บทที่ 3
กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 แนวความคิดเกี่ยวกับการพยากรณ์ข้อมูลอนุกรมเวลา

การวิเคราะห์อนุกรมเวลา (Time series analysis)

การวิเคราะห์นี้ต้องการแยกความเคลื่อนไหวต่างๆ ออกจากข้อมูลอนุมกรมเวลา การวิเคราะห์การเปลี่ยนแปลงเคลื่อนไหวของอนุกรมเวลาต่างๆ พบว่าการเปลี่ยนแปลงเคลื่อนไหวในอนุกรมเวลามักจะประกอบไปด้วย 4 ความเคลื่อนไหว

1. ค่าแนวโน้ม (trend : T)

ค่าแนวโน้มเป็นการเคลื่อนไหวในระยะเวลาที่ค่อนข้างจะยาวนาน ค่าแนวโน้มปกติแสดงถึงทิศทางอนุกรมเวลาชุดนั้นๆ มุ่งไปสู่ ค่าแนวโน้มอาจมีลักษณะเป็นเส้นตรง เส้นโค้ง หรือลักษณะอื่นใดก็ได้

2. การเคลื่อนไหวตามฤดูกาล (seasonal movement: S)

การเคลื่อนไหวประเภทนี้ เคลื่อนไหวขึ้นๆ ลงๆ ซึ่งในเวลาเดียวกัน กล่าวคือเคยสูงเคยต่ำในระยะเวลาใดก็มักจะสูงต่ำในระยะเวลานั้นต่อไป อิทธิพลของฤดูกาลนี้โดยปกติจะเกิดขึ้นซ้ำๆ ในลักษณะคล้ายกันทุกปี การเคลื่อนไหวนี้มักจะแสดงในลักษณะสัมพันธ์ คือเป็นจำนวนเปอร์เซ็นต์หรือเรียกว่าดัชนีฤดูกาล (seasonal index)

3. การเคลื่อนไหวตามวัฏจักร (cyclical movement: C)

เป็นการเคลื่อนไหวแบบขึ้นๆ ลงๆ ในระยะเวลานานกว่า 1 ปี การเปลี่ยนแปลงตามวัฏจักรโดยทั่วไปแบ่งออกเป็น 4 ระยะคือ

ระยะที่ 1 เป็นระยะฟื้นตัวหรือขยายตัว

ระยะที่ 2 เป็นระยะที่รุ่งเรือง

ระยะที่ 3 เป็นระยะหดตัว

ระยะที่ 4 เป็นระยะตกต่ำและในเวลาเดียวกันก็จะฟื้นตัวไประยะขยายตัว

4. ความเคลื่อนไหวผิดปกติ (irregular movement: I)

เป็นการเคลื่อนไหวเพิ่มขึ้นหรือลดลงโดยผิดปกติ เคลื่อนไหวอย่างมีลักษณะไม่แน่นอนโดยมีสาเหตุที่ทำให้เกิดขึ้น โดยไม่มีใครคาดหมาย หรือไม่อาจคาดการณ์ได้ล่วงหน้า เช่น ไฟไหม้ น้ำท่วม การนัดหยุดงาน สงครามโลก เป็นต้น

การทดสอบความนิ่งของข้อมูล (Unit Root Test)

แนวคิดการทดสอบความนิ่งของข้อมูล (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542) สามารถทดสอบได้โดยใช้การทดสอบ DF (Dicky-Fuller (DF) test) (Dicky and Fuller, 1981) และการทดสอบ ADF (Augmented Dicky-Fuller (ADF) test) (Said and Dickey 1984) โดยสมมุติฐาน (null hypothesis) ของการทดสอบ DF (DF test) คือ $H_0 : \rho = 1$ จากสมการ (1) ด้านล่าง

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{3.1}$$

ถ้า $|\rho| < 1$ X_t จะเป็น stationary data คือข้อมูลจะมีลักษณะนิ่ง และถ้า $\rho = 1$ แล้ว X_t จะเป็น non-stationary data อย่างไรก็ตามเราสามารถทดสอบได้อีกทางหนึ่งกล่าวคือ จากสมการ (1)

