่ถ้าค่า c.v. เพิ่มขึ้นแสดงว่าข้อมูลอนุกรมเวลาชุดนั้นเป็น stationary data

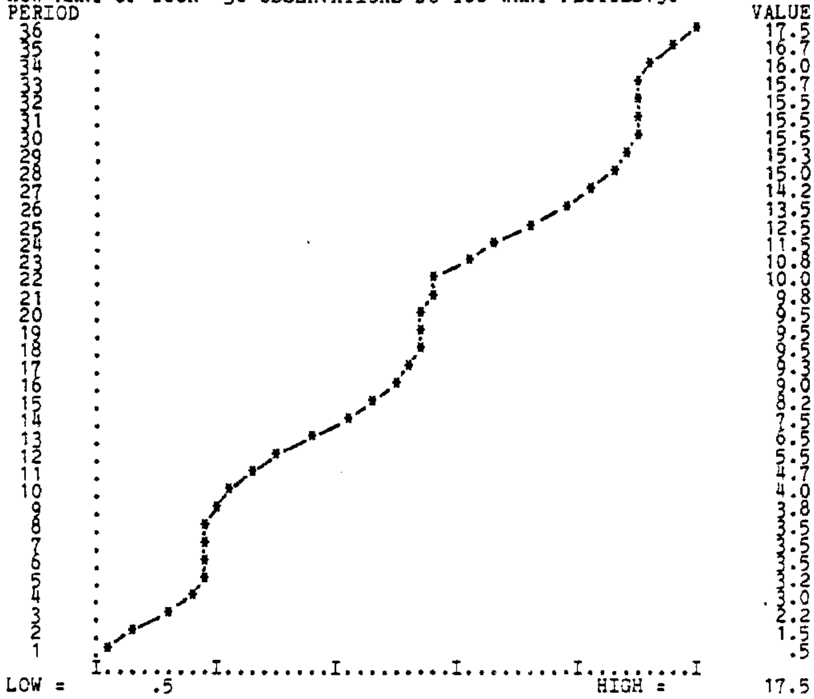
5. ใช้การทดสอบของไคสแควร์ (χ^2 - test) โดยตั้งสมมติฐาน คือ H_0 : ข้อมูลเป็น randomness H_1 : ข้อมูลมี pattern โดยจะปฏิเสธ H_0 เมื่อ $\chi^2_c > \chi^2_{k-1,1-\alpha}$ เมื่อ k เป็นจำนวน time lag สูงสุด

และ $\chi^2_c = n \sum_{s=1}^{k} r^2_{X(t)\,X(t-s)}$ โดยที่ n เป็นจำนวนข้อมูลทั้งหมด

6. ใช้กราฟของ autocorrelation พิจารณาว่าข้อมูลเป็นแบบ stationary หรือ nonstationary ดังได้กล่าวมาแล้วในหัวข้อ 5.1.1

ตัวอย่างที่ 5.7 ข้อมูล 36 ค่า จงวิเคราะห์โดยใช้ autocorrelation พิจารณาว่าข้อมูลมี pattern ใดบ้าง

```
DO YOU WANT A GRAPH OF YOUR DATA?Y
HOW MANY OF YOUR  36 OBSERVATIONS DO YOU WANT PLOTTED?36
```

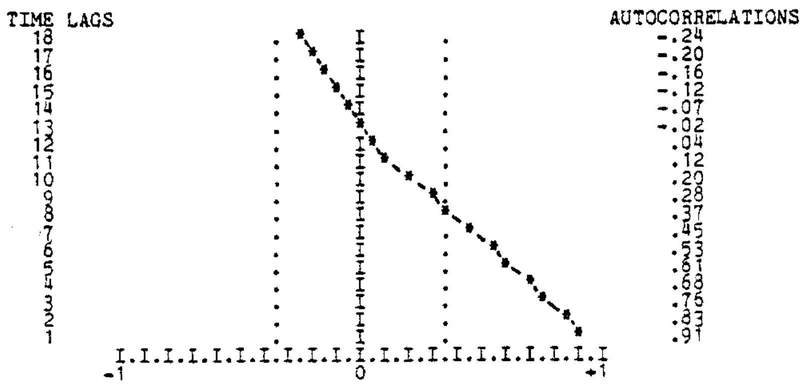


รูปที่ 5.11 แสดงกราฟของข้อมูลเดิม

จากกราฟของข้อมูลเดิมเมื่อเวลาเพิ่มขึ้นค่าข้อมูลจะเพิ่มขึ้นด้วย ทั้งนี้กราฟยังสะบัดไปมา จึงคาดเดาว่าข้อมูลอาจจะมีทั้ง trend และ seasonal จากนั้น plot กราฟของ autocorrelation โดย lag เพียง 18 วาระ ดังนี้

```
MEAN AUTOCORRELATION        .277      STANDARD ERROR=        .167
MEAN OF FIRST   9  VALUES=      .604  CHI SQUARE (COMPUTED)=     138.4
MEAN OF LAST    9  VALUES=     -.051  CHI SQUARE (FROM TABLE)=     8.7
COEFFICIENT OF VARIATION =      .548

DO YOU WANT A GRAPH OF THE AUTOCORRELATIONS?Y
```



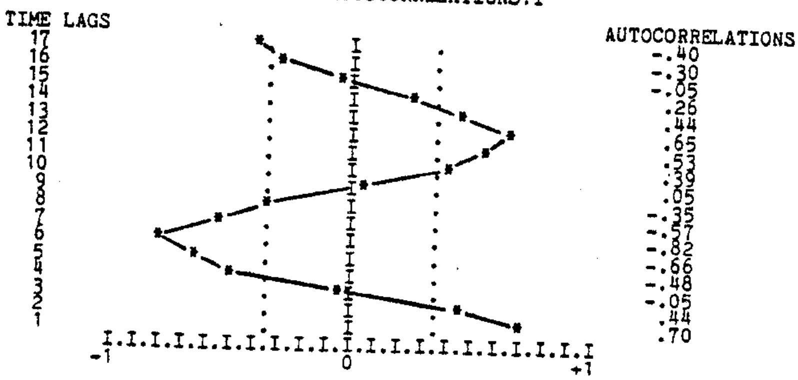

```
A STUDY OF THE AUTOCORRELATION COEFFICIENTS TELLS ME THAT THERE IS
SOME PATTERN IN YOUR DATA,I.E. THEY ARE NOT RANDOMLY DISTRIBUTED
AROUND THEIR MEAN (THIS IS BECAUSE    8 VALUES LIE OUTSIDE THE
CONTROL LIMITS--DOTTED LINES)  SEE PP11-19
```

รูปที่ 5.12 แสดงกราฟของ autocorrelation coefficient

จากรูปที่ 5.12 เมื่อพิจารณา $\bar{r} = 0.277$, $\hat{\sigma}_r = 0.167$, $\bar{r}_1 = 0.604$, $\bar{r}_2 = -0.051$, c.v. = 0.548 และ $\chi^2_c = 138.4 > \chi^2_{17,0.95} = 8.7$ ดังนั้นข้อมูลชุดนี้มี trend pattern อย่างแน่นอน เพราะค่า r ค่อย ๆ ไหลลดลง แต่อาจจะมี seasonal pattern ต้องกำจัด trend ออกจากข้อมูลจึงจะวิเคราะห์ผลได้

```
MEAN AUTOCORRELATION              -.014     STANDARD ERROR=              .169
MEAN OF FIRST   8   VALUES=       -.224     CHI SQUARE (COMPUTED)=      134.5
MEAN OF LAST    9   VALUES=        .173     CHI SQUARE (FROM TABLE)=      8.0
COEFFICIENT OF VARIATION =         .036

DO YOU WANT A GRAPH OF THE AUTOCORRELATIONS?Y
```

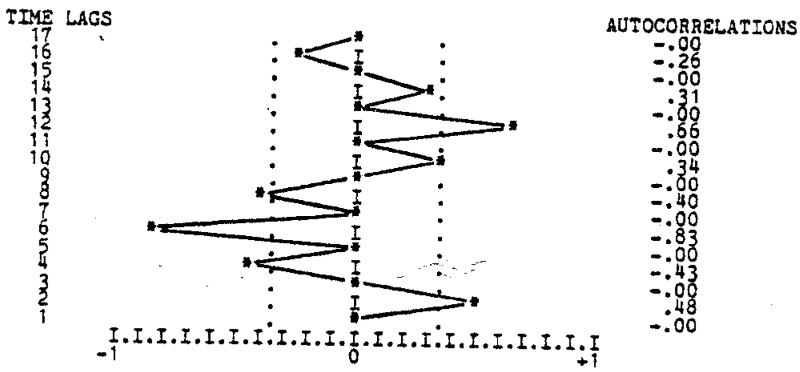


รูปที่ 5.13 แสดงกราฟของ autocorrelation หลังจากทำ first difference

จากรูปที่ 5.13 พิจารณาค่า $\bar{r} = -0.014$, $\hat{\sigma}_r = 0.169$, $\bar{r}_1 = -0.224$, $\bar{r}_2 = 0.173$, c.v. = 0.036 และ $\chi^2_c = 134.5 > \chi^2_{16,0.95} = 8.0$ ข้อมูลหลังจากทำ first difference ลดความผันแปรลงได้มาก แสดงว่า trend pattern ถูกกำจัดออกจากข้อมูล แต่กราฟยังมีจุดตกนอกเขตควบคุม (control limits หรือ confident band) หลายจุดอีกทั้งกราฟยังสะบัดไปมา แสดงว่าข้อมูล X'_t ยังมี pattern บางอย่างอยู่ ให้ทำ second difference ดังนี้

```
MEAN AUTOCORRELATION              -.007     STANDARD ERROR=              .171
MEAN OF FIRST   8   VALUES=       -.147     CHI SQUARE (COMPUTED)=       67.3
MEAN OF LAST    9   VALUES=        .117     CHI SQUARE (FROM TABLE)=      8.0
COEFFICIENT OF VARIATION =         .027

DO YOU WANT A GRAPH OF THE AUTOCORRELATIONS?Y
```




```
A STUDY OF THE AUTOCORRELATION COEFFICIENTS TELLS ME THAT THERE IS
SOME PATTERN IN YOUR DATA,I.E. THEY ARE NOT RANDOMLY DISTRIBUTED
AROUND THEIR MEAN (THIS IS BECAUSE    6 VALUES LIE OUTSIDE THE
CONTROL LIMITS--DOTTED LINES)  SEE PP11-19
```

```
YOUR DATA ARE STATIONARY AT THEIR       1ST DIFFERENCE

THEREFORE I AM LOOKING AT TABLE       2 TO DETERMINE THE LENGTH
OF SEASONALITY,IF ANY EXISTS

AUTOCORRELATIONS SIGNIFICANTLY DIFFERENT THAN ZERO  (ORDERED)
TIME LAG            AUTOCORRELATIONS
    6                    .82
    1                    .70
    5                    .66
   12                    .65
    7                    .57
   11                    .53
    4                    .48
   13                    .44
    2                    .44
   17                    .40
   10                    .39
    8                    .35

I CONSIDER THE LENGTH OF YOUR SEASONALITY IS =       6
```

รูปที่ 5.14 แสดงกราฟของ autocorrelation หลังจากทำ second difference

จากรูปที่ 5.14 พิจารณาค่า $\bar{r} = -0.007$, $\hat{\sigma}_r = 0.171$, $\bar{r}_1 = -0.147$, $\bar{r}_2 = 0.177$, c.v. = 0.027 และ $\chi^2_c = 67.3 > \chi^2_{16,0.95} = 8.0$ ค่า c.v. ของ X_t มากกว่า c.v. ของ X_t' และ c.v. ของ X_t' มากกว่า c.v. ของ X_t'' นอกจากนั้นยังมีจุดตกนอกเขตควบคุม 6 จุด ดังนั้นข้อมูล X_t'' ยังมี pattern อยู่ ซึ่งก็คือ seasonal pattern โดยมีช่วงกว้าง 6 วาระ

