บทที่ 2

ระเบียบวิธีการทางสถิติที่ใช้ในการวิจัย

การที่เราต้องการทราบค่าพยากรณ์ของสิ่งที่เกิดขึ้นในอนาคต ตามระยะเวลา เช่น ปี, เดือน, สัปดาห์ ฯลฯ วิธีทางสถิติที่ใช้ได้ดีสำหรับการพยากรณ์คือ การวิเคราะห์อนุกรมเวลา ซึ่งจะเน้นถึงการหาโมเดลหรือรูปแบบที่จะใช้พยากรณ์ในปีต่อ ๆ ไป หรืออาจกล่าวได้ว่าการวิเคราะห์อนุกรมเวลา เป็นการศึกษาลักษณะการเคลื่อนไหวของข้อมูลชุดหนึ่งที่เกิดขึ้นตามลำดับของเวลา ข้อมูลนั้น ๆ ประกอบด้วยปัจจัยต่าง ๆ 4 ประเภทคือ แนวโน้ม (Trend) ฤดูกาล (Season) วัฏจักร (Cycle) และเหตุการณ์ผิดปกติ (Irregularity)

2.1 ปัจจัยที่มีผลต่ออนุกรมเวลา

แนวโน้ม (Trend) แนวโน้มแสดงถึงทิศทางที่อนุกรมเวลา (ข้อมูล) ชุดนั้นพุ่งไปสู่ในระยะเวลาค่อนข้างยาว แนวโน้มอาจจะเป็นลักษณะเส้นตรง เส้นโค้ง ฯลฯ สำหรับค่าแนวโน้มจะใช้อักษรย่อ T

ฤดูกาล (Season) เป็นพฤติการณ์ที่เกิดขึ้นซ้ำ ๆ กันในช่วงระยะเวลาสั้น ๆ เช่น 1 ช.ม., 1 สัปดาห์, 1 เดือน ฯลฯ โดยจะเกิดขึ้นซ้ำกันในช่วงเวลาเดียวกัน จะแสดงการเปลี่ยนแปลงเกิดจากฤดูกาลในลักษณะของแบบ (pattern) ที่เกิดขึ้นซ้ำกันในแต่ละช่วง ซึ่งเรียกว่าดัชนีฤดูกาล (Seasonal Index) ใช้อักษรย่อ S หน่วยเป็นเปอร์เซ็น

วัฏจักร (Cycle) เป็นส่วนประกอบส่วนหนึ่งของอนุกรมเวลา โดยทั่วไปวัฏจักรจะประกอบด้วยระยะเวลาที่รุ่งเรือง (prosperity) แล้วค่อยตกต่ำ (depression)

เวลายาวมากในการที่จะมองเห็นเหตุการณ์เคลื่อนไหวของวัฏจักร สำหรับค่าวัฏจักรใช้อักษรย่อ C หน่วยเป็นเปอร์เซ็น

<u>เหตุการณ์ผิดปกติ</u> (Irregularity) เป็นเหตุการณ์ที่เกิดขึ้นโดยไม่คาดคิดมาก่อนล่วงหน้า เช่น การเกิดสงคราม การเกิดแผ่นดินไหว ฯลฯ ไม่ได้อยู่ภายใต้กฎที่แน่นอนเหมือน T, S, C เรียกได้อีกอย่างหนึ่งว่า การเปลี่ยนแปลงเชิงสุ่ม สำหรับค่าเหตุการณ์ผิดปกติใช้อักษรย่อ I หน่วยเป็นเปอร์เซ็น

การรวมตัวของปัจจัยทั้ง 4 ประเภท ที่ประกอบขึ้นเป็นอนุกรมเวลามีลักษณะของการรวมตัวได้ 2 แบบ

ก. การรวมตัวในเชิงคูณ หรือตัวแบบในเชิงคูณ (Multiplicative Model)

$$Y = T \times S \times C \times I$$

ข. การรวมตัวในเชิงบวก หรือตัวแบบในเชิงบวก (Additive Model)

$$Y = T+S+C+I$$

การพิจารณาว่าข้อมูลที่ศึกษามีการรวมตัวแบบใด จะต้องอาศัยประสบการณ์เพื่อพิจารณาว่าการเพิ่มหรือลดของปัจจัยแต่ละประเภทจะกระทบกระเทือนกันหรือไม่ ถ้ากระทบกระเทือนกันจะเป็นตัวแบบเชิงคูณ แต่ถ้าไม่กระทบกระเทือนกันจะเป็นตัวแบบในเชิงบวก ยกตัวอย่างเช่น ยอดขายของเดือนมิถุนายนของบริษัทแห่งหนึ่งมักจะสูงกว่ายอดขายของเดือนพฤษภาคมเกือบทุกปี ในปี พ.ศ.2515 ยอดขายเดือนมิถุนายนเท่ากับ 100,000 บาท และเดือนพฤษภาคมเท่ากับ 80,000 บาท ผลแตกต่างระหว่างยอดขายของสองเดือนนี้ เป็นที่เชื่อแน่ว่าเกิดจากปัจจัยทางฤดูกาล ยอดขายของเดือนมิถุนายนสูงกว่าเดือนพฤษภาคมเท่ากับ 20,000 บาท หรือสูงกว่า 25 เปอร์เซนต์ สำหรับปี พ.ศ.2515 ไม่มีความแตกต่างกันเลยไม่ว่าจะพยากรณ์แบบใด แต่สำหรับปี พ.ศ.2516 ถ้ายอดขายของเดือนพฤษภาคม มีเท่ากับ 110,000 บาท การพยากรณ์ยอดขายปี พ.ศ. 2516 เดือนมิถุนายน สามารถทำได้ 2 แบบคือ ถ้ายอดขายของเดือนมิถุนายนพยากรณ์

ได้เท่ากับ 130,000 บาท ถือว่าผลแตกต่างทางฤดูกาลมีค่าเป็นบวกเท่ากับ 20,000 บาท แต่ถ้าพยากรณ์ยอดขายเดือนมิถุนายนได้เท่ากับ 137,500 บาท แสดงว่าผลแตกต่างที่เกิดจากฤดูกาลเท่ากับ 25 เปอร์เซนต์ จึงคูณยอดขายเดือนพฤษภาคม ปี 2516 ด้วยดัชนีฤดูกาล .125 วิธีใดเป็นวิธีที่เหมาะสมขึ้นอยู่กับเหตุผลและประสบการณ์ของแต่ละบุคคล แต่ละบริษัท โดยทั่วไป สำหรับข้อมูลทางธุรกิจใช้ตัวแบบในเชิงคูณ ส่วนข้อมูลซึ่งเป็นจำนวนนักท่องเที่ยวที่เข้ามาท่องเที่ยวในประเทศไทยใช้ตัวแบบในเชิงคูณ

การพยากรณ์อาจจะใช้เฉพาะค่า T หรือถ้าต้องการให้ละเอียดขึ้นทำได้โดยการใช้ค่า S มาปรับค่า T ให้เป็นไปตามฤดูกาล ข้อมูลที่มีฤดูกาลส่วนใหญ่จะเป็นข้อมูลรายเดือน การหาค่า S จะทำได้ภายหลังจากการกำจัดค่า T,C และ I ออกจากข้อมูลเดิม การหาค่า S ทำได้หลายวิธีซึ่งจะกล่าวถึงในตอนที่ 2.3 ถ้าต้องการพยากรณ์ให้ละเอียดขึ้นจะต้องหาค่าพยากรณ์ของ CI ซึ่งจะทำภายหลังจากการกำจัดค่า T และ S จากข้อมูลเดิม โดย $CxI = \frac{Y}{TxS}$ และนำค่าพยากรณ์ CxI โดยอาจจะพยากรณ์ค่า C และพยากรณ์ค่า I นำมาปรับกับค่า TxS หรือจะนำค่าพยากรณ์ค่า C กับค่า I โดยไม่ต้องแยกพยากรณ์ และนำค่าพยากรณ์C รวมกับI มาปรับค่า TxS

2.2 วิธีการหาค่าแนวโน้ม(T)

มีหลายวิธี แต่จะกล่าวถึงเฉพาะวิธีที่จะใช้ในการวิเคราะห์อนุกรมเวลาในบทที่ 3 ดังต่อไปนี้

2.2.1 วิธีถัวเฉลี่ยเคลื่อนที่ (The Moving Average Method) การถัวเฉลี่ยเคลื่อนที่สามารถที่จะทำการเฉลี่ยเคลื่อนที่ 3 เดือน 6 เดือน 3 ปี ฯลฯ แล้วแต่กรณีสังเกตลักษณะของการเกิดซ้ำซาก ทุก 3 เดือน 6 เดือน 12 เดือน ฯลฯ ข้อดีของวิธีถัวเฉลี่ยเคลื่อนที่เป็นวิธีการที่ง่าย ค่าแนวโน้มที่ได้มักจะเคลื่อนตามความเคลื่อนไหวของอนุกรมชุดเดิม ข้อมูลไม่จำเป็นต้องมีลักษณะเป็นเส้นตรง ข้อเสียของวิธีนี้มี 2

2.2.2 วิธีกำลังสองน้อยที่สุด (Least Square Method) เป็นวิธีหาค่าแนวโน้มที่ให้ผลรวมของผลต่างระหว่างค่าจริงในข้อมูลและค่าแนวโน้มยกกำลังสอง หรือ $\sum(Y-\hat{Y})^2$ มีค่าน้อยที่สุด การจะทราบว่าข้อมูลชุดนั้นมีแนวโน้มในลักษณะใดจะทำได้โดยการพล็อตข้อมูลลงในกราฟ จะอยู่ในลักษณะใดลักษณะหนึ่ง เช่น โพลีโนเมียลกำลังหนึ่ง โพลีโนเมียลกำลังสอง โพลีโนเมียลกำลังสาม เอ็กซ์โพเนนเชียล ฯลฯ ข้อดีของวิธีนี้คือค่าแนวโน้มที่ได้จะเป็นตัวแทนที่ดีที่สุดของอนุกรมเวลาชุดนั้น ๆ เพราะผลรวมของผลต่างระหว่างค่าจริงในข้อมูล และค่าแนวโน้มยกกำลังสองมีค่าน้อยกว่าวิธีอื่น ๆ