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \tag{3.2}$$

สมมุติฐานของการทดสอบคือ $H_0 : \theta = 0$ และ $H_a : \theta < 0$ โดย $\rho = (1 + \theta)$ จากสมการที่ (1) และสมการที่ (2) จะได้ $X_t = (1+\theta) X_{t-1} + \varepsilon_t$ ถ้า θ ในสมการ (2) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (1) จะมีค่าน้อยกว่า 1 ดังนั้นเราจะปฏิเสธ $H_0 : \theta = 0$ ซึ่งเป็นการยอมรับ $H_a : \theta < 0$ หมายความว่า $\rho < 1$ และ X_t มี integration of order zero (Charemza and Deadman, 1992 : 131) นั่นคือข้อมูลมีความนิ่งคือมีความเป็น stationary data แต่ถ้าเรายอมรับ $H_0 : \theta = 0$ ได้ X_tจะเป็น non-stationary data

ถ้า X_t เป็นข้อมูลแบบแนวเดินเชิงสุ่มและมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) เราสามารถเขียนแบบจำลองได้ดังนี้

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{3.3}$$

และถ้า X_t มีแนวโน้มตามเวลาเชิงเส้น (linear time trend) เราสามารถเขียนแบบจำลองได้ดังนี้

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t \qquad \text{โดยที่ } t = \text{เวลา} \tag{3.4}$$

ซึ่งก็จะใช้สมมุติฐานในการทดสอบคือ $H_0 : \theta = 0$ โดยมี $H_a : \theta < 0$ โดยสรุปแล้วตัวแปรที่เราสนใจ คือ θ ถ้า $\theta = 0$; X_t จะเป็น stationary data โดยการเปรียบเทียบค่าสถิติ t (t-statistic) ที่คำนวณได้กับค่าที่เหมาะสมในตาราง Dickey-Fuller (Dickey-Fuller Tables) (Enders, 1995 : 221) หรือกับค่า วิกฤติ MacKinnon (MacKinnon critical values) (Gujarati, 1995 : 769)

ในกรณีที่ข้อมูลเป็น serial correlation ใน error term (ε_t) คือมีการเพิ่ม lagged change เราจะได้สมการเป็น

$$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.5}$$

$$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.6}$$

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.7}$$

(Enders, 1995 : 221 ; Gujarati, 1995 : 720)

จำนวนของ lagged change จะต้องมีมากพอที่จะทำให้ไม่เกิด autocorrelation ในส่วนของ error terms เราเรียกการทดสอบความนิ่งของข้อมูลในสมการ (5) – (7) ว่าการทดสอบ ADF (augmented Dickey – Fuller (ADF) test) ค่าสถิติทดสอบ ADF (ADF test statistic) มีการแจกแจงเชิงเส้นกำกับ (asymptotic distribution) เหมือนกับสถิติ DF (DF statistic) ดังนั้นก็สามารถใช้ค่าวิกฤติ (crititcal values) แบบเดียวกันได้ (Gujarati, 1995 ; 720)

การเลือก lag length (p-lag) ที่เหมาะสมนั้น ควรเริ่มต้นจาก lag length ที่สูงพอแล้วก็ลดขนาดของ lag length ลงโดยใช้ค่าสถิติทดสอบ t (T-test) ถ้าสัมประสิทธิ์ lag length ที่ p*นั้น ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติ เราก็จะทำการทดสอบ unit root ของตัวแปรนั้นโดยใช้ lag length ที่ p*-1 จน lag length ที่ใช้นั้นจะแตกต่างไปจากศูนย์อย่างมีนัยสำคัญทางสถิติ (Enders, 1995; 277)

การทดสอบความนิ่งแบบเป็นฤดูกาลของข้อมูล (Seasonal Unit Root Test)