5.2 การวิเคราะห์ตัวแบบของวิธี Box - Jenkins

ข้อมูลอนุกรมเวลาชุดหนึ่ง ๆ สามารถเขียนอยู่ในรูปของ linear combine ของ random error คือ $a_t, a_{t-1}, a_{t-2}, \ldots$ โดย $\psi_1, \psi_2, \psi_3, \ldots$ เป็นน้ำหนักที่ใช้ถ่วง $a_t, a_{t-1}, a_{t-2}, \ldots$ โดยค่าน้ำหนักที่ว่านี้อาจเป็นลำดับจำกัดหรือไม่ก็ได้ แต่ถ้าเป็นลำดับอนันต์ที่สามารถ converge เข้าหาค่าคงที่ค่าใดค่าหนึ่งได้จะเรียกสมการต่อไปนี้ว่า stationary stochastic process โดยค่า μ จะแสดงค่าระดับกลางซึ่งสามารถใช้เป็นค่าอ้างอิงของ X_t ได้ แต่ถ้าข้อมูลอนุกรมเวลาชุดนั้นเป็น nonstationary stochastic process ค่า μ จะเคลื่อนไหวขึ้นลงเสมอ สมการ คือ

$$X_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \ldots$$

กรณีตัวแบบที่ใช้เป็น stationary stochastic process มีที่ใช้กันอยู่ 3 แบบ และในแบบสุดท้ายเป็นสภาพกรณีที่ระดับไม่คงที่มีความผันแปรในข้อมูล ซึ่งเป็นแบบที่ดำเนินการให้ข้อมูลเข้าสู่สภาพ stationary process ตัวแบบทั้งหมด คือ

1. AR (Autoregressive Model)
2. MA (Moving Average Model)
3. ARMA (Mixed Autoregressive Moving Average Model)
4. ARIMA (Autoregressive Integrated Moving Average Model)

ทั้งเทคนิคการพยากรณ์วิธี Adaptive Filtering และ Box - Jenkins การพยากรณ์ต้องใช้ตัวแบบอย่างใดอย่างหนึ่งใน 4 แบบที่กล่าวมาแล้ว

5.2.1 Autoregressive Model (AR)

เป็นตัวแบบที่กำหนดตัวแปรเชิงสุ่ม X_t เขียนอยู่ในรูป linear combination ของข้อมูลอนุกรมเวลาที่ time lag เป็น t-1 , t-2 , . . . , t-p กับ error a_t (หรือใช้คำว่า shock a_t) โดยที่ a_t มีการแจกแจงแบบปกติมี mean เป็น 0 มี variance เป็น σ_a^2 และมี covariance เป็น 0

[$a_t \sim N(0, \sigma_a^2)$ และ $E(a_t a_{t-k}) = 0$] มีสมการ คือ

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \Phi_3 X_{t-3} + \ldots + \Phi_p X_{t-p} + a_t \quad ....(5.1)$$

ซึ่งเป็นสมการที่มี $\Phi_1, \Phi_2, \Phi_3, \ldots, \Phi_p$ เป็น autoregressive parameter และเรียกสมการ นี้ว่า autoregressive process of order p หรือ AR (p) จากสมการ 5.1 คูณด้วย X_{t-k} จากนั้น take expectation ได้สมการคือ

$$E(X_{t-k}X_t) = \Phi_1 E(X_{t-k}X_{t-1}) + \Phi_2 E(X_{t-k}X_{t-2}) + \Phi_3 E(X_{t-k}X_{t-3}) + \ldots +$$
$$\Phi_p E(X_{t-k}X_{t-p}) + E(X_{t-k}a_t) \quad ; \; k > 0 \quad ....(5.2)$$

ได้ $$\gamma_k = \Phi_1 \gamma_{k-1} + \Phi_2 \gamma_{k-2} + \ldots + \Phi_p \gamma_{k-p} \quad ....(5.3)$$

เมื่อ $E(X_{t-k}a_t) = E[(a_{t-k} + \psi_1 a_{t-k-1} + \psi_2 a_{t-k-2} + \ldots) a_t] = 0$

เรียก γ_k ว่าเป็น autocovariance ของตัวแปร 2 ตัว แต่ต่างกาลกัน k วาระ จาก

$E(X_{t-k}X_t) = E(X_{t-0}X_t) = E(X_t^2) = \sigma_X^2 = \gamma_0$ เมื่อ $k = 0$

จากสมการ 5.2 นำ $\sigma_X^2 = \gamma_0$ หารตลอด จะได้

$$\rho_k = \Phi_1 \rho_{k-1} + \Phi_2 \rho_{k-2} + \ldots + \Phi_p \rho_{k-p} \quad ....(5.4)$$

เรียกสมการ 5.3 ว่า autocovariance function

สมการ 5.4 ว่า autocorrelation function

ในสมการ 5.4 ถ้าให้ k เป็นกาลที่ lag ไปตั้งแต่ 1 ถึง p จะได้ดังนี้

$k = 1 \quad \rho_1 = \Phi_1 + \Phi_2 \rho_1 + \Phi_3 \rho_2 + \ldots + \Phi_p \rho_{p-1}$

$k = 2 \quad \rho_2 = \Phi_1 \rho_1 + \Phi_2 + \Phi_3 \rho_1 + \ldots + \Phi_p \rho_{p-2}$

.

.

.

$k = p \quad \rho_p = \Phi_1 \rho_{p-1} + \Phi_2 \rho_{p-2} + \Phi_3 \rho_{p-3} + \ldots + \Phi_p$

หรือสามารถเขียนในรูปของเมตริกซ์ได้ดังนี้

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & . & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & . & . & . & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & . & . & . & \rho_{p-3} \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ \rho_{p-1} & & & & & & \rho_p \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ . \\ \phi_p \end{bmatrix}$$

$$\underline{\rho_p} = \underline{P_p} \quad \underline{\phi_p}$$

โดยเมตริกซ์ P_p จะต้องเป็น positive definite จึงจะควบคุมให้เงื่อนไข nonstationary เป็นจริงเสมอ ค่าของ $\rho_1, \rho_2, \ldots$ เป็นดังนี้

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0 \text{ ได้ } 1 - \rho_1^2 > 0 \text{ คือ } -1 < \rho_1 < 1$$

$$\begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix} > 0 \text{ ได้ } 1 - \rho_2^2 > 0 \text{ คือ } -1 < \rho_2 < 1$$

$$\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix} > 0 \text{ ได้ } 1 + 2\rho_1^2\rho_2 - \rho_2^2 - 2\rho_1^2 > 0 \text{ คือ}$$

$$-1 < [(\rho_2 - \rho_1^2)/(1 - \rho_1^2)] < 1$$

จากสมการ 5.2 แทนค่า $k = 0$ จะได้

$$\sigma_X^2 = \gamma_0 = \Phi_1\gamma_{-1} + \Phi_2\gamma_{-2} + \ldots + \Phi_p\gamma_{-p} + E(X_t a_t)$$

หรือ $\sigma_X^2 = \gamma_0 = \Phi_1\gamma_1 + \Phi_2\gamma_2 + \ldots + \Phi_p\gamma_p + E(X_t a_t)$..............(5.5)

เมื่อ $\gamma_{-1} = \gamma_1$ หรือ $\gamma_{-2} = \gamma_2$ เมื่อค่าเป็นลบ หมายถึงกาล lag ไปในอดีต และค่าเป็นบวกจะหมายถึงกาล lead ไปในอนาคต

แต่จากสมการ 5.5 ได้ $E(X_t a_t) = E[(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots) a_t]$

$$E(X_t a_t) = E(a_t^2) = \sigma_a^2$$

ดังนั้น $\gamma_0 = \Phi_1\gamma_1 + \Phi_2\gamma_2 + \ldots + \Phi_p\gamma_p + \sigma_a^2 = \sigma_X^2$

หารด้วย γ_0 ตลอดได้ $1 = \Phi_1\rho_1 + \Phi_2\rho_2 + \ldots + \Phi_p\rho_p + (\sigma_a^2/\sigma_X^2)$

$\sigma_X^2 = \sigma_a^2 \;/\; (1 - \Phi_1\rho_1 - \Phi_2\rho_2 - \ldots - \Phi_p\rho_p)$

ก. AR (1) หรือ first order autoregressive model

ตัวแบบ คือ $X_t = \Phi_1 X_{t-1} + a_t$ ....(5.6)

จากสมการ (5.6) ได้ $E(X_{t-k}X_t) = \Phi_1 E(X_{t-k}X_{t-1}) + E(X_{t-k}a_t)$...........(5.7)

$\gamma_k = \Phi_1\gamma_{k-1}$ เพราะ $E(X_{t-k}a_t) = 0$

นำ γ_0 หารตลอดได้ $\rho_k = \Phi_1\rho_{k-1}$ เมื่อ $k > 0$

แทนค่า $k = 1,2,3,\ldots$ เมื่อ $\rho_0 = 1$ จะได้

$$\rho_1 = \Phi_1\rho_0 = \Phi_1$$

$$\rho_2 = \Phi_1\rho_1 = \Phi_1^2$$

$$\rho_3 = \Phi_1\rho_2 = \Phi_1^3$$

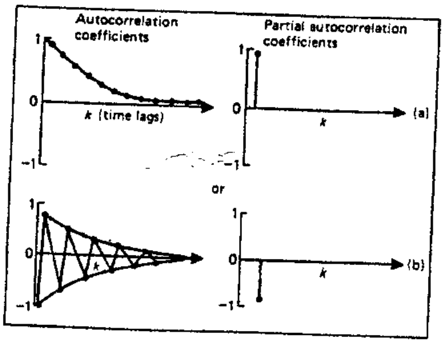
.

.

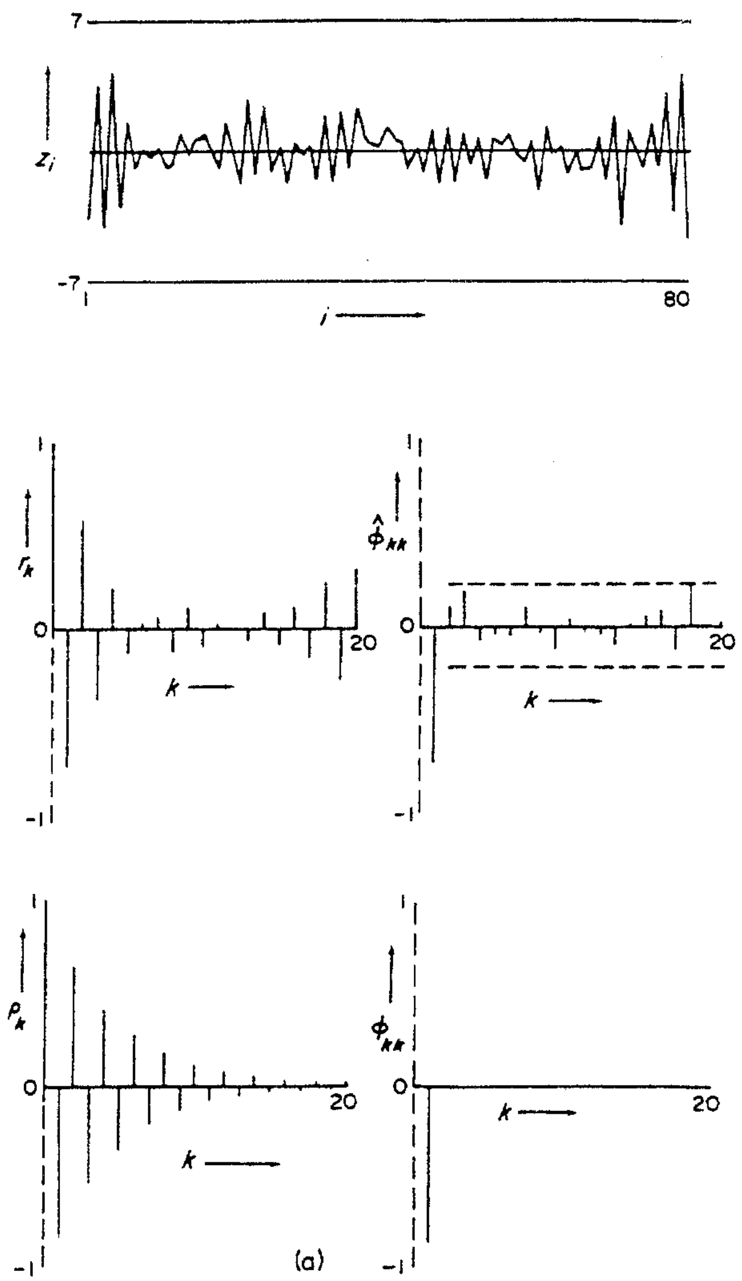
.