เมื่อข้อมูลมีลักษณะเป็นโพลีโนเมียลกำลัง n มีสมการดังนี้

$$Y = \alpha_0+\alpha_1X+\alpha_2X^2+\ldots\ldots\ldots\ldots+\alpha_nX^n+\epsilon$$

$$\hat{Y} = a_0+a_1X+a_2X^2+\ldots\ldots\ldots\ldots+a_nX^n$$

เมื่อ $a_0, a_1, a_2 \ldots, a_n$ เป็นค่าประมาณของ $\alpha_0, \alpha_1, \alpha_2 \ldots\ldots \alpha_n$ การจะหาค่าประมาณที่ทำให้ $\sum(Y-\hat{Y})^2$ มีค่าน้อยที่สุดโดยการดิฟเฟอเรนชิเอต (Differentiate) $\sum(Y-\hat{Y})^2$ เทียบกับค่าพารามิเตอร์และให้ผลจากการดิฟเฟอเรนชิเอตเท่ากับ 0 ซึ่งเรียกว่าสมการปกติ (Normal Equation) และจะหาค่าประมาณได้จากการแก้สมการปกติ เช่น สมการโพลีโนเมียลกำลังสาม (Hyperbolic Equation)

$$Y = \alpha_0+\alpha_1X+\alpha_2X^2+\alpha_3X^3+\epsilon$$

$$\hat{Y} = a_0+a_1X+a_2X^2+a_3X^3$$

เมื่อ a_0, a_1, a_2, a_3 เป็นค่าประมาณของ $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ จะได้สมการปกติ

$$\sum Y = Na_0+a_1\sum X+a_2\sum X^2+a_3\sum X^3$$

$$\sum XY = a_0\sum X+a_1\sum X^2+a_2\sum X^3+a_3\sum X^4$$

$$\sum X^2Y = a_0\sum X^2+a_1\sum X^3+a_2\sum X^4+a_3\sum X^5$$

$$\sum X^3Y = a_0\sum X^3+a_1\sum X^4+a_2\sum X^5+a_3\sum X^6$$

จะหาค่า a_0, a_1, a_2, a_3 ได้โดยการแก้สมการปกตินี้ เมื่อข้อมูลมี

$$Y = AB^X$$

$$\hat{Y} = ab^X$$

เมื่อ a, b เป็นค่าประมาณของ A,B ซึ่งสามารถเปลี่ยนให้อยู่ในรูปของเส้นตรงได้ โดยการใส่ log.

$$\log \hat{Y} = \log a + X \log b$$

ซึ่งก็อยู่ในรูปของสมการโพลีโนเมียลกำลังหนึ่งจะหาค่าประมาณได้ตามวิธีโพลีโนเมียล

2.2.3 วิธีอื่น ๆ นอกจากวิธีถัวเฉลี่ยเคลื่อนที่และวิธีกำลังสองน้อยที่สุด ยังมีวิธีการหาค่าแนวโน้มอื่น ๆ เช่น การเขียนแนวโน้มด้วยสายตา วิธีนี้ไม่ค่อยใช้เพราะไม่ค่อยมีเหตุผลเพียงพอที่จะทราบได้ว่า เส้นแนวโน้มที่เขียนขึ้นจะเป็นเส้นแนวโน้มที่เหมาะสมกับข้อมูลชุดนั้นมากที่สุด การเขียนแนวโน้มโดยการเลือกจุด 2 จุด วิธีนี้ต้องทราบว่าค่าแนวโน้มของข้อมูลเป็นเส้นตรงซึ่งจะหาค่าแนวโน้มได้หลายสมการ ทำให้ไม่สามารถที่จะตัดสินใจได้ว่าสมการไหนดีที่สุด วิธีหาค่าแนวโน้มกึ่งถัวเฉลี่ย (Semiaverage Method) วิธีนี้คล้ายกับการเขียนแนวโน้มโดยการเลือกจุด 2 จุด จะต่างกันเพียงแต่จุด 2 จุดที่เลือกเป็นค่าเฉลี่ยของอนุกรมเวลาโดยนำอนุกรมเวลาที่ต้องการศึกษาแบ่งออกเป็น 2 ชุด แล้วนำมาหาค่าเฉลี่ย จะได้จุด 2 จุดที่เส้นตรงจะผ่าน

2.2.4 การทดสอบโมเดล การทดสอบโมเดลจะทำขึ้นภายหลังจากการหาสมการสำหรับค่าแนวโน้มแบบต่าง ๆ เรียบร้อยแล้ว เพื่อต้องการดูว่าโมเดลใดจะเหมาะสมที่สุด การทดสอบโมเดลที่นิยมใช้มี 2 วิธีดังต่อไปนี้

2.2.4.1 การทดสอบโดยใช้หลักว่าโมเดลที่เหมาะสมที่สุดคือโมเดลที่ให้ค่า $\Sigma(Y-\hat{Y})^2$ น้อยที่สุด

2.2.4.2 การทดสอบ[4]โดยใช้ F-test ถ้าข้อมูลเป็นแบบโพลีโนเมียล การทดสอบโดยใช้ F - test จะอยู่ภายใต้ข้อสมมติที่ว่า Y_i เป็น random variable ϵ_i เป็น random variable $\sim N(0,\sigma^2)$ และ ϵ_i, ϵ_j ไม่มีความสัมพันธ์กัน คือ $Cov(\epsilon_i \epsilon_j) =$

4

จะทำการพิจารณาโดยเริ่มจากโพลีโนเมียลกำลังหนึ่ง การทดสอบโมเดลโดยการวิเคราะห์ความแปรปรวน จะพิจารณาโดยการทดสอบสมมติฐาน โดยการตั้งสมมติฐานจากสมการ

$$Y_i = \alpha_{i0} + \alpha_{i1}X + \dots + \alpha_{ii}X^i + \epsilon_i \qquad i = 1,2,\dots\dots$$

$$\hat{Y}_i = \hat{\alpha}_{i0} + \hat{\alpha}_{i1}X + \dots + \hat{\alpha}_{ii}X^i$$

Null Hypothesis (Ho) : $\alpha_{ii} = 0$

Alternative Hypothesis (H_A) : $\alpha_{ii} \neq 0$

ตารางที่ 2.1 สูตรในการวิเคราะห์ค่าแปรปรวนเพื่อทดสอบสมมติฐาน

Ho : $\alpha_{ii} = 0$ ในสมการโพลีโนเมียล

Source of Variation	Degree of Freedom	Sum of Square	Mean Square	F
$\alpha_{i0}, \alpha_{i1}, \dots, \alpha_{ii}$	i+1	$\hat{\alpha}_{i0}\Sigma Y + \hat{\alpha}_{i1}\Sigma XY + \dots + \hat{\alpha}_{ii}\Sigma X^i Y = (1)$		
$\alpha_{i0}, \alpha_{i1}, \dots \alpha_{ii-1}$	i	$\hat{\alpha}_{i-1,0}\Sigma Y + \hat{\alpha}_{i-1,1}\Sigma XY + \dots + \hat{\alpha}_{i-1,i-1}\Sigma X^{i-1}Y = (2)$		
$\alpha_{ii}/\alpha_{i0}, \dots \alpha_{ii-1}$	1	(1) - (2) = (3)	(3)	$F = \frac{(3)}{(5)/n-(i+1)}$
Residual	n-(i+1)	(4) - (1) = (5)	(5)/n-(i+1)	
Total (uncorrected)	n	$\Sigma Y^2 = (4)$		

n = จำนวนข้อมูลทั้งหมด

ที่ระดับนัยสำคัญ $\alpha = .05$ หรือ $\alpha = .10$ ถ้าค่า F ที่คำนวณได้จากการวิเคราะห์ความแปรปรวนน้อยกว่าค่า $F_{(1,\ n-(i+1),\ 1-\alpha)}$ ที่ได้จากตารางสถิติ

คำนวณได้มากกว่า $F_{(1,n-(i+1),\ 1-\alpha)}$ ที่ได้จากตารางสถิติ จะปฏิเสธสมมติฐาน $\alpha_{ii} = 0$ ซึ่งหมายความว่า α_{ii} มีความสำคัญเพียงพอที่จะอยู่ในสมการ ทดสอบโพลีโนเมียลกำลังสูงขึ้นต่อไป จนกว่าจะยอมรับสมมติฐาน เมื่อข้อมูลมีลักษณะเป็นเอ็กซ์โพเนนเชียล การทดสอบก็ทำเช่นเดียวกับโพลีโนเมียลกำลังหนึ่ง เพราะเมื่อเปลี่ยนสมการเอ็กซ์โพเนนเชียล โดยการใส่ log จะได้สมการใหม่ ซึ่งเป็นสมการโพลีโนเมียลกำลังหนึ่ง

2.3 <u>วิธีการหาดัชนีฤดูกาล</u> (S)

ดัชนีฤดูกาล เป็นดัชนีแสดงถึงอัตราร้อยละของการเข้ามาของนักท่องเที่ยวที่เข้ามาท่องเที่ยวในประเทศไทยในแต่ละเดือน การหาดัชนีฤดูกาลทำได้หลายวิธี ดังต่อไปนี้

2.3.1 วิธีถัวเฉลี่ยอย่างง่าย (The Method of Simple Average) ข้อดีของวิธีนี้คือ การคำนวณง่ายไม่ยุ่งยากอย่างวิธีอื่น ๆ แต่มีข้อเสียคือ เป็นวิธีที่ไม่ละเอียดพอ