แบบจำลองที่ใช้ในการศึกษาของข้อมูลอนุกรมเวลา สำหรับทดสอบว่าตัวแปรแต่ละตัวมีลักษณะนิ่ง (stationary) หรือไม่ โดยความนิ่งที่ทดสอบนั้นจะมีด้วยกัน 3 แบบ คือ ความนิ่งแบบมาตรฐาน (standard unit root) ความนิ่งแบบฤดูกาลเป็นรายครึ่งปี (semiannual root) และความนิ่งตามฤดูกาลแบบรายไตรมาส (seasonal root at the quarterly frequency) โดยใช้แบบจำลองของ Hylleberg et al. 1990 (Patterson, 2000) ดังนี้

$$(1 - L^4)X_t = \gamma_1 X_{1t-1} - \gamma_2 X_{2t-1} + \gamma_3 X_{3t-1} \gamma_4 X_{3t-2} + \varepsilon_t$$

โดยสมมุติฐานว่าง (null hypothesis) ของการทดสอบว่ามีความไม่นิ่งแบบมาตรฐาน H_0 : $\gamma_1 = 0$ ถ้าการทดสอบพบว่า $\gamma_1 \neq 0$ (ปฏิเสธสมมุติฐานว่าง) แสดงว่า $(1-L^4)X_t$ มีลักษณะนิ่งแบบมาตรฐาน ถัดไปคือการทดสอบความนิ่งแบบฤดูกาลเป็นรายครึ่งปี โดยสมมุติฐานว่าง คือ H_0 : $\gamma_2 = 0$ และถ้าจากการทดสอบพบว่า $\gamma_2 = 0$ (ยอมรับสมมุติฐานว่าง) แสดงว่า $(1-L^4)X_t$ มีลักษณะไม่นิ่งแบบรายครึ่งปีและสุดท้ายคือการทดสอบความนิ่งแบบรายปี โดยใช้การทดสอบ F-test สมมุติฐานว่างคือ $\gamma_3 = \gamma_4 = 0$ ถ้าจากการทดสอบได้ค่า F-test น้อยกว่าค่าอาณาเขตวิกฤต(ยอมรับสมมุติฐานว่าง) แสดงว่า $(1-L^4)X_t$ มีลักษณะไม่นิ่งแบบรายปี โดยตัวแปร X_{1t-1}, X_{2t-1} , X_{3t-1} และ X_{3t-2} หาได้จาก

$$X_{1t-1} = (1+L+L^2+L^3)X_{t-1} = X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4}$$
$$X_{2t-1} = -(1-L+L^2-L^3)X_{t-1} = X_{t-1} - X_{t-2} + X_{t-3} - X_{t-4}$$
$$X_{3t-1} = -(1-L^2)X_{t-1} = X_{t-1} - X_{t-3}$$
$$X_{3t-2} = (1-L^3)X_{t-1} = X_{t-2} - X_{t-4}$$

หมายเหตุ: L คือ lag operater

ซึ่งค่าสถิติที่ใช้ทดสอบจะใช้ของ Hylleberg et al. 1990 (Patterson, 2000 : 276)

ตาราง 3.1 ค่าสถิติของ Hylleberg et al.1990

H_0 :	$\gamma_1 = 0$	$\gamma_2 = 0$	$\gamma_3 = \gamma_4 = 0$
H_a :	$\gamma_1 \neq 0$	$\gamma_2 \neq 0$	$\gamma_3 \neq 0, \gamma_4 \neq 0$
Intercept	-2.89	-1.91	3.00
Intercept plus seasonal dummies	-2.94	-2.90	7.66
Intercept plus seasonal dummies plus time trend	-3.52	-2.93	6.62

3.2 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลาโดยวิธี Box-Jenkins

เมื่อข้อมูลของเรามีความนิ่งแล้ว เราก็จะทำการพยากรณ์โดยวิธี Box-Jenkins ซึ่งมี 4 ขั้นตอนด้วยกัน

1. การกำหนดแบบจำลอง (identification) เป็นการหารูปแบบ ARIMA(p,q) ที่คาดว่าเหมาะสมให้กับอนุกรมเวลาโดยที่ autocorrelation : ρ_k คือการวัดความสัมพันธ์ของแต่ละช่วงเวลาโดยมีช่วงที่ย้อนหลังไป k หน่วยเวลา โดยที่ $|\rho|<1$ โดยพิจารณาเปรียบเทียบค่า autocorrelation (R_k) ของอนุกรมตัวอย่าง กับค่า autocorrelation (ρ_k) ของอนุกรมเวลาที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา

$$R_k = \frac{\sum_{t-a}^{n-k}(X_{t-q})(X_{t+k-q})}{\sum_{t-a}^{n}(X_{t-q})^2} \tag{3.8}$$