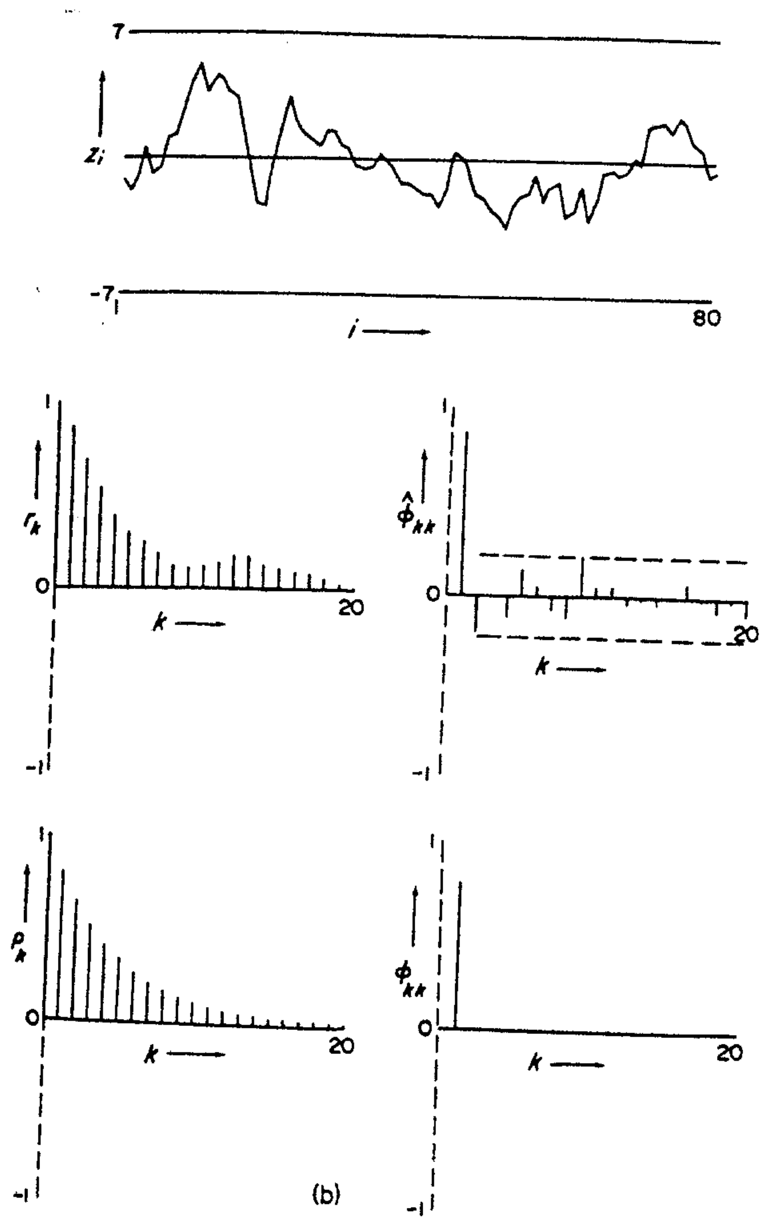
$$\rho_k = \Phi_1\rho_{k-1} = \Phi_1^k \quad \text{เมื่อ } k > 0 \text{ และ } -1 < \Phi_1 < 1$$

ซึ่งสมการ $\rho_k = \Phi_1^k$ เรียกว่า autocorrelation function ของ AR (1) เมื่อแทนค่า Φ_1 ด้วย r_1 จะได้ $\rho_0 = 1$, $\rho_1 = r_1$, $\rho_2 = r_1^2$ และเมื่อนำคู่ลำดับ (k, ρ_k) มา plot กราฟโดยค่า k เป็นแกนนอน ρ_k เป็นแกนตั้ง จะได้กราฟของ AR (1) เมื่อ r_1 เป็นบวก กราฟของ ρ_k จะมีลักษณะแบบ exponential ถ้า เมื่อ r_1 เป็นลบ กราฟของ ρ_k จะสะบัดรอบแกนนอน ขณะที่กราฟของ partial autocorrelation function เป็น bar chart เดียว ดังรูป



รูปที่ 5.15 แสดงกราฟของ AR (1)


7
-7
z_i
i
1
80
r_k
$\hat{\phi}_{kk}$
ρ_k
ϕ_{kk}
0
-1
1
20
k
(a)



Simulated AR(1) processes (of length 80)
(a) $z_i = -0{\cdot}8z_{i-1} + a_i$: *(b)* $z_i = 0{\cdot}8z_{i-1} + a_i$

รูปที่ 5.16 แสดงกราฟของ AR (1) กราฟของ autocorrelation และ partial autocorrelation

จากสมการ 5.7 เมื่อ $k = 0$ จะได้

$$E(X_t^2) = \Phi_1 E(X_t X_{t-1}) + E(X_t a_t)$$

และ $E(X_t a_t) = E[(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots) a_t] = \sigma_a^2$

ได้ $\sigma_X^2 = \Phi_1 \gamma_1 + \sigma_a^2$

นำ $\gamma_0 = \sigma_X^2$ หารตลอด ได้ $1 = \Phi_1 \rho_1 + (\sigma_a^2 / \sigma_X^2)$

$$\sigma_X^2 = \sigma_a^2 / (1 - \Phi_1 \rho_1) = \sigma_a^2 / (1 - \Phi_1^2)$$

หรือแทนค่า Φ_1 ด้วย r_1 จะได้

$$\sigma_X^2 = \sigma_a^2 / (1 - r_1^2)$$

ข. AR (2) หรือ second order autoregressive model

ตัวแบบ คือ $X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + a_t$

$$E(X_{t-k} X_t) = \Phi_1 E(X_{t-k} X_{t-1}) + \Phi_2 E(X_{t-k} X_{t-2}) + E(X_{t-k} a_t) \qquad \text{..........(5.8)}$$

$$\gamma_k = \Phi_1 \gamma_{k-1} + \Phi_2 \gamma_{k-2} \qquad \text{......................................(5.9)}$$

เมื่อ $E(X_{t-k} a_t) = E[(a_{t-k} + \psi_1 a_{t-k-1} + \psi_2 a_{t-k-2} + \ldots) a_t] = 0$

สมการ (5.9) นำ γ_0 หารตลอดได้ $\rho_k = \Phi_1 \rho_{k-1} + \Phi_2 \rho_{k-2}$

แทนค่า $k = 1,2$ และ $\rho_0 = 1$ จะได้

$\rho_1 = \Phi_1 + \Phi_2 \rho_1$ จะได้ $\rho_1 = \Phi_1 / (1 - \Phi_2)$

$\rho_2 = \Phi_1 \rho_1 + \Phi_2$

ได้ $\rho_2 = [\Phi_1^2 / (1 - \Phi_2)] + \Phi_2$

จากสมการ (5.8) เมื่อ $k = 0$ จะได้

$$\sigma_x^2 = \Phi_1 \gamma_1 + \Phi_2 \gamma_2 + E(X_t a_t)$$

แต่ $E(X_t a_t) = E[(a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots) a_t] = E(a_t a_t) = \sigma_a^2$

$$\sigma_x^2 = \Phi_1 \gamma_1 + \Phi_2 \gamma_2 + \sigma_a^2$$

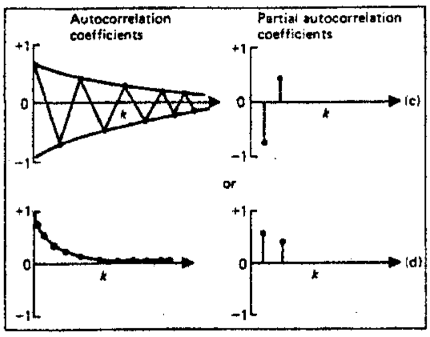
$$1 = \Phi_1 \rho_1 + \Phi_2 \rho_2 + (\sigma_x^2 / \sigma_a^2)$$

$$\sigma_x^2 = \sigma_a^2 / (1 - \Phi_1 \rho_1 - \Phi_2 \rho_2) \qquad \text{...........................(5.10)}$$

แทนค่า $\rho_1 = \Phi_1 / (1 - \Phi_2)$, $\rho_2 = \Phi_2 + [\Phi_1^2 / (1 - \Phi_2)]$ ลงในสมการ (5.10)ได้

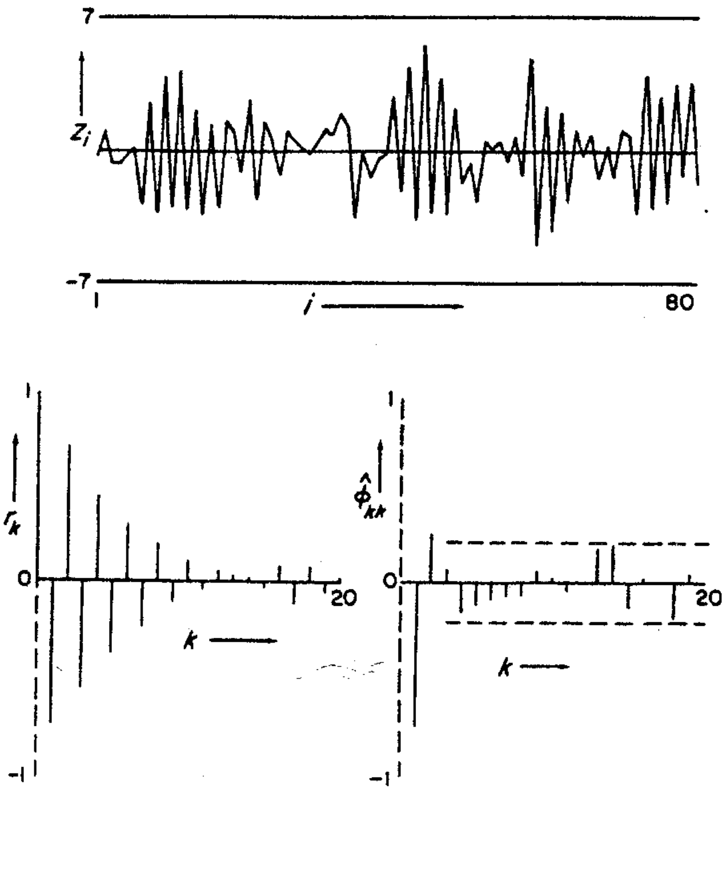
$$\sigma_x^2 = [(1 - \Phi_2)/(1 + \Phi_2)] [\sigma_a^2 / \{(1 - \Phi_2)^2 - \Phi_1^2\}]$$

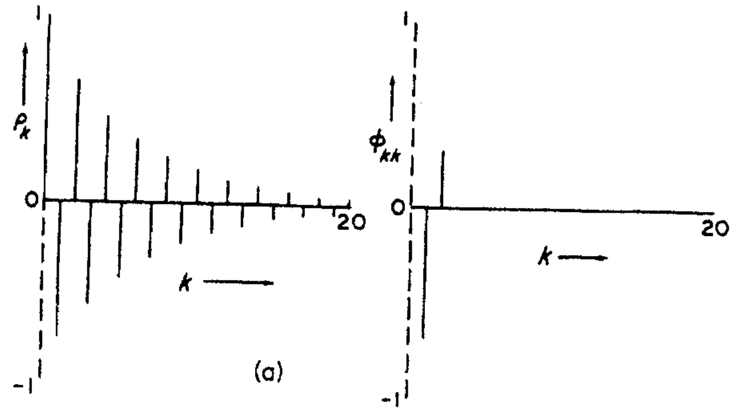
ถ้านำคู่ลำดับ (k , ρ_k) ไป plot กราฟของ autocorrelation function และสามารถ plot กราฟของ partial autocorrelation function ได้ดังนี้