2.3.2 วิธีอัตราส่วนต่อค่าเฉลี่ยเคลื่อนที่ (Ratio-to-Moving-Average Method) ข้อดีของวิธีนี้คือ หลีกเลี่ยงความผิดพลาดที่เกิดจากการนำเอาการเปลี่ยนแปลงอันเนื่องจากฤดูกาลไปกละกับค่าแนวโน้ม และถูกต้องกว่าวิธีอื่น ในกรณีที่ค่าแนวโน้มมิได้มีลักษณะเป็นเส้นตรง ในกรณีที่วัฏจักรครอบคลุมระยะเวลาที่ค่อนข้างสั้น เช่น 5 ปี 10 ปี วิธีนี้จะแสดงให้เห็นถึงการเปลี่ยนแปลงอันเนื่องจากวัฏจักร และค่าแนวโน้มได้ง่ายและดีกว่าเส้นโค้งทางคณิตศาสตร์อื่น

2.3.3 วิธีอื่น ๆ เช่นวิธีอัตราส่วนต่อค่าเฉลี่ยเคลื่อนที่อย่างง่าย (Ratio-to-Simple Average Method) หรือวิธีอัตราส่วนต่อค่าแนวโน้ม (The Ratio-to Trend-Method) ทั้งสองวิธีเป็นวิธีที่ค่อนข้างง่าย และจะให้ผลดีในกรณีที่ค่าแนวโน้มเป็นเส้นตรง แต่ถ้าค่าแนวโน้มไม่ได้เป็นเส้นตรงจะให้ผลไม่ดีพอ

การสร้างดัชนีฤดูกาลโดยวิธีอัตราส่วนต่อค่าเฉลี่ยเคลื่อนที่

เริ่มด้วยการคำนวณค่าเฉลี่ยเคลื่อนที่ 12 เดือน ที่ตกอยู่กับกลางเดือนของแต่ละเดือนเพื่อที่จะกำจัดการเปลี่ยนแปลงอันเนื่องจากฤดูกาล และเหตุการณ์ผิดปกติออกจากข้อมูล ดังนั้นค่าเฉลี่ยเคลื่อนที่ที่ได้จึงเป็นค่าประมาณของค่าแนวโน้มและการเปลี่ยนแปลงอันเนื่องจากวัฏจักร เมื่อได้ค่าเฉลี่ยเคลื่อนที่ของแต่ละเดือนแล้วหารข้อมูลเดิมของแต่ละเดือน ด้วยค่าเฉลี่ยเคลื่อนที่ของเดือนนั้น ๆ ขั้นสุดท้ายคือการหาค่าเฉลี่ยของแต่ละเดือน และปรับให้ผลรวมของค่าเหล่านั้น ซึ่งมี 12 ค่าให้เท่ากับ 1,200 เพราะดัชนีฤดูกาลปกติแต่ละเดือนมีอัตราร้อยละ 100

2.4 วิธีการหาค่าเปลี่ยนแปลงอันเนื่องมาจากวัฏจักร

ค่าวัฏจักรมักจะเป็นอัตราร้อยละโดยหาได้จากการกำจัดค่า T และ S ในอนุกรมเวลาโดย $CxI = \frac{Y}{TxS}$ ซึ่ง Y แทนอนุกรมเวลาเดิม ค่า TxS เรียกว่าค่าปกติ และค่า CxI เรียกว่าอัตราร้อยละของค่าปกติ ขั้นต่อไปคือการหาค่า C จากค่า CxI อย่างไรก็ตามบางครั้งวัฏจักรไม่ปรากฏให้เห็นเด่นชัด เมื่อพิจารณาอนุกรมเวลาที่เกิดในช่วงเวลายาวนาน จึงไม่จำเป็นที่จะต้องแยกค่า C ออกจากค่า CxI

การหาค่าวัฏจักร (C)

จะหาค่า ได้จากการกำจัดค่า I ออกจากค่า CxI โดยวิธีถัวเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก น้ำหนักที่ใช้ถ่วงในการคำนวณค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักโดยมากใช้สัมประสิทธิ์ของไบโนเมียล (Binomial Coefficient) เช่น 1,2,1 หรือ 1,4,6,4,1 ฯลฯ การใช้ค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก มีข้อดีคือทำให้อนุกรมเวลานั้นราบเรียบขึ้น ถ้าต้องการให้ราบเรียบมากจะต้องเลือกสัมประสิทธิ์ของไบโนเมียลที่สูงขึ้นด้วยค่าที่ได้จากวิธีนี้ ยังไม่สามารถหาค่าพยากรณ์ C ได้ เพราะไม่มีสมการที่แน่นอนสำหรับใช้พยากรณ์ ถ้าต้องการหาค่าพยากรณ์จะต้องหาสมการที่แน่นอนที่สามารถพยากรณ์ได้ วิธี

ในบางครั้งการพยากรณ์ค่า C ค่าเดียวอาจจะไม่ถูกต้อง เพราะลำบากที่จะแยกค่า C ออกจากค่า I จึงพยากรณ์ค่า C และ I พร้อมกัน โดยไม่ต้องแยกพยากรณ์แต่ละค่า การพยากรณ์ค่า C และ I พร้อมกันจะเรียกว่าการพยากรณ์ CI ซึ่งนิยมใช้คือการพยากรณ์โดยเทคนิคของบ๊อกซ์และเจนกินส์

2.5 วิธีการประมาณค่าการเปลี่ยนแปลงอันเนื่องจากเหตุการณ์ผิดปกติ

การหาค่าเหตุการณ์ผิดปกติ (I) จะหาได้จากการกำจัดค่า T,S,C ออกจากอนุกรมเวลาเดิม ซึ่ง $I = \frac{Y}{TxSxC} = \frac{CxI}{C}$ ค่า I ที่ได้ไม่มีรูปแบบที่แน่นอนสำหรับการพยากรณ์ จึงต้องหารูปแบบสำหรับค่า I โดยเทคนิคของบ๊อกซ์และเจนกินส์

2.6 วิธีการพยากรณ์โดยเทคนิคของบ๊อกซ์และเจนกินส์

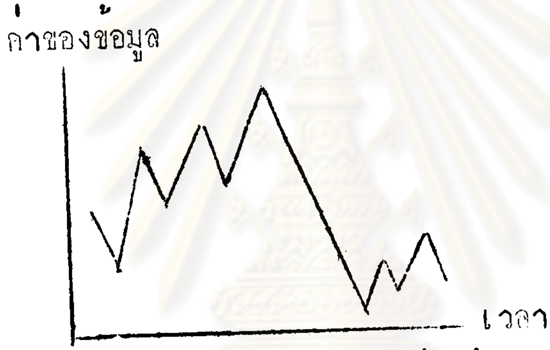
อนุกรมเวลา หมายถึงข้อมูลที่เกิดขึ้นในระยะเวลาต่าง ๆ กัน ซึ่งเรียงลำดับโดยคำนึงถึง เวลาของการเกิด โดยระยะเวลาของการเกิดเป็นระยะเวลาเท่า ๆ กัน อาจจะเป็นรายเดือน รายสัปดาห์ รายวัน ฯลฯ การพยากรณ์ค่าในอนาคตเป็นวัตถุประสงค์ข้อใหญ่ในการวิเคราะห์อนุกรมเวลา วิธีพยากรณ์มีหลายวิธีในที่นี้จะกล่าวถึงวิธีพยากรณ์วิธีหนึ่งที่นิยมมาก โดยใช้เทคนิคของบ๊อกซ์และเจนกินส์

2.6.1 ประเภทของอนุกรมเวลา ที่สามารถกำหนดรูปแบบโดยเทคนิคของบ๊อกซ์และเจนกินส์ แบ่งเป็น 3 ประเภท

2.6.1.1 อนุกรมเวลาคงที่ (Stationary Time Series) หมายถึง ข้อมูลของอนุกรมเวลาที่มีคุณสมบัติทางสถิติ (เช่นค่าเฉลี่ย ค่าแปรปรวน หรือฟังชั่นความน่าจะเป็นในการเกิดข้อมูลชุดใดๆ) คงที่ไม่เปลี่ยนแปลงไปตามเวลาที่ผ่านไป ดังแสดงในรูปที่ 2.1

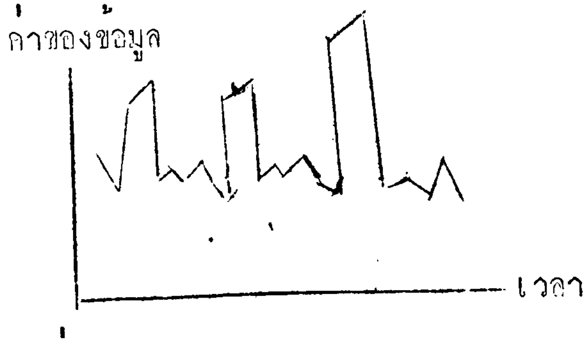
ค่าของข้อมูล

2.6.1.2 อนุกรมเวลาไม่คงที่ (Non Stationary Time Series)
ข้อมูลของอนุกรมเวลามีค่าของคุณสมบัติบางอย่างจะเปลี่ยนแปลงไปตามเวลา หมายถึงข้อมูลที่เคลื่อนไหวไม่แน่นอน เปลี่ยนแปลงไปตามระยะเวลา ดังแสดงในรูปที่ 2.2



รูปที่ 2.2 ลักษณะอนุกรมเวลาไม่คงที่

2.6.1.3 อนุกรมเวลาฤดูกาล (Seasonal Time Series)
หมายถึงข้อมูลที่มีการเคลื่อนไหวขึ้นลงตามระยะเวลาเป็นช่วงที่แน่นอน และลักษณะการเคลื่อนไหวในระยะเวลาหนึ่ง เช่น 3 เดือน 6 เดือน 1 ปี ฯลฯ จะคล้าย ๆ กันคือ จะขึ้นหรือลงเหมือน ๆ กันระยะเวลาเดียวกันในช่วงเวลาที่ซ้ำ ๆ กัน ดังแสดงในรูปที่ 2.3