โดยที่ $X_t = \sum_{t-a}^{n}(X_t)$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

partial autocorrelation : ρ_{kk} คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลาโดยพิจารณาเปรียบเทียบค่า partial autocorrelation (R_{kk}) ของอนุกรมเวลาตัวอย่างกับค่า partial autocorrelation (ρ_{kk}) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$R_{kk} = \frac{\sum_{j=1}^{k-1}(R_{k-1,j})(R_{k-j})}{\sum_{j=1}^{k-1}(R_{k-1,j})(R_j)} \tag{3.9}$$

การพิจารณาเปรียบเทียบแต่ละรูปแบบ ต้องพิจารณา R_k , R_{kk} กับ P_k , P_{kk} พร้อมกันหลายๆ ค่า จึงมักจะพิจารณาจากรูปที่เรียกว่าคอเรลโลแกรม ที่ได้จากการพล็อต R_k , R_{kk} กับ P_k , P_{kk} ในช่วงเวลาที่ k ดังนั้นการพิจารณาการเปรียบเทียบ จะเป็นการเปรียบเทียบคอเรลโลแกรมของค่า autocorrelation ของอนุกรมเวลาตัวอย่าง (R_k) กับค่า autocorrelation ของอนุกรมเวลาของประชากร (P_k) และคอเรลโลแกรมของค่า partial autocorrelation ของอนุกรมเวลาตัวอย่าง (R_{kk}) กับค่า partial autocorrelation ของอนุกรมเวลาประชากร (P_{kk}) สำหรับแต่ละรูปแบบจะมีคอเรลโลแกรม ของ P_k และ P_{kk} ต่างกัน อนุกรมเวลาที่จะนำมากำหนดรูปแบบจะต้องเป็นอนุกรมเวลาที่เป็น stationary เท่านั้น หากไม่เป็น stationary จะต้องแปลงให้เป็น stationary เสียก่อน

วิธีการวิเคราะห์ด้วย Box-Jenkins เป็นการวิเคราะห์ด้วยวิธีทางอนุกรมเวลา เพื่อหาสมการที่มีรูปแบบเหมาะสม โดยใช้ค่า autocorrelation function (ACF) partial autocorrelation function (PACF) เป็นหลักในการพิจารณารูปแบบที่เลือกใช้จะอยู่ในกลุ่มของรูปแบบ ARIMA (p,d,q) หรือเรียกว่า auto regressive – moving average order p and q ซึ่งเป็นรูปแบบที่ใช้ในการพยากรณ์ค่าในอนาคต โดยจะแบ่งเป็นรูปแบบ AR(p) และ MA(q) เข้าด้วยกัน รูปแบบ AR(p) จะหมายถึง ค่าที่สังเกต Y จะขึ้นอยู่กับ Y ก่อนหน้าเท่ากับ p และ MA(q) นั้นก็คือค่าที่สังเกต Y จะขึ้นอยู่กับค่าความคาดเคลื่อนก่อนหน้า เท่ากับ q ซึ่งก็จะได้รูปแบบ ARIMA(p,q) โดยที่จะกำหนดรูปแบบดังนี้

AR (p) คือ $X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t$

MA (q) คือ $X_t = \alpha + \varepsilon_t - \theta_t \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$

AR MA (p,q) คือ $X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_t \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$

การกำหนด p, q ในแบบจำลอง (identifying the dependence order of model) ขั้นตอนคือการระบุว่าแบบจำลองนี้ควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใด และ moving average, q เท่าใด โดยพิจารณาจาก ACF และ PACF