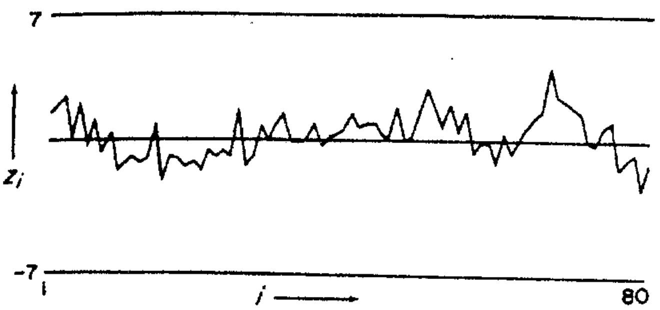
AR (2) Model

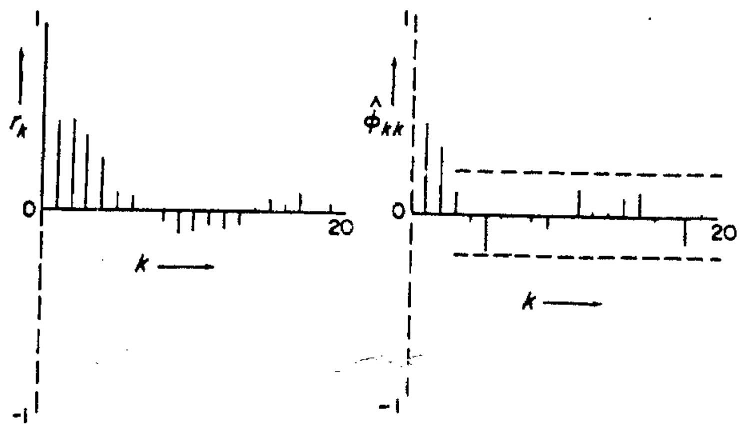
รูปที่ 5.17 แสดงกราฟของ AR (2)

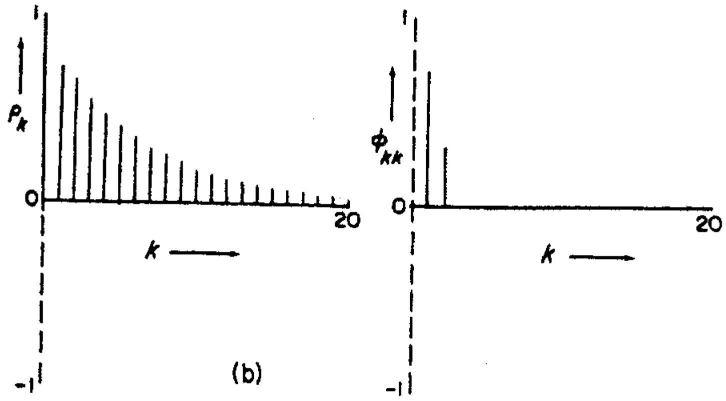



1
ρ_k
0
20
k
-1
1
ϕ_{kk}
0
20
k
-1

(a)


7
z_i
-7
1
i
80


1
r_k
0
20
k
-1
1
$\hat{\phi}_{kk}$
0
20
k
-1



Simulated AR(2) processes
(a) $z_i = -0{\cdot}5z_{i-1} + 0{\cdot}3z_{i-2} + a_i$; (b) $z_i = 0{\cdot}5z_{i-1} + 0{\cdot}3z_{i-2} + a_i$
((c) and (d) overleaf)

รูปที่ 5.18 แสดงกราฟของ autocorrelation และ partial autocorrelation ของ AR (2)

ค. AR (p) หรือ p order autoregressive model

มีตัวแบบ คือ $X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t$

$$E(X_{t-k}X_t) = \Phi_1 E(X_{t-k}X_{t-1}) + \Phi_2 E(X_{t-k}X_{t-2}) + \ldots + \Phi_p E(X_{t-k}X_{t-p}) + E(X_{t-k}a_t)$$

ได้ $\gamma_k = \Phi_1\gamma_{k-1} + \Phi_2\gamma_{k-2} + \ldots + \Phi_p\gamma_{k-p} \quad ; \quad k \geq 0$

นำ γ_0 หารตลอดได้ $\rho_k = \Phi_1\rho_{k-1} + \Phi_2\rho_{k-2} + \ldots + \Phi_p\rho_{k-p}$

เป็น autocorrelation function ซึ่งโดยปกติแล้ว AR จะกล่าวถึงตัวแบบเพียง AR (1) , AR (2)และ AR (3) ซึ่ง AR (1) จะมี $\rho_k = r_1^2$ โดย k จะมีค่าน้อยกว่าหรือเท่ากับ n / 4 ซึ่ง AR (p) จะหา autocorrelation function ได้ คือ $\rho_0, \rho_1, \rho_2, \ldots, \rho_{n/4}$ เมื่อ $r_k \sim N(0, 1/n)$ และ $\Phi_k \sim N(0, 1/n)$ ดังนั้น 95 % confident band คือ $[-1.96/\sqrt{n}\ ,\ 1.96/\sqrt{n}]$

5.2.2 Moving Average Model (MA)

ก. MA (1) หรือ first order moving average model

ตัวแบบ คือ $X_t = a_t - \theta_1 a_{t-1}$

จากตัวแบบสามารถ derive หา autocorrelation function ได้ดังนี้

$$E(X_t X_t) = E[(a_t - \theta_1 a_{t-1})(a_t - \theta_1 a_{t-1})]$$

$$\gamma_0 = E(a_t^2) - 2\theta_1 E(a_t a_{t-1}) + \theta_1^2 E(a_{t-1}^2)$$

$$\gamma_0 = \sigma_a^2 + \theta_1^2\sigma_a^2 = \sigma_a^2(1+\theta_1^2) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(5.11)$$

$E(X_{t-1}X_t) = E[(a_{t-1} - \theta_1 a_{t-2})(a_t - \theta_1 a_{t-1})]$

$E(X_{t-1}X_t) = E(a_{t-1}a_t) - \theta_1 E(a_{t-1}^2) - \theta_1 E(a_{t-2}a_t) + \theta_1^2 E(a_{t-1}a_{t-2})$

$$\gamma_1 = -\theta_1 \sigma_a^2 \qquad \text{.....................(5.12)}$$

นำ (5.12) / (5.11) ได้

$$\rho_1 = -\theta_1 / (1+\theta_1^2)$$

ซึ่ง $\rho_0 = 1$ และ $\rho_k = 0$ เมื่อ $k \geq 2$

ใน MA (1) สามารถเขียน X_t อยู่ในรูป linear combination ของข้อมูลที่ lag ไปหลาย ๆ วาระได้ หรือ MA (1) มี infinite lagged dependent variable คือ

$$X_t = a_t - \theta_1 (X_{t-1} + \theta_1 a_{t-2}) \quad \text{เมื่อ} \quad a_{t-1} = X_{t-1} + \theta_1 a_{t-2}$$

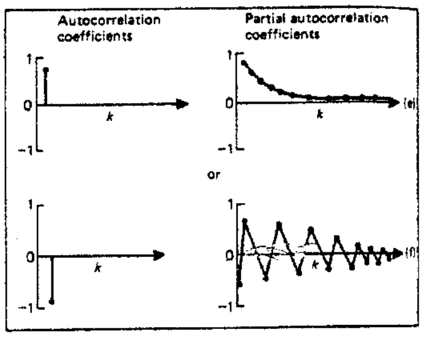
$$X_t = a_t - \theta_1 X_{t-1} - \theta_1^2 a_{t-2}$$

$$X_t = a_t - \theta_1 X_{t-1} - \theta_1^2 (X_{t-2} + \theta_1 a_{t-3}) \quad \text{เมื่อ} \quad a_{t-2} = X_{t-2} + \theta_1 a_{t-3}$$

$$X_t = a_t - \theta_1 X_{t-1} - \theta_1^2 X_{t-2} + \theta_1^3 a_{t-3}$$

หรือ $X_t = a_t - \theta_1 X_{t-1} - \theta_1^2 X_{t-2} - \theta_1^3 X_{t-3} - \ldots$

แต่ใน AR (1) สามารถได้เป็น finite lagged dependent variable คือ $X_t = \Phi_1 X_{t-1} + a_t$ แต่มี autocorrelation function ที่ infinite คือ $\rho_k = \Phi_1^k = r_1^k$ เมื่อ $k \geq 0$ ซึ่งใน MA (1) จะหา autocorrelation function ได้จำนวนไม่เกิน จำนวน order ของ MA คือ MA (1) จะได้ $\rho_0 = 1$, $\rho_1 = -\theta_1 / (1+\theta_1^2)$ เมื่อเขียนกราฟของ autocorrelation coefficient จะเป็น bar chart แท่งเดียว และกราฟของ partial autocorrelation coefficient จะเป็นกราฟลักษณะ exponential หรือสลับฟันปลา ดังรูป



MA (1) Model

รูปที่ 5.19 แสดงกราฟ acf และ pacf ของ MA (1)

ข. MA (2) หรือ second order moving average model

ตัวแบบ คือ $X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

จากตัวแบบสามารถ derive หา autocorrelation function ได้ดังนี้

$$E(X_t X_t) = E[(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})]$$

$$\gamma_0 = E(a_t^2) - 2\theta_1 E(a_t a_{t-1}) + -2\theta_2 E(a_t a_{t-2}) + \theta_1^2 E(a_{t-1}^2) + 2\theta_1\theta_2 E(a_{t-1} a_{t-2}) + \theta_2^2 E(a_{t-2}^2)$$

$$\gamma_0 = \sigma_a^2 + \theta_1^2\sigma_a^2 + \theta_2^2\sigma_a^2 = \sigma_a^2(1+\theta_1^2+\theta_2^2)$$

$$E(X_{t-1} X_t) = E[(a_{t-1} - \theta_1 a_{t-2} - \theta_2 a_{t-3})(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})]$$

$$E(X_{t-1} X_t) = E(a_{t-1} a_t) - \theta_1 E(a_{t-1}^2) - \theta_2 E(a_{t-1} a_{t-2}) - \theta_1 E(a_t a_{t-2}) + \theta_1^2 E(a_{t-1} a_{t-2}) + \theta_1\theta_2 E(a_{t-2}^2) - \theta_2 E(a_{t-3} a_t) + \theta_1\theta_2 E(a_{t-1} a_{t-3}) + \theta_2^2 E(a_{t-3} a_{t-2})$$

$$\gamma_1 = -\theta_1\sigma_a^2 + \theta_1\theta_2\sigma_a^2 = \sigma_a^2(-\theta_1 + \theta_1\theta_2)$$

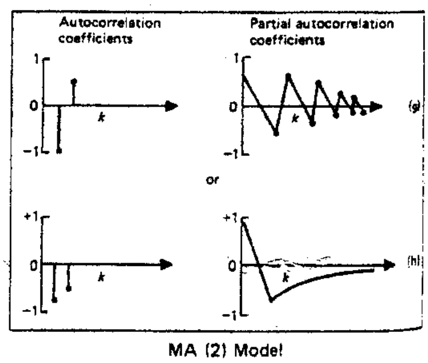
$$E(X_{t-2} X_t) = E[(a_{t-2} - \theta_1 a_{t-3} - \theta_2 a_{t-4})(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2})]$$

$$\gamma_2 = -\theta_2\sigma_a^2$$

ได้ $\rho_1 = (-\theta_1 + \theta_1\theta_2)/(1+\theta_1^2+\theta_2^2)$

$\rho_2 = (-\theta_2)/(1+\theta_1^2+\theta_2^2)$

เมื่อ plot กราฟของ autocorrelation coefficient จะได้เป็น bar chart สองแท่ง และกราฟของ partial correlation coefficient จะเป็นรูปสลับฟันปลาขึ้นลง ดังรูป



รูปที่ 5.20 แสดงกราฟ acf และ pacf ของ MA (2)

ก. MA (q) หรือ q order moving average model

ตัวแบบ คือ $X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$

จากตัวแบบสามารถ derive หา autocorrelation function ได้ดังนี้

$E(X_t X_t) = E[(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q})(a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q})]$

$\gamma_0 = E(a_t^2) + \theta_1^2 E(a_{t-1}^2) + \theta_2^2 E(a_{t-2}^2) + \ldots + \theta_q^2 E(a_{t-q}^2)$

$\gamma_0 = \sigma_a^2 (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)$

ได้ autocorrelation function คือ

$\rho_k = (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \ldots + \theta_{q-k} \theta_q) / (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)$