2.6.2 รูปแบบทั่ว ๆ ไปของอนุกรมเวลา (A general class of time series models)

สมมติให้ $Y_1, Y_2, \ldots, Y_n$ เป็นอนุกรมเวลา

$e_1, e_2, \ldots, e_n$ เป็น random shocks หรือ residual มีการแจกแจงแบบปกติด้วยค่า Mean เท่ากับ 0 และ Variance เท่ากับ σ_e^2

รูปแบบทั่วไปของอนุกรมเวลาที่ศึกษาโดยบ็อกซ์และเจนกินส์คือ

$$W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} \cdots - \phi_p W_{t-p} = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \cdots - \theta_q e_{t-q} \cdots (1)$$

W_t คือผลต่างครั้งที่ d ของความเบี่ยงเบนของ Y_t จากค่า μ

เช่น ถ้า $d = 0$; $W_t = Y_t - \mu$

ถ้า $d = 1$; $W_t = (Y_t - \mu) - (Y_{t-1} - \mu)$

$= Y_t - Y_{t-1}$

และ $d = 2$; $W_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

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

โดยที่ $\phi_1, \phi_2, \ldots, \phi_p$ คือ พารามิเตอร์ใน Autoregressive Model

$\theta_1, \theta_2, \ldots, \theta_q$ คือ พารามิเตอร์ใน Moving Average Model

จะเรียกสมการ (1) ว่า Autoregressive Moving Average Model อันดับที่ p,d,q ARIMA (p,d,q) โดย

p คือ อันดับที่ของ Autoregressive Model

d คือ อันดับที่ของผลต่าง

q คือ อันดับที่ของ Moving Average Model

ARIMA (p,d,q) คือรูปแบบทั่ว ๆ ไป จากรูปแบบนี้ จะหารูปแบบต่าง ๆ ได้โดยกำหนดค่า p,d และ q จุดประสงค์ของการกำหนดรูปแบบก็เพื่อ

ใช้พยากรณ์โดยเทคนิคของบ๊อกซ์และเจนกินส์ เพื่อหารูปแบบของอนุกรมเวลาคงที่ รูปแบบอนุกรมเวลาไม่คงที่ และอนุกรมเวลาฤดูกาล ภายใต้ข้อสมมติที่ว่า $E(e_t) = 0$ $Var(e_t) = \sigma^2_e$, $E(e_t e_{t'}) = 0$ สำหรับ $t \neq t'$, $E(e_t Y_{t'}) = 0$ สำหรับ $t' < t$

2.6.3 รูปแบบที่สามารถใช้พยากรณ์โดยเทคนิคของบ๊อกซ์และเจนกินส์

2.6.3.1 รูปแบบคงที่ (Stationary Models)

รูปแบบคงที่ต่าง ๆ หาได้จากรูปแบบทั่วไป ARIMA(p,d,q) โดยกำหนดให้ d=0 ซึ่งพิจารณา 5 รูปแบบ ด้วยกันคือ

ก. The First Order Moving Average Model หรือMA(1)

รูปแบบนี้ได้จากการแทนค่า p = 0,d = 0หรือ q=1 ดังนั้น MA (1) ก็คือ ARIMA (0,0,1) มีสมการดังนี้คือ

$$Y_t - \mu = e_t - \theta_1 e_{t-1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

โดยที่ $|\theta_1| < 1$

ข. The Second Order Moving Average Modelหรือ MA(2)

รูปแบบนี้ได้จากการแทนค่า p = 0, d = 0 และq = 2 ดังนั้น MA(2) ก็คือ ARIMA (0,0,2)มีสมการดังนี้คือ

$$Y_t - \mu = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \qquad \ldots\ldots\ldots\ldots(3)$$

โดยที่ $\theta_2 + \theta_1 < 1$

$\theta_2 - \theta_1 < 1$

$-1 < \theta_2 < 1$

ค. The First Order Autoregressive Model หรือ AR(1)

รูปแบบนี้ได้จากการแทนค่า p = 1, d = 0 และ q = 0 ดังนั้น AR(1) ก็คือ ARIMA(1,0,0) มีสมการดังนี้คือ

$$(Y_t - \mu) - \phi_1(Y_{t-1} - \mu) = e_t \qquad \text{โดยที่ } |\phi_1| < 1$$

ง. The Second Order Autoregressive Model หรือAR(2)

รูปแบบนี้ได้จากการแทนค่า p = 2, d = 0 และ q = 0 ดังนั้น AR(2) คือ ARIMA (2,0,0) ซึ่งมีสมการดังนี้คือ

$$(Y_t-\mu)-\phi_1(Y_{t-1}-\mu)-\phi_2(Y_{t-2}-\mu) = e_t \qquad \text{.....(5)}$$

โดยที่ $\phi_2+\phi_1<1$

$\phi_2-\phi_1<1$

จ. The First Order Autoregressive-First Order Moving Average Model หรือ ARMA (1,1)

รูปแบบนี้ได้จากการแทนค่า p = 1, d = 0 และ q = 1 ดังนั้น ARMA(1,1) คือ ARIMA (1,0,1) ซึ่งมีสมการดังนี้คือ

$$(Y_t-\mu)-\phi_1(Y_{t-1}-\mu) = e_t-\theta_1 e_{t-1} \qquad \text{.....(6)}$$

โดยที่ $|\phi_1|<1$

$|\theta_1|<1$

2.6.3 <u>รูปแบบไม่คงที่</u> (Non-Stationary Models)

รูปแบบไม่คงที่ จะหาได้จากการแทนค่า d $\geq$ 1 ใน ARIMA(p,d,q) ในที่นี้จะให้ d = 1 ซึ่งพิจารณา 6 รูปแบบคือ

ก. The Integrated Moving Average Model of Order (0,1,1) หรือ IMA (0,1,1)

รูปแบบนี้ได้จากการแทนค่า p = 0, d = 1 และ q = 1 ดังนั้น IMA(0,1,1) ก็คือ ARIMA(0,1,1) มีสมการดังนี้คือ

$$Y_t-Y_{t-1} = e_t-\theta_1 e_{t-1} \qquad \text{......(7)}$$

โดยที่ $|\theta_1|<1$

ข. The Integrated Moving Average Model of Order (0,1,2) หรือ IMA (0,1,2)

รูปแบบนี้ได้จากการแทนค่า p = 0, d = 1 และ q = 2 ดังนั้น IMA(0,1,2) ก็คือ ARIMA(0,1,2) มีสมการดังนี้ คือ

$$Y_t - Y_{t-1} = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-1} \qquad \ldots\ldots (8)$$

โดยที่ $\theta_2 + \theta_1 < 1$

$\theta_2 - \theta_1 < 1$

$-1 < \theta_2 < 1$

ค. The Nonstationary First Order Autoregressive Model ARIMA (1,1,0)

รูปแบบนี้ได้จากการแทนค่า p = 1, d = 1 และ q = 0 มีสมการดังนี้

$$(Y_t - Y_{t-1}) - \phi_1 (Y_{t-1} - Y_{t-2}) = e_t \qquad \ldots\ldots (9)$$

โดยที่ $|\phi_1| < 1$

ง. The Nonstationary Second Order Autoregressive Model หรือ ARIMA (2,1,0)

รูปแบบนี้ได้จากการแทนค่า p = 2, d = 1 และ q = 0 มีสมการดังนี้

$$(Y_t - Y_{t-1}) - \phi_1 (Y_{t-1} - Y_{t-2}) - \phi_2 (Y_{t-2} - Y_{t-3}) = e_t \qquad \ldots\ldots (10)$$

โดยที่ $\phi_2 + \phi_1 < 1$

$\phi_2 - \phi_1 < 1$

$-1 < \phi_2 < 1$

จ. The First Order Autoregressive Integrated First Order Moving Average Model หรือ ARIMA (1,1,1)

รูปแบบนี้ได้จากการแทนค่า p = 1, d = 1 และ q = 1 ซึ่งมีสมการดังนี้

$$(Y_t - Y_{t-1}) - \phi_1 (Y_{t-1} - Y_{t-2}) = e_t - \theta_1 e_{t-1} \qquad \ldots\ldots (11)$$

ฉ. The Random Walk Model หรือ ARIMA (0,1,0)

รูปแบบนี้ได้จากการแทนค่า p = 0, d = 1 และ q = 0 มีสมการดังนี้

$$Y_t - Y_{t-1} = e_t \quad \ldots\ldots\ldots(12)$$

2.6.3.3 <u>รูปแบบที่มีฤดูกาล</u> (Seasonal Model)

ถ้าข้อมูลที่สนใจมีการขึ้นลงตามฤดูกาล จะหารูปแบบที่ใช้พยากรณ์อนุกรมเวลาฤดูกาล (Seasonal Time Series) ได้จาก ARIMA (p,d,q) ซึ่งแสดงผลเกี่ยวข้องกับข้อมูลที่อยู่ในช่วงเวลาที่ติดต่อกัน อาจจะเป็นเดือน เป็นปี ฯลฯ อนุกรมเวลาฤดูกาลนั้นนอกจากข้อมูลที่เกี่ยวข้องกันระหว่างเดือนแล้วข้อมูลยังเกี่ยวข้องกันระหว่างปีด้วย ซึ่งหารูปแบบได้ดังนี้ ถ้าสมมติใช้ ARIMA (0,1,1) แสดงความสัมพันธ์ของข้อมูลที่อยู่ห่างกัน 12 ช่วง หรือ 12 เดือน มีสมการดังนี้

$$Y_t - Y_{t-12} = e_t - \Theta^* e_{t-12}$$

โดยที่ $Y_t - Y_{t-12}$ หมายถึงข้อมูลที่ห่างกัน 12 ช่วง

Θ^* คือ ค่าพารามิเตอร์ใน Seasonal Moving Average Model

และถ้าสมมติ ใช้ ARIMA (0,1,1) แสดงความสัมพันธ์ของข้อมูลที่อยู่ห่างกัน 1 ช่วง หรือ 1 เดือน มีสมการดังนี้

$$a_t - a_{t-1} = \gamma_t - \Theta \gamma_{t-1}$$

โดยที่ $a_t - a_{t-1}$ หมายถึงข้อมูลที่อยู่ห่างกัน 1 ช่วง

$\gamma_t - \gamma_{t-1}$ หมายถึง random shock ที่อยู่ห่างกัน 1 ช่วง

จาก Seasonal Model ข้อมูลที่เกี่ยวข้องกันในแต่ละปี และในปีเดียวกันยังเกี่ยวข้องกันในแต่ละเดือนด้วย มีสมการดังนี้คือ

$$(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) = (e_t - \Theta e_{t-1}) - \Theta^*(e_{t-12} - \Theta e_{t-13})$$