ตาราง3.2 การพิจารณา ACF และ PACF

ชนิดของแบบจำลอง	รูปแบบของ ACF	รูปแบบของ PACF
AR (p)	ลู่โค้งเข้าหาแกน (tails off)	เกิดค่าที่ชัดเจนเพียง p ค่าแล้วหายไป (cut off after lag p)
MA (q)	เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป (cut off after lag p)	ลู่โค้งเข้าหาแกน (tails off)
ARMA (p,q)	ลู่โค้งเข้าหาแกน (tails off)	ลู่โค้งเข้าหาแกน (tails off)

ที่มา : Gujarati (2003)

จากตาราง เมื่อคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ คอเรลโลแกรม PACF จะมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมาให้นับเป็นค่าที่ p ของ AR (p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรม เกิดขึ้น 1 แท่ง คือแบบจำลองควรมีลักษณะเป็น AR(1) สำหรับ MA(q) นั้นก็จะมี AFC ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบ ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สามารถสรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(2) และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรจะเป็น ARMA (p,q) และเมื่อรวมกันกับการทดสอบ stationary ในขั้นตอนที่ 1 แล้ว จะสามารถหาค่าของ difference ได้ ซึ่งผลจากการ difference จำนวน d ครั้งนั้น ก็จะได้แบบจำลอง ARIMA(p,d,q)

2. การประมาณค่าพารามิเตอร์ในรูปแบบ (estimation) จะทำได้โดยการสร้างสมการที่ได้มาจากความสัมพันธ์ระหว่าง P_k และพารามิเตอร์ โดยสมการที่สร้างขึ้นจะมีจำนวนพารามิเตอร์ที่ต้องการประมาณ ส่วนค่าประมาณที่ได้จากการวิเคราะห์จะได้จากการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด ขั้นตอนของการวิเคราะห์จะต้องมีการกำหนดค่าประมาณเริ่มต้น เมื่อการวิเคราะห์สิ้นสุดจะใช้ค่าประมาณที่ได้ที่นำไปใช้ประโยชน์ในการสร้างสมการพยากรณ์ โดยจะพิจารณาจาก

adjusted R^2 คือการวัดค่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใดหากค่านี้เท่ากับ 1 หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% ถ้าค่านี้มีค่าเท่ากับ 0 แสดงว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย

$$\text{adjusted } R^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1))} = 1 - (1 - R^2)\frac{n-1}{n-k} \qquad (3.10)$$

Durbin-Watson statistic (DW) ใช้ทดสอบ autocorrelation ว่ามีสหสัมพันธ์ในตัวเองหรือไม่ โดยถ้าค่า DW เข้าใกล้ 2 ก็จะถือว่าไม่มีสหสัมพันธ์ในตัวเอง

Akaike Information Criterion (AIC) คือค่าที่อธิบายว่า สมการที่ได้ มีค่าความคลาดเคลื่อนของผลการพยากรณ์มากน้อยแค่ไหน โดยหากค่านี้ยิ่งมีค่าน้อย แบบจำลองที่ได้นั้นยิ่งสามารถเป็นตัวแทนข้อมูลจริงได้ดี

$$AIC = \left(\frac{2k}{n}\right) + \log\left(\frac{\sum \hat{u}_i^2}{n}\right) \qquad (3.11)$$

กำหนดให้ $\sum \hat{u}_i^2$ = ผลรวมของค่าความคลาดเคลื่อน

n = ค่าสังเกตทั้งหมด

Schwarz's Bayeian Information Criterion (SBC) คือการวัดค่าความคลาดเคลื่อนของผลการพยากรณ์มากน้อยแค่ไหนอีกวิธีหนึ่ง คล้ายกับวิธี AIC โดยมีสมการเป็นดังนี้

$$SBC = \log\left(\frac{\sum \hat{u}_i^2}{n}\right) + \left(\frac{2k \log n}{n}\right) \qquad (3.12)$$

จากค่า R^2 และ AIC จะนำมาใช้การพิจารณาเลือกแบบจำลอง ARIMA (p, d, q) ที่เหมาะสม เพื่อใช้ในการคัดเลือกในขั้นตอนที่ 3 ต่อไป