เมื่อ $k = 1,2,\ldots,q$

5.2.3 Mixed Autoregressive Moving Average Model (ARMA)

ก. ARMA (1,1) หรือ first order autoregressive - first order moving average model

ตัวแบบ คือ $X_t = \Phi_1 X_{t-1} + a_t - \theta_1 a_{t-1}$

$E(X_t X_t) = E[X_t(\Phi_1 X_{t-1} + a_t - \theta_1 a_{t-1})]$

$\gamma_0 = \Phi_1 \gamma_1 + E(X_t a_t) - \theta_1 E(X_t a_{t-1})$

เมื่อ $E(X_t a_t) = E[(\Phi_1 X_{t-1} + a_t - \theta_1 a_{t-1}) a_t] = \sigma_a^2$

$E(X_t a_{t-1}) = E[(\Phi_1 X_{t-1} + a_t - \theta_1 a_{t-1}) a_{t-1}]$

$E(X_t a_{t-1}) = E[(\Phi_1(\Phi_1 X_{t-2} + a_{t-1} - \theta_1 a_{t-2}) + a_t - \theta_1 a_{t-1}) a_{t-1}]$

$E(X_t a_{t-1}) = \Phi_1 \sigma_a^2 - \theta_1 \sigma_a^2 = (\Phi_1 - \theta_1)\sigma_a^2$

ได้ $\gamma_0 = \Phi_1 \gamma_1 + \sigma_a^2 - \theta_1(\Phi_1 \sigma_a^2 - \theta_1 \sigma_a^2)$

ได้ $\gamma_0 = \Phi_1 \gamma_1 + (1 - \theta_1 \Phi_1 + \theta_1^2)\sigma_a^2$ ....(5.13)

$E(X_{t-1} X_t) = E[X_{t-1}(\Phi_1 X_{t-1} + a_t - \theta_1 a_{t-1})]$

$\gamma_1 = \Phi_1 \gamma_0 - \theta_1 \sigma_a^2$ ....(5.14)

$E(X_{t-2} X_t) = E[X_{t-2}(\Phi_1 X_{t-1} + a_t - \theta_1 a_{t-1})]$

$\gamma_2 = \Phi_1 \gamma_1$ ....(5.15)

แก้สมการ (5.13) และ (5.14) จะได้

$\gamma_0 = [(1 - 2\theta_1 \Phi_1 + \theta_1^2)\sigma_a^2] / (1 - \theta_1^2)$

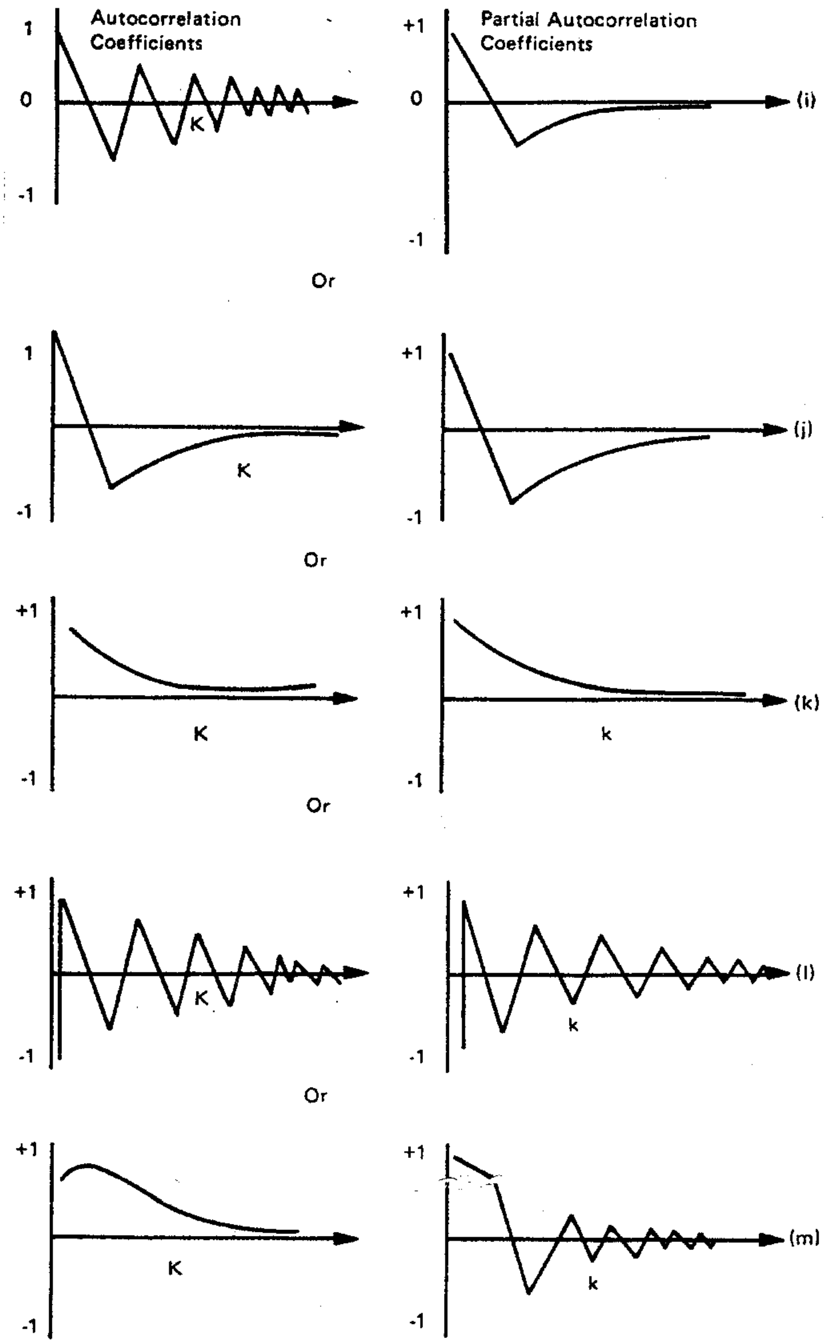
$\gamma_1 = [(\Phi_1 - \Phi_1^2 \theta_1 + \Phi_1 \theta_1^2 - \theta_1)\sigma_a^2] / (1 - \Phi_1^2)$

จะได้ $\rho_1 = \gamma_1 / \gamma_0 = [(1 - \theta_1 \Phi_1)(\Phi_1 - \theta_1)] / (1 - 2\theta_1 \Phi_1 + \theta_1^2)$

$\rho_2 = \gamma_2 / \gamma_0 = \Phi_1 \rho_1$

$\rho_3 = \gamma_3 / \gamma_0 = \Phi_1 \rho_2$ จนถึง $\rho_k = \gamma_k / \gamma_0 = \Phi_1 \rho_{k-1}$ เมื่อ $k \geq 2$

ประมาณค่า ρ_1 และ ρ_2 ด้วย r_1 และ r_2 ตามลำดับ เราย่อมหาค่าเริ่มต้นของ Φ_1 และ θ_1 ได้ ซึ่งเมื่อแทนค่า Φ_1 และ θ_1 ลงในตัวแบบก็จะได้สมการของ ARMA (1,1) ที่ต้องการและไป plot กราฟของ autocorrelation กับ partial autocorrelation coefficient ของ ARMA (1,1) จะได้ดังนี้



รูปที่ 5.21 แสดงกราฟ acf และ pacf ของ ARMA (1,1)

ในการ plot กราฟของ autocorrelation จะเริ่มจากค่า ρ_1 ถ้าค่า Φ_1 มีค่าเป็นบวก กราฟของ autocorrelation จะเรียบ แต่ถ้า Φ_1 เป็นลบ เส้นโค้งของ autocorrelation จะสะบัดรอบแกน เช่นเดียวกัน ถ้าค่า θ_1 มีค่าเป็นบวก กราฟของ partial autocorrelation จะเรียบ แต่ถ้า θ_1 เป็นลบ เส้นโค้งของกราฟ partial autocorrelation จะสะบัดรอบแกน โดยที่ ค่าของ Φ_1 และ θ_1 จะมีค่าอยู่ในช่วง (-1 , 1)

ข. ARMA (p , q)

ตัวแบบ คือ

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

$$E(X_{t-k} X_t) = E[X_{t-k}(\Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q})]$$

จะได้ autocovariance function คือ

$$\gamma_k = \Phi_1 \gamma_{k-1} + \Phi_2 \gamma_{k-2} + \ldots + \Phi_p \gamma_{k-p}$$

และได้ autocorrelation function ดังนี้

$$\rho_k = \Phi_1 \rho_{k-1} + \Phi_2 \rho_{k-2} + \ldots + \Phi_p \rho_{k-p}$$

ใน model ARMA (p , q) เมื่อ p = 0,1,2,... และ q = 0,1,2,... กรณีที่ p = 1 , q = 0 ตัวแบบ ก็คือ $X_t = \Phi_1 X_{t-1} + a_t$ หรือก็คือ ตัวแบบของ AR (1) หรือ ARMA (1,0) ถ้า p = 2 , q = 0 ก็คือ ตัวแบบของ AR (2) หรือ ARMA (2,0) คือ $X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + a_t$ เมื่อ p = 0 , q = 1 เขียนได้เป็น MA (1) หรือ ARMA (0,1) ตัวแบบคือ $X_t = a_t - \theta_1 a_{t-1}$ เมื่อ p = 0 , q = 2 เขียนได้เป็น MA (2) หรือ ARMA (0,2) ตัวแบบคือ $X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

5.2.4 Autoregressive Integrated Moving Average (ARIMA)

กระบวนการ ARIMA (p , d , q) เป็นกระบวนการของ ARMA (p , q) ที่ทำการ d^{th} difference โดยมีตัวแบบจากการ derived สมการของ

$$\Phi(B)(1-B)^d X_t = \theta(B) a_t$$

เมื่อ $\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p$

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

$(1-B)^d X_t$ เป็นการ difference ครั้งที่ d ของอนุกรมเวลา X_t

ในการวิเคราะห์กราฟของ autocorrelation coefficient เมื่อกาลเพิ่มขึ้น ลักษณะกราฟจะค่อย ๆ ลดลงช้า ๆ เราสามารถตรวจสอบโดยการ difference ถ้ากราฟของ autocorrelation coefficient มีลักษณะไหลลงอย่างรวดเร็วแล้ว แสดงว่าข้อมูลอนุกรมเวลาชุดนี้เป็น nonstationary เราสามารถหาตัวแบบของ ARIMA ได้ดังนี้

ARIMA (0,1,1) $\quad (1-B)X_t = (1-\theta_1 B)a_t$

$$X_t - X_{t-1} = a_t - \theta_1 a_{t-1}$$

$$X_t = X_{t-1} + a_t - \theta_1 a_{t-1}$$

ARIMA (1,1,0) $\quad (1-\Phi_1 B)(1-B)X_t = a_t$

$$(1-\Phi_1 B - B + \Phi_1 B^2)X_t = a_t$$

$$X_t - \Phi_1 X_{t-1} - X_{t-1} + \Phi_1 X_{t-2} = a_t$$

$$X_t = (\Phi_1 + 1)X_{t-1} - \Phi_1 X_{t-2} + a_t$$

ARIMA (0,2,2) $\quad (1-B)^2 X_t = (1-\theta_1 B)^2 a_t$

$$(1-2B+B^2)X_t = (1-2\theta_1 B + \theta_1^2 B^2)a_t$$

$$X_t - 2X_{t-1} + X_{t-2} = a_t - 2\theta_1 a_{t-1} + \theta_1^2 a_{t-2}$$

$$X_t = 2X_{t-1} - X_{t-2} + a_t - 2\theta_1 a_{t-1} + \theta_1^2 a_{t-2}$$

ARIMA (1,1,1) $\quad (1-\Phi_1 B)(1-B)X_t = (1-\theta_1 B)a_t$

$$(1-\Phi_1 B - B + \Phi_1 B^2)X_t = (1-\theta_1 B)a_t$$

$$X_t - \Phi_1 X_{t-1} - X_{t-1} + \Phi_1 X_{t-2} = a_t - \theta_1 a_{t-1}$$

$$X_t = (\Phi_1 + 1)X_{t-1} - \Phi_1 X_{t-2} + a_t - \theta_1 a_{t-1}$$

ARIMA (2,1,1) $\quad (1-\Phi_1 B)^2(1-B)X_t = (1-\theta_1 B)a_t$

$$(1-\Phi_1 B - \Phi_2 B^2)(1-B)X_t = (1-\theta_1 B)a_t$$

$$(1-\Phi_1 B - \Phi_2 B^2 - B + \Phi_1 B^2 + \Phi_2 B^3)X_t = a_t - \theta_1 a_{t-1}$$