หรือ $Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} = e_t - \Theta e_{t-1} - \Theta^* e_{t-12} + \Theta\Theta^* e_{t-13}$ ....(13)

2.6.4 <u>ขั้นตอนในการวิเคราะห์อนุกรมเวลาโดยเทคนิคของบ็อกซ์และเจนกินส์</u>

เพื่อหารูปแบบที่เหมาะสม จะแบ่งการวิเคราะห์อนุกรมเวลาโดยเทคนิคของบ็อกซ์และเจนกินส์ เป็น 4 ขั้นตอน

2.6.4.1 การกำหนดรูปแบบ (Identification)

2.6.4.2 การประมาณค่าพารามิเตอร์ (Parameter Estimation)

2.6.4.3 การตรวจสอบรูปแบบ (Diagnostic Checking)

2.6.4.4 การพยากรณ์ (Forecasting) **และการปรับปรุงค่าพยากรณ์**

2.6.4.1 การกำหนดรูปแบบ ค่าสถิติที่ใช้เป็นเครื่องมือในการเลือกรูปแบบ ก็คือค่าสัมประสิทธิ์ออโตคอริเลชั่น (Autocorrelation Coefficient) ซึ่ง...

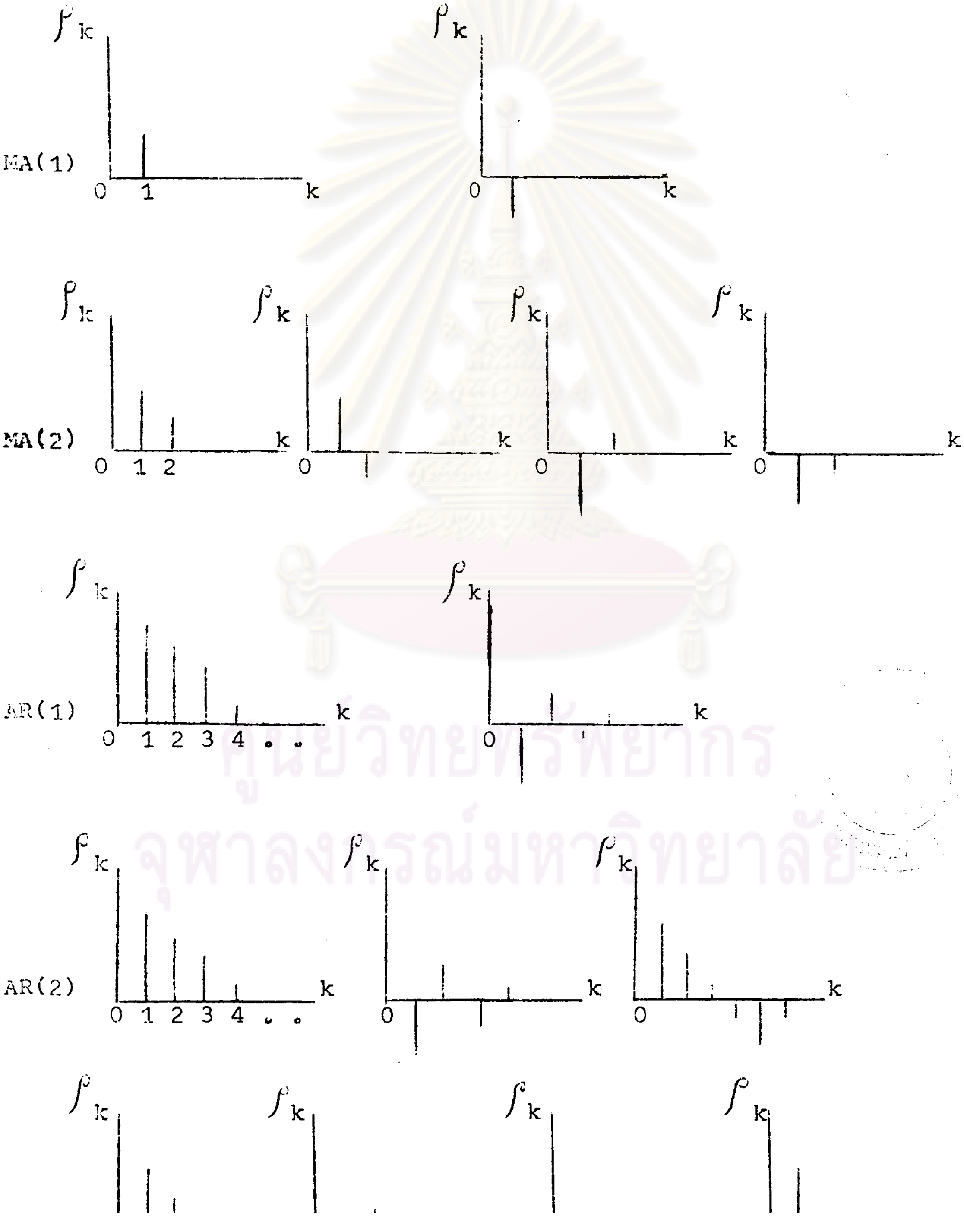
$$\rho_k = \frac{E(Y_t-\mu)(Y_{t+k}-\mu)}{Var(Y_t)} \quad , \quad k = 1,2, \quad \ldots\ldots\ldots(14)$$

$$E(Y_t-\mu)(Y_{t+k}-\mu) = \frac{\sum_{t=1}^{N-k}(Y_t-\mu)(Y_{t+k}-\mu)}{N-k}$$

$$Var(Y_t) = \frac{\sum_{t=1}^{N}(Y_t-\mu)^2}{N}$$

โดยที่ค่าสัมประสิทธิ์ออโตคอริเลชั่น ρ_k นี้จะแสดงความสัมพันธ์ของข้อมูลที่อยู่ห่างกัน k ช่วงเวลา

รูปที่ 2.4 แสดงลักษณะของ ρ_k สำหรับรูปแบบคงที่ (Stationary Models) ต่าง ๆ



ในทางปฏิบัติเมื่อต้องการกำหนดรูปแบบ ไม่สามารถจะหาค่า ρ_k ได้ จึงต้องคำนวณหาค่าสัมประสิทธิ์ออโตคอริเลชั่นของตัวอย่าง r_k แทน ρ_k ซึ่ง

$$r_k = \frac{\frac{1}{n-k}\sum_{t=1}^{n-k}(Y_t-\bar{Y})(Y_{t+k}-\bar{Y})}{\frac{1}{n}\sum_{t=1}^{n}(Y_t-\bar{Y})^2} \quad ; k = 1,2, \quad \ldots(15)$$

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

เมื่อนำค่า r_k ที่ได้ ไปพล็อตกราฟ และนำไปเปรียบเทียบกับรูปแบบมาตรฐานที่แสดงความสัมพันธ์ของ k กับ ρ_k ตามรูปที่ 2.4 ทำให้เลือกรูปแบบที่เหมาะสมได้

รูปแบบมาตรฐานที่แสดงความสัมพันธ์ ของ ρ_k กับค่าของ k สำหรับรูปแบบ AR(1), AR(2) และ ARMA(1,1) จะมีรูปแบบที่มีลักษณะคล้ายกัน ถ้าหาก r_k ที่ได้จากการพล็อตกราฟ มีรูปแบบที่เป็นไปได้ 3 รูปแบบคือ AR(1), AR(2), ARMA(1,1) จะต้องนำข้อมูลชุดนั้นไปหาประมาณค่าพารามิเตอร์ ทั้ง 3 รูปแบบ แล้วจึงนำไปตรวจสอบว่ารูปแบบไหนเหมาะสมกับข้อมูลชุดนั้น โดยวิธีทำการตรวจสอบรูปแบบที่จะกล่าวตอนต่อไป

การพิจารณาว่าข้อมูลชุดไหนเป็นอนุกรมเวลาคงที่หรือไม่นั้น จะดูได้จากค่า r_k ของข้อมูล ถ้า r_k มีค่าลดน้อยมากเมื่อ k มีค่ามาก เมื่อ k มีค่ามาก แสดงว่าข้อมูลชุดนั้นเป็นอนุกรมเวลาแบบไม่คงที่

ถ้าเป็นอนุกรมเวลาไม่คงที่ อนุกรมเวลาของผลต่างมักจะเป็นอนุกรมเวลาคงที่ จึงจำเป็นต้องพิจารณาผลต่างครั้งที่ 1 ของข้อมูลชุดนั้น ถ้าผลต่างครั้งแรกเป็นอนุกรมเวลาคงที่ ไม่จำเป็นที่จะต้องพิจารณาผลต่างครั้งที่ 2 แต่ถ้าผลต่างครั้งแรกไม่คงที่ จำเป็นต้องพิจารณาผลต่างครั้งที่ 2 ต่อไปว่าเป็นอนุกรมเวลาคงที่หรือไม่ แต่ในทางปฏิบัติมักจะหาผลต่างไม่เกิน 2 ครั้ง แล้วนำค่า r_k ของอนุกรมเวลาของผลต่างที่เป็นอนุกรมเวลาแบบคงที่ นำไปเปรียบเทียบกับรูปแบบมาตรฐานซึ่งแสดงความสัมพันธ์ของ ρ_k กับ k