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - X_{t-1} + \Phi_1 X_{t-2} + \Phi_2 X_{t-3} = a_t - \theta_1 a_{t-1}$$

$$X_t = (1+\Phi_1)X_{t-1} + (\Phi_2 - \Phi_1)X_{t-2} - \Phi_2 X_{t-3} + a_t - \theta_1 a_{t-1}$$

ARIMA (1,2,1) $\quad (1-\Phi_1 B)(1-B)^2 X_t = (1-\theta_1 B)a_t$

$$(1-2B+B^2-\Phi_1 B + 2\Phi_1 B^2 - \Phi_1 B^3)X_t = a_t - \theta_1 a_{t-1}$$

$$X_t - 2X_{t-1} + X_{t-2} - \Phi_1 X_{t-1} + 2\Phi_1 X_{t-2} - \Phi_1 X_{t-3} = a_t - \theta_1 a_{t-1}$$

$$X_t = (2+\Phi_1)X_{t-1} - (1+2\Phi_1)X_{t-2} + \Phi_1 X_{t-3} + a_t - \theta_1 a_{t-1}$$

ARIMA (1,1,2) $\quad (1-\Phi_1 B)(1-B)X_t = (1-\theta_1 B - \theta_2 B^2)a_t$

$$(1-\Phi_1 B - B + \Phi_1 B^2)X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

$$X_t - \Phi_1 X_{t-1} - X_{t-1} + \Phi_1 X_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

$$X_t = (1+\Phi_1)X_{t-1} - \Phi_1 X_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

ARIMA (2,2,2)

$$(1-\Phi_1 B - \Phi_2 B^2)(1-B)^2 X_t = (1-\theta_1 B - \theta_2 B^2)a_t$$

$$(1-\Phi_1B-\Phi_2B^2-2B+2\Phi_1B^2+2\Phi_2B^3+B^2-\Phi_1B^3-\Phi_2B^4)X_t = a_t-\theta_1a_{t-1}-\theta_2a_{t-2}$$

$$X_t=(2+\Phi_1)X_{t-1}+(\Phi_2-2\Phi_1-1)X_{t-2}+(\Phi_1-2\Phi_2)X_{t-3}+\Phi_2X_{t-4}+a_t-\theta_1a_{t-1}-\theta_2a_{t-2}$$

นอกจากนี้การเขียนตัวแบบดังกล่าวแล้ว เรายังสามารถนำรูปแบบของ ARIMA เสนอออกมาได้ 3 รูปแบบ คือ

ก. เสนอในรูปของ difference equation ดังได้แสดงให้แล้วข้างต้น เช่น ARIMA (0,1,1) มีสมการ

คือ $X_t = X_{t-1}+a_t-\theta_1a_{t-1}$

ข. เสนอในรูปของ linear filter form รูปแบบนี้ช่วยในการหาค่าความแปรปรวนของค่าพยากรณ์ได้ มีรูปแบบ คือ

$$X_t = a_t+\psi_1a_{t-1}+\psi_2a_{t-2}+\ldots$$

หรือ $X_t = (1+\psi_1B+\psi_2B^2+\ldots)a_t$

ค. เสนอในรูปของค่าข้อมูลในอดีตกับ shock ของปัจจุบัน เมื่อมีตัวถ่วงน้ำหนัก คือ $\Pi(B)$ มีรูปแบบคือ

$$1-\Pi_1X_{t-1}-\Pi_2X_{t-2}-\Pi_3X_{t-3}-\ldots = a_t$$

หรือ $(1-\Pi_1B-\Pi_2B^2-\Pi_3B^3-\ldots)X_t = a_t$

$$\Pi(B) = 1/(1+\psi_1B+\psi_2B^2+\ldots)$$

$$\Pi(B) = 1-\Pi_1B-\Pi_2B^2-\Pi_3B^3-\ldots$$

5.3 ตัวแบบในวิธี Adaptive filtering

5.3.1 AR (p)

ตัวแบบ คือ $X_t=\Phi_1X_{t-1}+\Phi_2X_{t-2}+\ldots+\Phi_pX_{t-p}+e_t$

เมื่อ e_t = ค่าความคลาดเคลื่อน ณ กาล t

Φ_i = ค่าถ่วงน้ำหนักที่ i หรือ $\Phi_i = w_i$

ได้ $e_t = X_t-\Phi_1X_{t-1}-\Phi_2X_{t-2}-\ldots-\Phi_pX_{t-p}$

$$e_t^2 = (X_t-\Phi_1X_{t-1}-\Phi_2X_{t-2}-\ldots-\Phi_pX_{t-p})^2$$

$$\frac{\partial e_t^2}{\partial \Phi_i} = 2e_t\frac{\partial e_t}{\partial \Phi_i}$$ เมื่อ $i=1,2,\ldots,p$

$$\frac{\partial e_t^2}{\partial \Phi_i} = \nabla e_t^2 = \text{gradient ของ } e_t^2 = 2e_t(-X_{t-i}) = -2e_tX_{t-i}$$

ได้ค่าถ่วงน้ำหนักที่ปรับใหม่ คือ $\Phi_i' = \Phi_i - k\nabla e_t^2$

$$\Phi_i' = \Phi_i + 2ke_tX_{t-i} \qquad \text{เมื่อ } i = 1, 2, \ldots, p$$

หรือ

$$w_i' = w_i + 2ke_tX_{t-i} \qquad t = p+1, p+2, \ldots, n$$

5.3.2 MA (q)

ตัวแบบ คือ $X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$

หรือ $X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}$

$$e_t = X_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

$$e_t^2 = (X_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q})^2$$

$$\frac{\partial e_t^2}{\partial \theta_i} = 2e_te_{t-i}$$

ค่าถ่วงน้ำหนักที่ปรับใหม่ คือ $\theta_i' = \theta_i - k\nabla e_t^2$

$$\theta_i' = \theta_i - 2ke_te_{t-i}$$

หรือ

$$w_i' = w_i - 2ke_te_{t-i}$$

5.3.3 ARMA (p,q)

ตัวแบบคือ $X_t = \Phi_1X_{t-1} + \Phi_2X_{t-2} + \ldots + \Phi_pX_{t-p} - e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}$

$$e_t = X_t - \Phi_1X_{t-1} - \Phi_2X_{t-2} - \ldots - \Phi_pX_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

$$\frac{\partial e_t^2}{\partial \Phi_i} = -2e_tX_{t-i} \quad \text{และ} \quad \frac{\partial e_t^2}{\partial \theta_i} = 2e_te_{t-i}$$

ได้ θ_i และ Φ_i ที่ปรับค่าใหม่คือ

$$\Phi_i' = \Phi_i + 2ke_tX_{t-i} \qquad \text{เมื่อ } i = 1,2,\ldots,p \ ; \ t = (p+1)(1)n$$

$$\theta_i' = \theta_i - 2ke_te_{t-i} \qquad i = 1,2,\ldots,q \ ; \ t = (q+1)(1)n$$

5.4 ขั้นตอนการดำเนินการในวิธีของ Box - Jenkins

วิธีการของ Box - Jenkins มีประสิทธิภาพและกระบวนการในทางปฏิบัติสำหรับใช้ autoregressive หรือ moving average เพื่อประยุกต์ในการพยากรณ์หลักการกระจายของ Box - Jenkins เป็นการพัฒนากระบวนการสำหรับแยกแยะตัวแบบ ARMA ที่เหมาะสมที่สุดของข้อมูลชุดหนึ่ง ๆ และสำหรับทดสอบความถูกต้องของตัวแบบ รูปที่ 5.22 แสดงขั้นตอนในกระบวนการของ Box - Jenkins ถ้าประยุกต์อย่างถูกต้อง ขั้นตอนที่ 3 จะให้ตัวแบบที่ถูกต้องสำหรับการพยากรณ์ ซึ่งขั้นตอนการดำเนินงานมีดังนี้

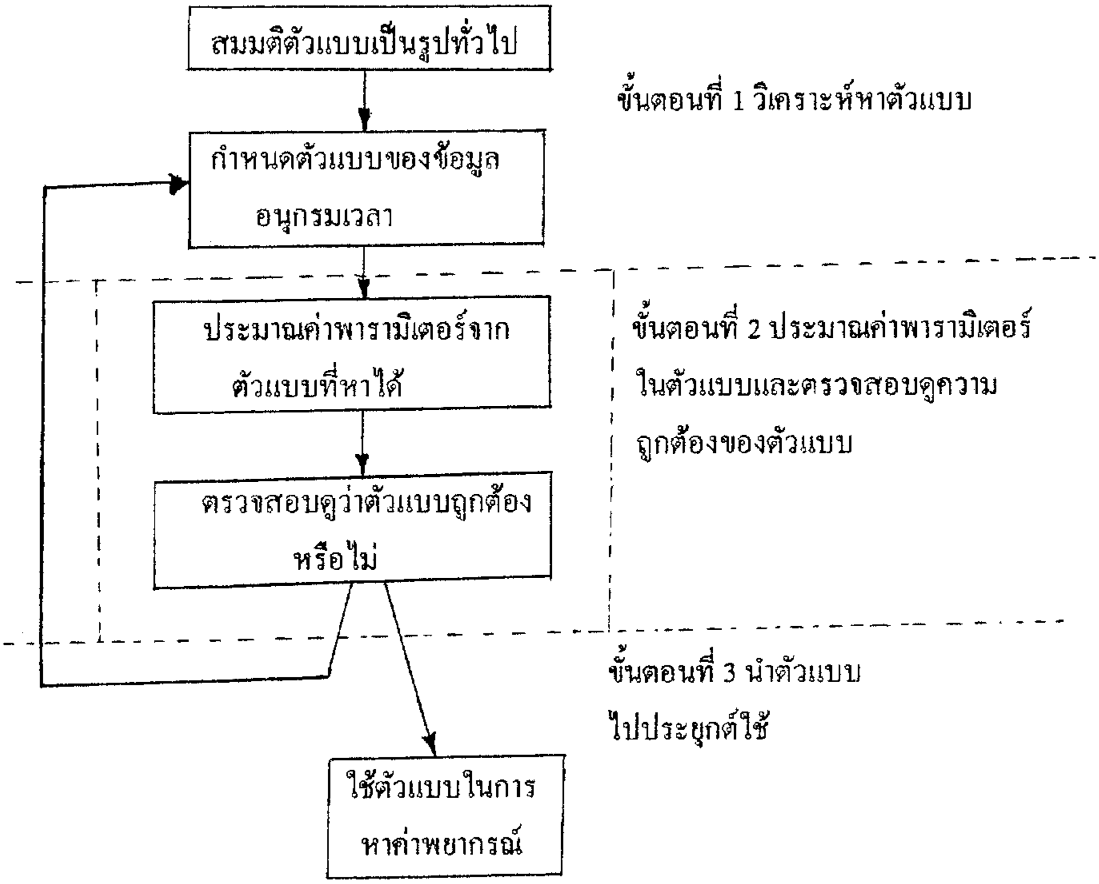
ขั้นตอนที่ 1 วิเคราะห์หาตัวแบบ

ก. สมมติตัวแบบในรูปทั่วไป

รูปทั่วไปของตัวแบบ ARMA คือ

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

จากสมการจะใช้ได้เหมาะสมสำหรับข้อมูลอนุกรมเวลาที่เป็น stationary งานของผู้ใช้คือต้องตัดสินใจสำหรับดีกรีของ p และ q ในตัวแบบ ARMA