2.6.4.2 การประมาณค่าพารามิเตอร์ (Parameter Estimation)

เมื่อกำหนดรูปแบบได้แล้ว จะหาค่าประมาณของพารามิเตอร์ได้จากสูตรที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับรูปแบบ ARIMA(p,d,q) ดังแสดงในตารางที่ 2.2

ตารางที่ 2.2 สูตรที่ใช้ประมาณค่าพารามิเตอร์สำหรับรูปแบบ ARIMA(p,d,q)

ลำดับ (p,d,q)	พารามิเตอร์ที่จะประมาณ	สูตรที่ใช้ในการประมาณ	ขอบเขตของค่าพารามิเตอร์
(0,d,1)	θ_1	$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$	$-1<\theta_1<1$
(0,d,2)	θ_1 และ θ_2	$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_1^2}$ $\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$-1<\theta_2<1$ $\theta_2-\theta_1<1$ $\theta_2-\theta_1<1$
(1,d,0)	ϕ_1	$\phi_1 = \rho_1$	$-1<\phi_1<1$
(2,d,0)	ϕ_1 และ ϕ_2	$\phi_1 = \frac{(1-\rho_2)}{1-\rho_1^2}$ $\phi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1}$	$-1<\phi_2<1$ $\phi_2+\phi_1<1$ $\phi_2-\phi_1<1$
(1,d,1)	ϕ_1 และ θ_1	$\rho_1 = \frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$ $\rho_2 = \rho_1\phi_1$	$-1<\phi_1<1$ $-1<\theta_1<1$

ϕ_1, ϕ_2, θ_1, และ θ_2 คือพารามิเตอร์ของรูปแบบต่าง ๆ ที่ต้องการประมาณ

ไม่สามารถจะหาได้ จึงใช้ค่า r_1 และ r_2 เป็นค่าประมาณของ ρ_1 และ ρ_2 แทน เพื่อช่วยให้การประมาณค่าพารามิเตอร์สะดวกขึ้น จึงมีตารางที่ 2.3[5] และรูปที่ 2.5 2.6, 2.7 แสดงความสัมพันธ์ของพารามิเตอร์กับค่าของ ρ_1, ρ_2 สำหรับรูปแบบ ARIMA (0,d,1), ARIMA(0,d,2), ARIMA (2,d,0) และ ARIMA (1,d,1) การประมาณของพารามิเตอร์ที่ได้จากตาราง 2.3 และรูปที่ 2.5, 2.6, 2.7 เป็นค่าประมาณเบื้องต้น ซึ่งจะต้องนำไปหาค่าประมาณค่าสุดท้าย ค่าประมาณค่าสุดท้ายที่ดีที่สุดที่จะนำเข้าไปใช้ในแต่ละรูปแบบหาได้จากค่า Sum Square ของรูปแบบนั้น ที่มีค่าน้อยที่สุด ค่า Sum Square ของรูปแบบต่าง ๆ ที่น้อยที่สุดเป็นค่า $\hat{Y}$ ที่มีส่วนเบี่ยงเบนจาก Y_t น้อยที่สุด ในทุก ๆ เวลาที่ t การคำนวณหาค่า Sum Square ที่น้อยที่สุดนี้ จะคำนวณโดยใช้เครื่องคอมพิวเตอร์ Sum Square แต่ละรูปแบบคือ

AR(1) : Sum Square $= \sum_{t=2}^{n} \left[Y_t - \mu - \phi_1(Y_{t-1} - \mu)\right]^2$

AR(2) : Sum Square $= \sum_{t=3}^{n} \left[Y_t - \mu - \phi_1(Y_{t-1} - \mu) - \phi_2(Y_{t-2} - \mu)\right]^2$

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

MA(2) : Sum Square $= \sum_{t=3}^{n} \left[Y_t - \mu - \theta_1(Y_{t-1} - \mu) + \theta_2 e_{t-1}\right]^2$

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

และในทำนองเดียวกัน ถ้าเป็น Nonstationary Model ใช้ค่าผลต่างครั้งที่ 1 หรือครั้งที่ 2 แทนค่า $Y_t - \mu$ เช่นผลต่างครั้งที่ 1 เป็น Stationary Model ใช้ค่า $Y_t - Y_{t-1}$ แทนค่า $Y_t - \mu$ เมื่อคำนวณหา Sum Square น้อยที่สุดของอนุกรมเวลา

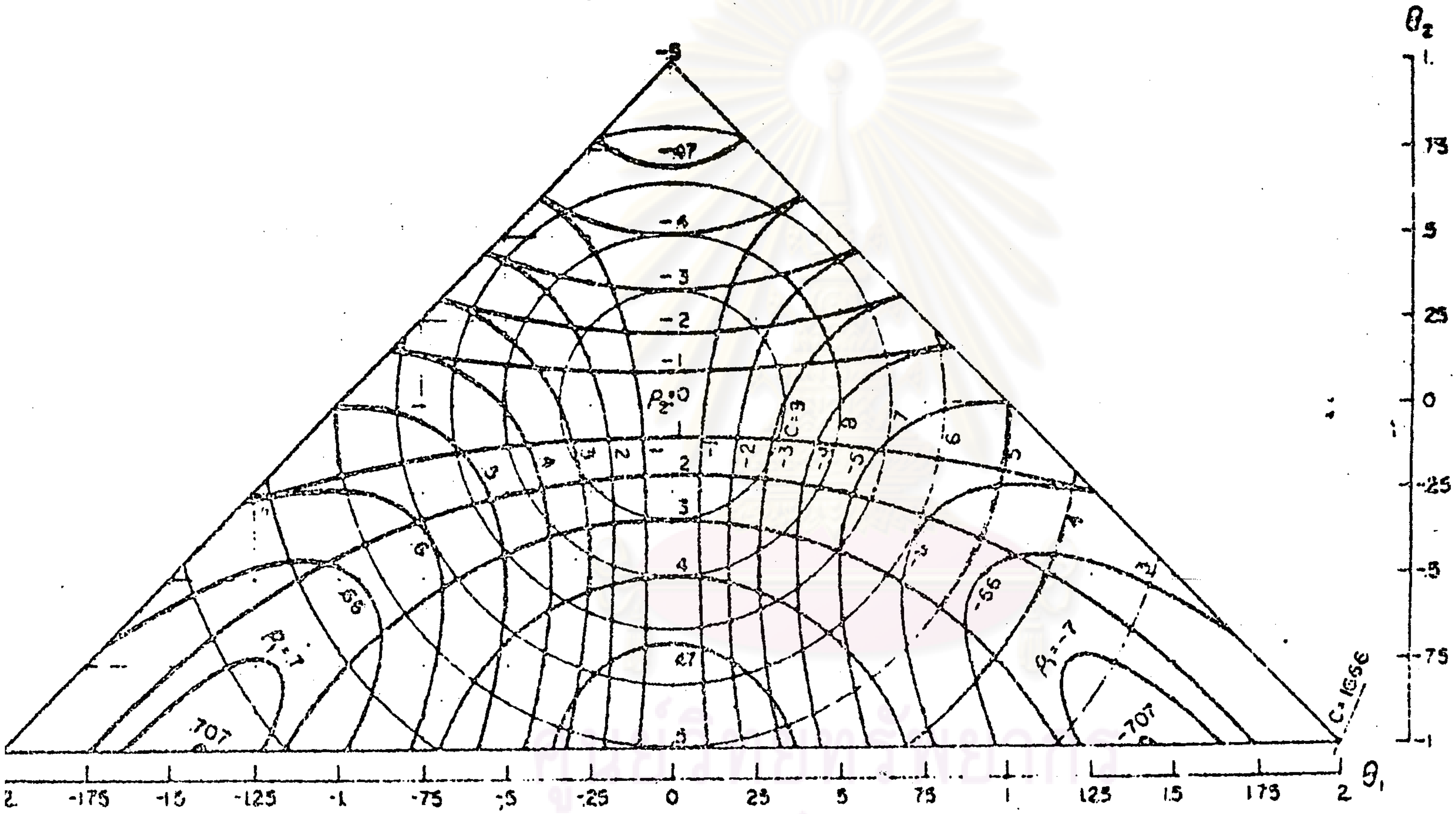
---

[5] Collection of Tables and Chart page 517-520 Box, G.E., P and G.M. Jenkins, Time Series Analysis Forecasting and Control.

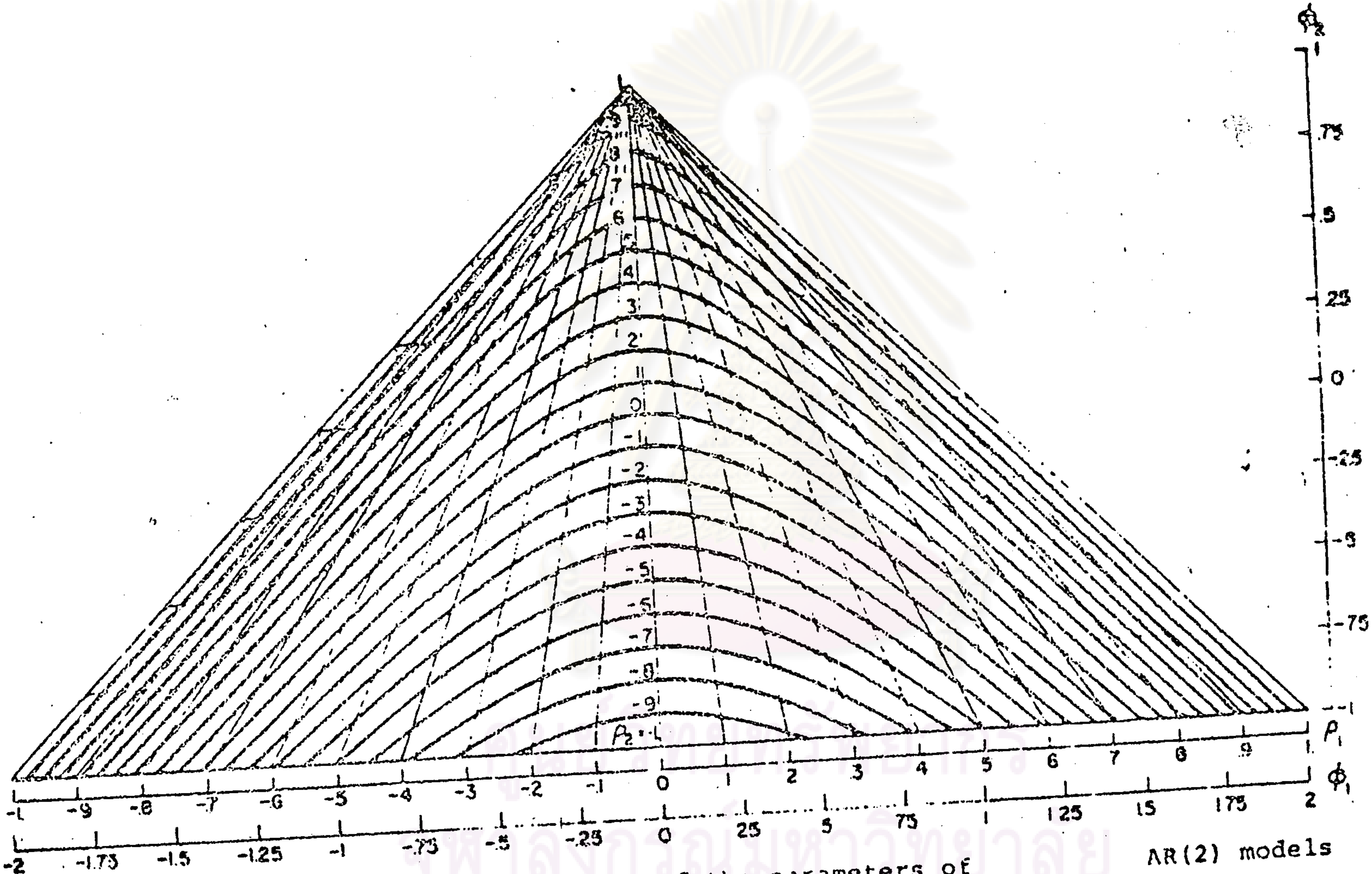
แต่ละชุด โดยการแทนค่าประมาณเบื้องต้นพารามิเตอร์และค่าเปลี่ยนแปลงของพารามิเตอร์ที่กำหนดให้ เครื่องคอมพิวเตอร์จะคำนวณค่าพารามิเตอร์ที่ให้ค่า Sum Square น้อยที่สุดของรูปแบบนั้น ๆ ให้ และค่าประมาณพารามิเตอร์ที่ได้จาก Sum Square น้อยที่สุดในแต่ละรูป คือค่าประมาณพารามิเตอร์สุดท้ายที่ใช้ในรูปแบบเพื่อที่จะได้รูปแบบที่เหมาะสม

ตารางที่ 2.3 แสดงความสัมพันธ์ระหว่าง ρ_1 กับ θ_1 สำหรับ First order moving average processes MA(1)

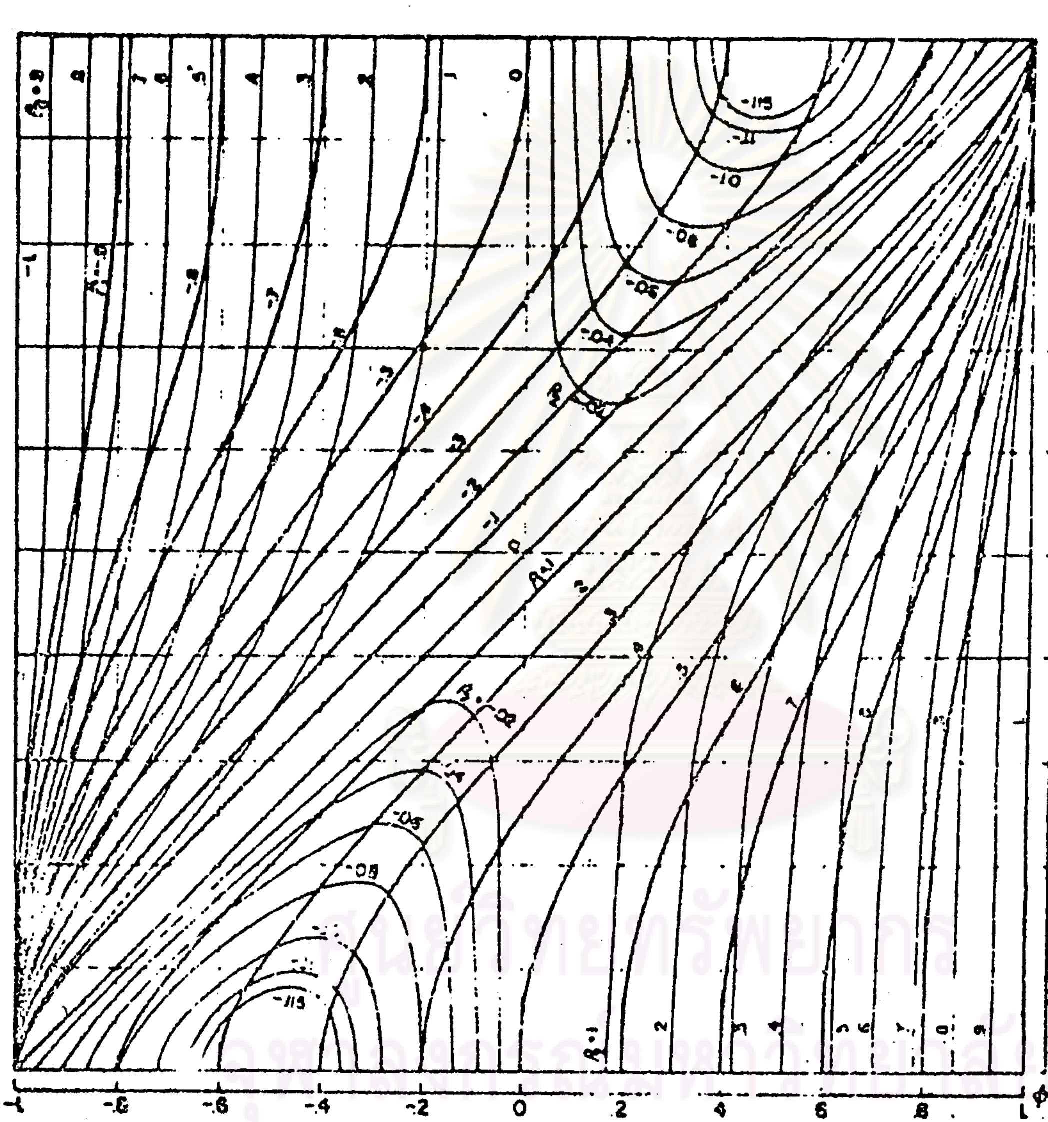
θ_1	ρ_1	θ_1	ρ_1
0.00	0.000	-0.00	0.000
0.05	-0.050	-0.05	0.050
0.10	-0.099	-0.10	0.099
0.15	-0.147	-0.15	0.147
0.20	-0.192	-0.20	0.192
0.25	-0.235	-0.25	0.235
0.30	-0.275	-0.30	0.275
0.35	-0.315	-0.35	0.315
0.40	-0.349	-0.40	0.349
0.45	-0.374	-0.45	0.374
0.50	-0.400	-0.50	0.400
0.55	-0.422	-0.55	0.422
0.60	-0.441	-0.60	0.441
0.65	-0.457	-0.65	0.457
0.70	-0.468	-0.70	0.468
0.75	-0.480	-0.75	0.480
0.80	-0.488	-0.80	0.488
0.85	-0.495	-0.85	0.495
0.90	-0.497	-0.90	0.497
0.95	-0.499	-0.95	0.499



รูปที่ 2.5 Chart to find the initial estimates of the parameters of MA(2) models



รูปที่ 2.6 Chart to find the estimates of the parameters of AR(2) models



รูปที่ 2.7 Chart to find the initial estimates of the parameters of ARMA(1,1) model.