รูปที่ 5.22 แสดงขั้นตอนการดำเนินการใน Box - Jenkins

ข. กำหนดตัวแบบให้ข้อมูลอนุกรมเวลา

ควรพิจารณาดูว่าข้อมูลเป็นลักษณะ stationary หรือ nonstationary ถ้าเป็น nonstationary ควรปรับให้เป็น stationary โดยการทำ d^{th} difference (ปกติ d = 0,1,2) เมื่อข้อมูลอนุกรมเวลาเป็น stationary จึงตรวจสอบดูว่าเป็นตัวแบบใด คือ AR (p) หรือ MA (q) หรือ ARMA (p , q) โดยใช้กราฟของ autocorrelation coefficient (acf) พร้อมทั้งกราฟของ partial autocorrelation coefficient (pacf) พิจารณาว่าเป็น ARMA (p,q) ที่ p , q order เท่าไร เมื่อได้ตัวแบบแล้วให้ดูช่วงความเชื่อมั่นของกราฟ autocorrelation และ partial autocorrelation ว่ามีกี่ตัว ที่มีค่าเกิน confident band ซึ่ง

การทดสอบสมมติฐานทางสถิติ คือ $H_0: \rho_k = 0 \quad H_1: \rho_k \neq 0$

เมื่อตรวจสอบนัยสำคัญทางสถิติได้จำนวนตัวที่อยู่นอกช่วงความเชื่อมั่น ก็จะเป็นจำนวน order ของตัวแบบนั้น

<u>ขั้นตอนที่ 2</u> ประมาณค่าพารามิเตอร์ในตัวแบบและตรวจสอบความถูกต้อง

ก. เมื่อได้ตัวแบบของข้อมูลอนุกรมเวลา เราต้องประมาณค่าพารามิเตอร์ โดยทำการเปลี่ยนแปลงค่าไปจาก -1 ถึง 1 เรื่อย ๆ (เรียกว่า iteration) โดยยึดถือเกณฑ์ที่ว่าค่าพารามิเตอร์ตัวใดให้ค่า MSE หรือ MAPE ต่ำที่สุด จะเป็นค่าประมาณของพารามิเตอร์ในตัวแบบ

ข. การตรวจสอบความถูกต้องของตัวแบบ ให้นำค่า residual จากตัวแบบที่ประมาณค่าแล้วมาทำการวิเคราะห์กราฟ autocorrelation coefficient ว่ามีลักษณะของกราฟเป็น random หรือไม่ โดยพิจารณาว่าค่า autocorrelation ของ residual มีค่าเข้าใกล้ 0 หรือไม่ มีกี่ค่า ที่ตกนอกแถบช่วงความเชื่อมั่น (confident band) ถ้าค่าส่วนใหญ่เข้าใกล้ 0 และไม่มีจุดตกนอกแถบช่วงความเชื่อมั่น แสดงว่า residual มีลักษณะ randomness ให้นำตัวแบบที่ได้ไปประยุกต์ใช้งานในขั้นตอนต่อไปได้ แต่ถ้าไม่มีลักษณะ randomness แสดงว่า ขั้นตอนในการหาตัวแบบผิดพลาดให้ไปวิเคราะห์หาตัวแบบใหม่อีกครั้ง โดยดำเนินการเช่นนี้เรื่อย ๆ จนกว่าจะได้ตัวแบบที่เหมาะสมกับข้อมูลอนุกรมเวลา

<u>ขั้นตอนที่ 3</u> ประยุกต์ตัวแบบไปใช้งาน

จากค่าความคลาดเคลื่อน (residual) ว่าเป็น random ของขั้นตอนที่ 2 เขียนได้เป็น $e_t = X_t - F_t$ ได้จากตัวแบบ ARIMA (p ,d ,q) เราสามารถประมาณค่าพยากรณ์ล่วงหน้า ต้องทราบค่าต่าง ๆ จากตัวแบบ ตัวอย่างเช่น ใน ARMA (1,1) ได้ตัวแบบเป็น

$$F_{t+1} = 0.35\ X_t + a_{t+1} - 0.20\ a_t$$

กรณีตัวแบบนี้ต้องทราบค่า X_t และ a_t จึงจะหาค่าพยากรณ์ของวาระที่ t+1 ได้

ตัวอย่างที่ 5.8 จากข้อมูล 45 period จงวิเคราะห์ตัวแบบที่เหมาะสมโดยวิธี Box - Jenkins แล้วหาค่าพยากรณ์ใน period ที่ 46 - 50

Period	ค่าสังเกต	Period	ค่าสังเกต	Period	ค่าสังเกต
1	2426.411	7	2624.803	13	2802.829
2	2457.842	8	2653.578	14	2832.128
3	2492.729	9	2682.360	15	2860.593
4	2526.982	10	2712.987	16	2888.414
5	2561.318	11	2743.418	17	2916.086
6	2593.570	12	2774.085	18	2944.661

Period	ค่าสังเกต	Period	ค่าสังเกต	Period	ค่าสังเกต
19	2974.686	28	3230.139	37	3491.975
20	3004.561	29	3258.426	38	3522.099
21	3032.981	30	3288.609	39	3550.525
22	3060.397	31	3317.237	40	3578.137
23	3088.676	32	3342.874	41	3606.044
24	3115.502	33	3370.009	42	3633.347
25	3141.658	34	3399.791	43	3662.882
26	3168.636	35	3434.208	44	3693.069
27	3197.796	36	3465.827	45	3725.575

จากข้อมูลนำไปหาค่า autocorrelation coefficient และ partial autocorrelation coefficient ของข้อมูลเดิม เมื่อ plot กราฟจะได้ว่ากราฟของ autocorrelation ไม่เข้าใกล้ 0 เมื่อ lag ไป 8 วาระ ดังนั้นข้อมูลอนุกรมเวลาชุดนี้เป็น nonstationary data กราฟของ partial autocorrelation ไม่ได้ใช้ประโยชน์อะไรเลย เมื่อข้อมูลเป็น nonstationary data

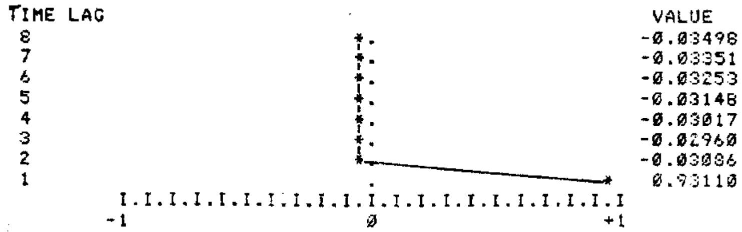
Autocorrelations and Partial Autocorrelations of the Original Data

```
ENTER TOTAL ORDER OF DIFFERENCING0
HOW MANY AUTOCORRELATIONS DO YOU WANT?8


TIME LAG                                                  VALUE    STD.ERR.
 8                        .        *                      0.47261    0.41566
 7                        .          *                    0.53545    0.40653
 6                        .           *                   0.59913    0.39413
 5                        .            *                  0.66370    0.37762
 4                        .              *                0.72916    0.35587
 3                        .               *               0.79552    0.32721
 2                        .                *              0.86284    0.28886
 1                        .                  *            0.93110    0.23518
        I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I
        -1                0                   +1

STANDARD ERROR (Q=0) = .149071
```

PARTIAL AUTOCORRELATIONS



รูปที่ 5.23 แสดงกราฟของ autocorrelation และ partial autocorrelation ของข้อมูลเดิม

จากนั้นให้นำข้อมูลทำ 1st difference คือ $(1-B)X_t = X_t - X_{t-1}$ เมื่อ $t = 2,3,\ldots,n$ เมื่อได้ข้อมูลหลังจากการทำ 1st difference ให้นำไปหา autocorrelation coefficient และ partial autocorrelation coefficient จากรูปที่ 5.24 กราฟของ autocorrelation ค่าส่วนใหญ่เข้าใกล้ 0 ยกเว้นค่าในวาระแรกและวาระที่สอง จึงสรุปได้ว่า ข้อมูลเป็น stationary แล้วหลังจากทำ 1st difference ค่า autocorrelation ในวาระที่ 7 และ 8 มีค่าสูงเล็กน้อย แต่ต่ำกว่า 2 เท่าของ standard error คือ 2 (0.1508) = 0.3016 ดังนั้น จึงถือว่าค่าส่วนใหญ่เข้าใกล้ 0 หลังจากนั้นให้นำกราฟของ autocorrelation และ partial autocorrelation เปรียบเทียบว่าเข้าลักษณะ model ใด เมื่อพิจารณารูปที่ 5.15 กราฟในรูปที่ 5.24 มีลักษณะใกล้เคียงที่สุด จึงได้ตัวแบบที่สอดคล้องกับข้อมูลอนุกรมเวลาชุดนี้ คือ ARIMA (1,1,0) คือ

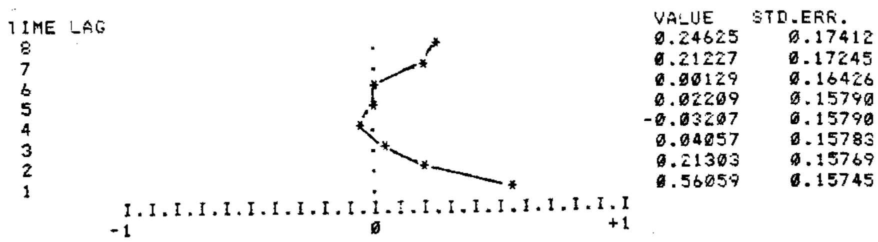
$$(1-\Phi_1 B)(1-B)X_t = a_0 + a_t \quad \text{เมื่อ } a_0 \text{ เป็น constant term}$$

$$X_t - \Phi_1 X_{t-1} - X_{t-1} + \Phi_1 X_{t-2} = a_0 + a_t$$

$$X_t = (1-\Phi_1)X_{t-1} - \Phi_1 X_{t-2} + a_0 + a_t$$

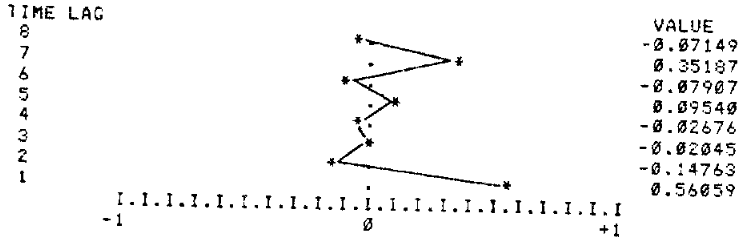
จากการ run โปรแกรมประมาณค่า $\Phi_1 = 0.5593$, constant term = 13.01 variance = 3.56

First Differences of the Data



STANDARD ERROR (Q=0) = .150756

PARTIAL AUTOCORRELATIONS



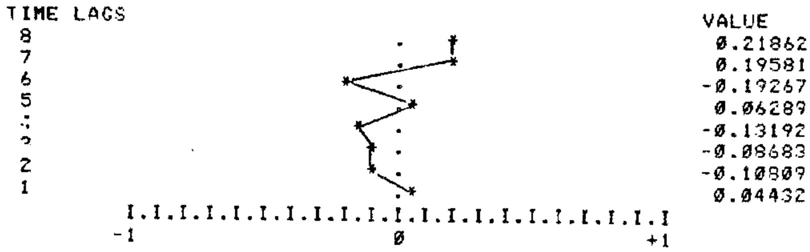

```
WHAT ORDER OF AR PROCESS TO BE FITTED
1

WHAT ORDER OF MA PROCESS TO BE FITTED ESTIMATE OF OVERALL CONSTANT TERM= 12.9742
                                      ESTIMATE OF WHITE NOISE VARIANCE= 3.55348

0
ESTIMATED AUTOREGRESSIVE PARAMETERS
NO. 1     = .560592
```

รูปที่ 5.24 แสดงกราฟของ autocorrelation และ partial autocorrelation หลังจากทำ 1st difference

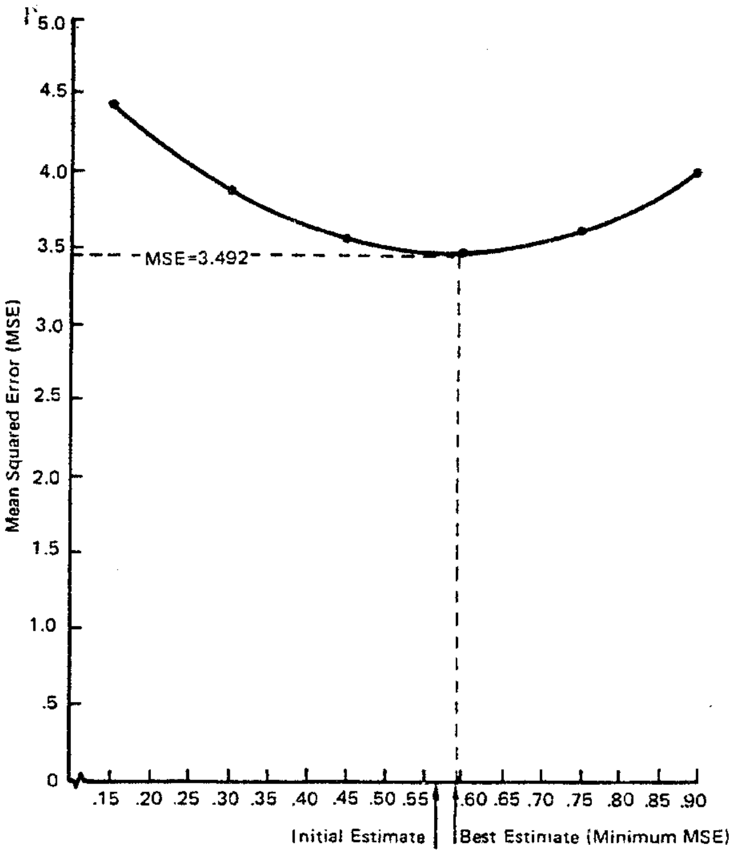
งานที่สำคัญ คือ ต้องตรวจสอบว่าตัวแบบที่กำหนดให้กับข้อมูลอนุกรมเวลาชุดนี้ มีความเหมาะสมถูกต้องหรือไม่ โดยหาค่า autocorrelation ของ residual หรือคือ white noise นั่นเอง ตรวจสอบได้โดยพิจารณากราฟของ autocorrelation ของ residual ค่าส่วนใหญ่เข้าใกล้ 0 จึงสรุปได้ว่ากราฟของ residual เป็น random ตัวแบบที่กำหนดให้กับข้อมูลอนุกรมเวลาชุดนี้จึงมีความเหมาะสม




```
CHI-SQUARED GOODNESS OF FIT STATISTIC= 14.7552
ON 24    DEGREES OF FREEDOM
```

รูปที่ 5.25 แสดงกราฟ autocorrelation ของ residuals

หน้าที่สำคัญคือพยายาม search หาค่าพารามิเตอร์ Φ_1 ที่เหมาะสม คือให้ค่า MSE ต่ำที่สุด จากรูปที่ 5.26 แสดงถึงค่า MSE จากค่าประมาณ Φ_1 ที่แตกต่างกัน ค่า $\Phi_1 = 0.58$ จะให้ค่า MSE ต่ำที่สุด คือ MSE = 3.492 ขณะที่ครั้งแรกที่เราประมาณค่า $\Phi_1 = 0.5593$ มีค่า MSE = 3.56



รูปที่ 5.26 แสดงค่าของ mean square error

เมื่อ Φ_1 = 0.58 constant term = 12.401 ได้ตัวแบบของข้อมูลชุดนี้ คือ

$$X_t = a_0 + (1 + 0.58) X_{t-1} - 0.58 X_{t-2} + a_t$$

ค่าพยากรณ์วาระที่ 46 - 50 คือ

$$F_{46} = a_0 + 1.58 X_{45} - 0.58 X_{44}$$

$$F_{46} = 12.401 + 1.58 (3725.575) - 0.58(3693.069) = 3756.83$$

$$F_{47} = 12.401 + 1.58 (3756.83) - 0.58(3725.575) = 3787.36$$

$$F_{48} = 3817.47$$

$$F_{49} = 3847.33$$

$$F_{50} = 3877.05$$

แบบฝึกหัด

ข้อมูลต่อไปนี้เป็นยอดขาย (หน่วยพันฟรังส์) ในอุตสาหกรรมกระดาษระหว่างปี 1963 - 1972 ใช้ข้อมูลต่อไปนี้วิเคราะห์หาตัวแบบที่เหมาะสมที่สุดและตรวจสอบตัวแบบที่ได้ว่าเหมาะสมถูกต้องหรือไม่ ถ้าตัวแบบถูกต้องให้หาค่าพยากรณ์ใน 12 period ถัดไปของปี 1973

period	observation	period	observation	period	observation
1	56.27	41	70.11	81	74.20
2	59.90	42	79.01	82	84.72
3	66.85	43	59.46	83	73.17
4	59.78	44	23.07	84	89.85
5	57.99	45	61.72	85	77.81
6	66.82	46	69.14	86	85.61
7	49.92	47	70.11	87	93.88
8	21.52	48	70.58	88	81.30
9	55.58	49	74.76	89	78.34
10	58.69	50	77.34	90	82.81
11	54.61	51	81.38	91	65.73
12	57.11	52	76.67	92	31.00
13	63.47	53	72.89	93	78.00
14	63.93	54	74.92	94	86.00
15	71.22	55	68.10	95	78.00
16	62.16	56	24.14	96	80.80
17	62.10	57	68.02	97	89.52
18	67.60	58	70.83	98	85.61
19	50.13	59	69.42	99	89.33
20	22.03	60	77.21	100	87.50
21	56.07	61	79.53	101	83.51
22	60.25	62	78.84	102	93.46
23	62.64	63	89.00	103	83.25
24	60.55	64	79.74	104	30.00

period	observation	period	observation	period	observation
25	64.68	65	75.10	105	79.14
26	65.84	66	82.13	106	90.00
27	71.29	67	69.16	107	78.17
28	68.77	68	29.07	108	88.00
29	72.39	69	72.71	109	87.50
30	70.72	70	86.84	110	99.30
31	62.90	71	81.24	111	97.68
32	23.75	72	79.96	112	96.87
33	61.33	73	84.30	113	87.17
34	73.04	74	84.70	114	100.69
35	73.49	75	94.20	115	83.20
36	65.18	76	80.43	116	34.56
37	67.62	77	84.03	117	84.95
38	74.82	78	87.15	118	91.39
39	81.07	79	65.63	119	86.87
40	72.94	80	37.05	120	99.37

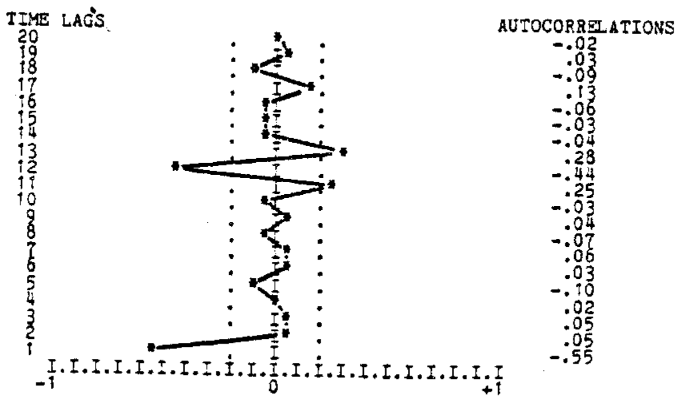
เฉลย นำค่าข้อมูลเดิมไปหาค่า autocorrelation coefficient และ partial autocorrelation coefficient แล้ว plot กราฟได้ดังรูป

Autocorrelation Coefficients for 1 Short and 1 Long Difference

*** AUTOCORRELATION COEFFICIENTS ***

STANDARD ERROR = .155 CHI SQUARED = 84.09

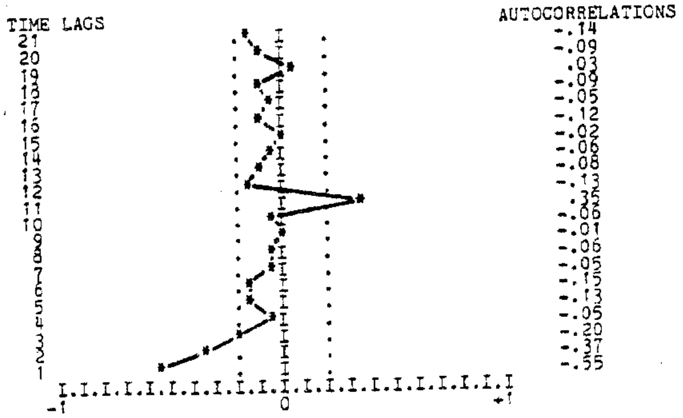
DEGREES OF FREEDOM = 32



*** PARTIAL AUTOCORRELATION COEFFICIENTS ***

STANDARD ERROR = .097 CHI SQUARED= 93.65

DEGREES OF FREEDOM = 32



จากกราฟ autocorrelation coefficient โดย lag ไป 20 วาระ ค่าส่วนใหญ่พุ่งเข้าสู่ 0 ยกเว้นค่าใน period ที่ 12 โดยกราฟของ partial autocorrelation coefficient เข้าลักษณะ exponential พุ่งเข้าสู่ 0 เมื่อพิจารณากราฟจะคล้ายกราฟของ MA(1) ในรูปที่ 5.19 จากนั้นไป plot กราฟของ autocorrelation coefficient ของ residual จะได้ว่าทุกค่าเข้าใกล้ศูนย์ ดังนั้น ตัวแบบ MA(1) จึงถูกต้องเหมาะสมกับข้อมูลอนุกรมเวลาชุดนี้

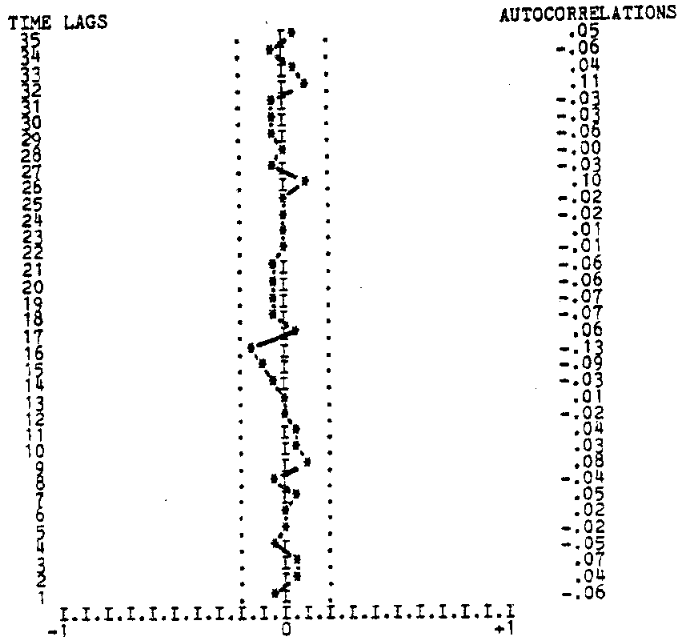
Output for MA(1) Model, Seasonal in MA

```
SHORT TERM DIFFERENCES =  1  LONG TERM    =  1
LENGTH OF SEASONALITY  = 12  MODEL NUMBER=  6

   1 ST PARAMETER =       .82553
   2 ND PARAMETER =            0
   3 RD PARAMETER =       .61228

   MSE =                    19.7
   NUMBER OF ITERATIONS =     39
```

```
*** RESIDUAL AUTOCORRELATION COEFFICIENTS ***
STANDARD ERROR =    .097        CHI SQUARED =        11.91
      DEGREES OF FREEDOM = 32
```



ค่าพยากรณ์ของ period ที่ 121 - 132 ได้ดังนี้

Forecasts for Periods 122 to 132

```
MEAN PC ERROR (MPE) OR BIAS          =         -1.67%
MEAN SQUARED ERROR (MSE)             =          19.4
MEAN ABSOLUTE PC ERROR     (MAPE)    =           5.6%

FORECAST FOR HOW MANY PERIODS AHEAD (0=NONE,24=MAX)?12
```

PER	FORECAST	95 PC.BOUNDS	
121	94.6	85.7	103.5
122	99.7	88.2	111.2
123	103.0	91.5	114.5
124	98.2	86.7	109.7
125	92.9	81.4	104.5
126	102.2	90.7	113.8
127	86.9	75.4	98.4
128	42.3	30.8	53.9
129	89.5	77.9	101.0
130	97.9	86.3	109.4
131	90.7	79.2	102.3
132	99.9	88.4	111.4

```
DO YOU WANT TO TRY ANOTHER SET OF DIFFERENCES?N
MODEL NO. (0=NO MORE)?0
```