2.6.4.3 การตรวจสอบรูปแบบ (Diagnostic Checking)

การตรวจสอบรูปแบบที่เลือกว่าถูกต้องเหมาะสมกับข้อมูลชุดนั้นหรือไม่ ค่าความคลาดเคลื่อน e_t ซึ่งเป็นผลต่างระหว่างค่าจริงของข้อมูลกับค่าคาดหมาย เท่ากับ $Y_t - \hat{Y}_t$ เป็นค่าที่มีความสำคัญมากในการตรวจสอบ

การตรวจสอบรูปแบบที่เหมาะสมทำได้ 2 วิธี

2.6.4.3.1 ตรวจสอบว่า ค่า e_t จะมีความสัมพันธ์กันหรือไม่ กล่าวคือ $r_k(\hat{e})$ เท่ากับ 0 หรือไม่ ภายใต้ข้อสมมติที่ว่า $e_t \sim N(0, \sigma_e^2)$

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

โดยที่ n = จำนวน ค่าคลาดเคลื่อน และ $k = 1,2,\ldots\ldots K$ ให้ $K \leq n$ และต้องไม่น้อยกว่า 12 ค่า

ถ้า $r_k(\hat{e})$ เท่ากับ 0 แสดงว่ารูปแบบที่ตรวจสอบถูกต้องและเหมาะสม ถ้า $r_k(\hat{e})$ มีค่าไม่เท่ากับ 0 แสดงว่ารูปแบบที่ตรวจสอบยังไม่เหมาะสมต้องหารูปแบบใหม่ ซึ่ง $r_k(\hat{e})$ จะเท่ากับ 0 หรือไม่ขึ้นอยู่กับการเปรียบเทียบ $r_k(\hat{e})$ กับค่าประมาณของส่วนเบี่ยงเบนมาตรฐานของ $r_k(\hat{e})$ ต่อค่า $\hat{\sigma}_{r_k(\hat{e})}$ ที่ระดับนัยสำคัญ α หมายความว่าถ้า $|r_k(\hat{e})| \leq Z_{\alpha/2}\,\hat{\sigma}_{r_k(\hat{e})}$ ทุกค่า k แสดงว่ายอมรับว่า $r_k(\hat{e}) = 0$, ใช้ $\alpha = .05$

$|r_k(\hat{e})| \leq 1.96\,\hat{\sigma}_{r_k(\hat{e})}$ แสดงว่ายอมรับ $r_k(\hat{e}) = 0$ โดยที่ $k = 1,2,\ldots K$

บางครั้งในทางปฏิบัติ ใช้ 2 เป็นค่าประมาณแทนค่า 1.96 สำหรับที่ระดับนัยสำคัญ $\alpha = .05$ เนื่องง่ายและสะดวกต่อการคำนวณ

$$\hat{\sigma}_{r_k(\hat{e})} = \sqrt{Var(r_k(\hat{e}))}$$

ตารางที่ 2.4 แสดงค่า $\hat{\sigma}_{r_k(\hat{e})}$ ในแต่ละรูปแบบ

$Var[r_k(\hat{e})]$	ARIMA(1,d,0)	ARIMA(2,d,0)	ARIMA(0,d,1)	ARIMA(0,d,2)	ARIMA(1,d,1
k=1	$\frac{\phi_1^2}{n}$	$\frac{\phi_2^2}{n}$	$\frac{\theta_1^2}{n}$	$\frac{\theta_2^2}{n}$	$\frac{\theta_1^2}{n}$
k=2	$\frac{1-\phi_1^2+\phi_1^4}{n}$	$\frac{\phi_2^2+\phi_1^2(1+\phi_2)^2}{n}$	$\frac{1-\theta_1^2+\theta_1^4}{n}$	$\frac{\theta_2^2+\theta_1^2(1+\theta_2)^2}{n}$	$\frac{1-\theta_1^2+\theta_1^4}{n}$
k=3	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$

2.6.4.3.2 ตรวจสอบรูปแบบโดยใช้ ไค-สแควร์ (χ^2-test)

รูปแบบที่กำหนดจะเหมาะสมถ้า $Q \leq \chi^2_{(1-\alpha),K-a}$ และรูปแบบที่กำหนดจะไม่เหมาะสม ถ้า $Q > \chi^2_{(1-\alpha),K-a}$ โดยที่

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

$K \leq \sqrt{n}$ โดยที่ k ต้องไม่น้อยกว่า 12 ค่า

a = จำนวนพารามิเตอร์ในรูปแบบที่ต้องการทดสอบ

n = จำนวนข้อมูลทั้งหมด

2.6.4.4 การพยากรณ์จากรูปแบบที่กำหนดโดยเทคนิคของบ็อกซ์และเจนกินส์

เมื่อได้รูปแบบที่เหมาะสม โดยเทคนิคของบ็อกซ์และเจนกินส์แล้ว การพยากรณ์ตามรูปแบบที่เหมาะสมนั้น ต้องทราบค่าเริ่มต้นของการพยากรณ์ (Forecast Origin) ด้วย ซึ่งปกติจะใช้ช่วงเวลา n คือ ช่วงเวลาสุดท้ายของอนุกรมเวลาที่มีอยู่เป็นค่าเริ่มต้นพยากรณ์ช่วงเวลาที่ n+1, n+2, ...จะทำการพยากรณ์ล่วงหน้าไว้กี่ช่วงก็ได้ แต่ตามปกติไม่นิยมพยากรณ์ไว้หลายช่วง เพราะเมื่อค่าในช่วง n+1 เกิดขึ้นจริง ๆ นำค่า n+1 ที่เกิดขึ้นจริงนี้ เป็นค่าเริ่มต้นของการพยากรณ์ในรูปแบบนั้น เพื่อหา

แต่ในวิทยานิพนธ์นี้จะไม่ทำการปรับปรุงค่าพยากรณ์ และค่าพยากรณ์ที่นำมาใช้ คือค่าพยากรณ์ที่เป็นค่าเดียว (Point Forecast Value)

2.7 การหาค่าพยากรณ์วิธีการต่าง ๆ ซึ่งจะนำมาเปรียบเทียบ

2.7.1 การพยากรณ์โดยใช้ค่าแนวโน้ม (T) จากข้อมูลเดิม ซึ่งหาได้จาก วิธีการหาแนวโน้ม ที่กล่าวถึงในตอนที่ 2.2

2.7.2 การพยากรณ์โดยใช้ค่าแนวโน้มและปรับด้วยดัชนีฤดูกาล (TxS) ซึ่งหาได้จากวิธีการหาค่าแนวโน้ม ที่กล่าวถึงในตอนที่ 2.2 และการหาดัชนีฤดูกาลในตอนที่ 2.3

2.7.3 การพยากรณ์โดยใช้ค่าแนวโน้มปรับด้วยค่าดัชนีฤดูกาลและปรับด้วยค่าวัฏจักร และค่าเหตุการณ์ผิดปกติโดยไปแยกค่าวัฏจักรออกจากเหตุการณ์ผิดปกติ (TxSxCI) การหาค่าแนวโน้ม ตามวิธีที่กล่าวถึงในตอน 2.2 การหาดัชนีฤดูกาลตามวิธีที่กล่าวถึงในตอน 2.3 การหาค่าวัฏจักรรวมเหตุการณ์ผิดปกติหาได้จากการกำหนดรูปแบบโดยเทคนิคของบ็อกซ์และเจนกินส์ตามวิธีที่กล่าวถึงในตอน 2.6

2.7.4 การพยากรณ์โดยใช้ค่าแนวโน้ม ปรับด้วยค่าดัชนีฤดูกาล และปรับด้วยค่าวัฏจักร และเหตุการณ์ผิดปกติ โดยแยกพยากรณ์ค่าวัฏจักรและเหตุการณ์ผิดปกติ (TxSxCxI) การหาค่าแนวโน้มตามวิธีที่กล่าวถึงในตอน 2.2 การหาดัชนีฤดูกาลตามวิธีที่กล่าวถึงในตอน 2.3 การหาค่าวัฏจักรและเหตุการณ์ผิดปกติ ตามวิธีที่กล่าวถึงในตอนที่ 2.4, 2.5 และ 2.6

2.7.5 การพยากรณ์โดยใช้เทคนิคของบ็อกซ์และเจนกินส์จากข้อมูลเดิม โดยนำข้อมูลเดิมมากำหนดรูปแบบหาค่าพยากรณ์ตามวิธีที่กล่าวถึงในตอน 2.6

2.8 เกณฑ์ (Criterion) ที่จะใช้ในการเปรียบเทียบวิธีการพยากรณ์จากวิธีต่าง ๆ ที่กล่าวถึง

ในตอนที่ 2.7 คือ ค่า $\Sigma(Y-\hat{Y})^2$ หมายถึงผลรวมของผลต่างระหว่างค่าจริง

ตารางที่ 3.1 แสดงจำนวนนักท่องเที่ยวที่เข้ามาท่องเที่ยวในประเทศไทย ตั้งแต่ปี พ.ศ.2506 - 2520

เดือน / ปี	ม.ค.	ก.พ.	มี.ค.	เมย.	พ.ค.	มิย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.
2506	14,532	13,085	19,589	19,886	14,483	14,717	17,659	15,440	13,303	17,384	18,113	16,585
2507	15.320	15,600	18,420	18,899	17,211	14,783	17,904	18,303	16,592	20,464	19,736	18,692
2508	14,363	15,826	18,918	18,779	18,466	16,182	18,289	20,134	18,782	21,703	21,665	21,918
2509	16,592	17,706	23,601	23,628	23,241	21,248	27,880	26.031	19,936	29,744	29,274	26,236
2510	27,350	25,794	31,365	23,500	28,131	20,057	24,321	26,100	29,820	33,253	30,466	35,688
2511	29,590	28,927	29,305	32,613	30,695	26,765	33,601	35,599	28,393	34,604	32,471	34,699
2512	32,960	32,629	39,460	41,946	36,216	35,707	41,766	39,381	37,798	45,520	44,081	42,320
2513	41,866	42,879	50,862	55,149	57,138	49,943	62,984	62,613	51,476	47,179	49,335	57,247
2514	48,606	46,997	53,030	52,999	50,360	45,034	52,793	54,924	49,732	60,356	62,850	61,057
2515	59,010	62,291	66,905	62,642	60,590	57,109	67,618	69,963	62,701	77,723	86,818	87,388
2516	79,188	78,066	83,155	87,838	76,332	72,966	93,127	96,980	77,557	89,635	97,296	105,597
2517	108,276	86,579	90,857	94,793	80,960	77,240	85,865	95,044	79,431	97,810	101,065	109,472
2518	100,839	101,526	108,163	95,829	95,341	78,771	88,901	103,075	101,721	93,714	99,641	112,554
2519	97,132	95,560	97,518	98,584	82,414	74,611	87,460	96,547	79,732	90,248	97,964	100,672
2520	97,706	97,278	98.192	92,491	92,579	82,942	99,044	110,095	92,622	110,273	124,491	122,959