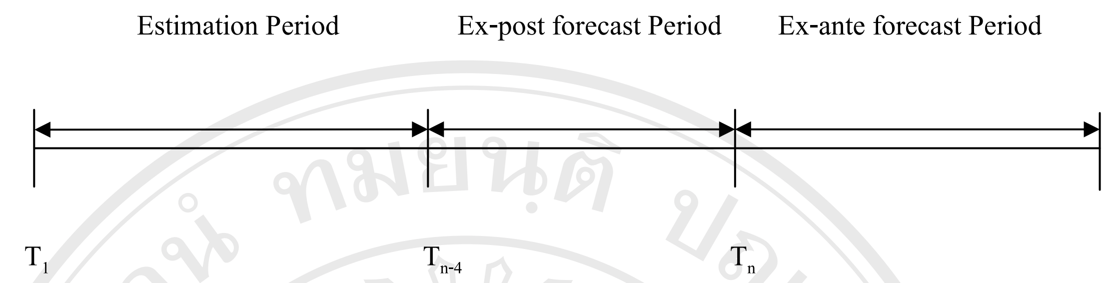
3. การตรวจสอบแบบจำลอง (diagnostic checking) ในการตรวจสอบความถูกต้องของแบบจำลองนั้นจะพิจารณาคุณสมบัติความเป็นเชิงสุ่ม(white Noise) ของค่าประมาณความคลาดเคลื่อน(estimated residual, $\hat{e}_t$) โดยใช้ค่า Q-statistic ของ Box-Pierce ซึ่งกำหนดสมมุติฐาน $H_0: P_1(\hat{e}_t) = P_2(\hat{e}_t) = \ldots = P_k(\hat{e}_t) = 0$ ถ้าค่า Q-statistic ของอนุกรมเวลาไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ 0.10 แสดงว่า $\hat{e}_t$ มีคุณสมบัติความเป็นเชิงสุ่ม(white noise) หรือมีการกระจายแบบปกติ (normal distribution) มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนเท่ากับ $\sigma^2 I\left[\hat{e}_t \sim NID(0,\sigma^2 I)\right]$ แสดงว่า $\hat{e}_t$ ไม่มีสหสัมพันธ์ในตัวเอง และมีความแปรปรวนไม่แตกต่างกัน (heteroscedasticity) หมายความว่าอนุกรมเวลาดังกล่าวได้ผ่านการวินิจฉัยและมีความเหมาะสมที่จะใช้ในการพยากรณ์ต่อไป แต่หากพบว่าแบบจำลองที่ได้ไม่เหมาะสมจะต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

$$\text{Q-statistic} = N\sum_{k=1}^{m}\rho_k^2 \qquad (3.13)$$

โดย N คือ จำนวนของข้อมูล

m คือ ค่า lag length

4. การพยากรณ์ (forecasting) ในการพยากรณ์นั้นจะใช้สมการที่สร้างจากรูปแบบการพยากรณ์ที่กำหนด และผ่านการตรวจสอบในขั้นตอนที่ผ่านมาแล้ว โดยพิจารนาจากค่า root mean squared error และค่า Theil's inequality coefficient ที่มีค่าต่ำสุด ซึ่งการพยากรณ์จะพิจารณาได้เป็น 3 ช่วงคือ



รูป 3.1 ช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา : Pindyck and Rubinfeld (1997)

ช่วงที่ 1 Estimation Period คือช่วง T_1 ถึงช่วง T_{n-4} เป็นการพยากรณ์เพื่อเปรียบเทียบกับค่าข้อมูลจริง ในการเลือกรูปแบบที่มีความเหมาะสม จะพิจารณาจากค่า Theil's inequality coefficient ที่มีค่าต่ำสุด และค่า root mean squared error ที่ดีที่สุด

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(X_t^s - X_t^a\right)} \tag{3.14}$$

โดยให้ X_t^s = ค่าที่ประมาณจากแบบจำลอง

X_t^a = ค่าข้อมูลจริง

T = จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

และค่าสัมประสิทธิ์ Theil (Theil's inequality coefficient, U) คือ

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(X_t^s - X_t^a\right)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(X_t^s\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(X_t^a\right)^2}} \tag{3.15}$$

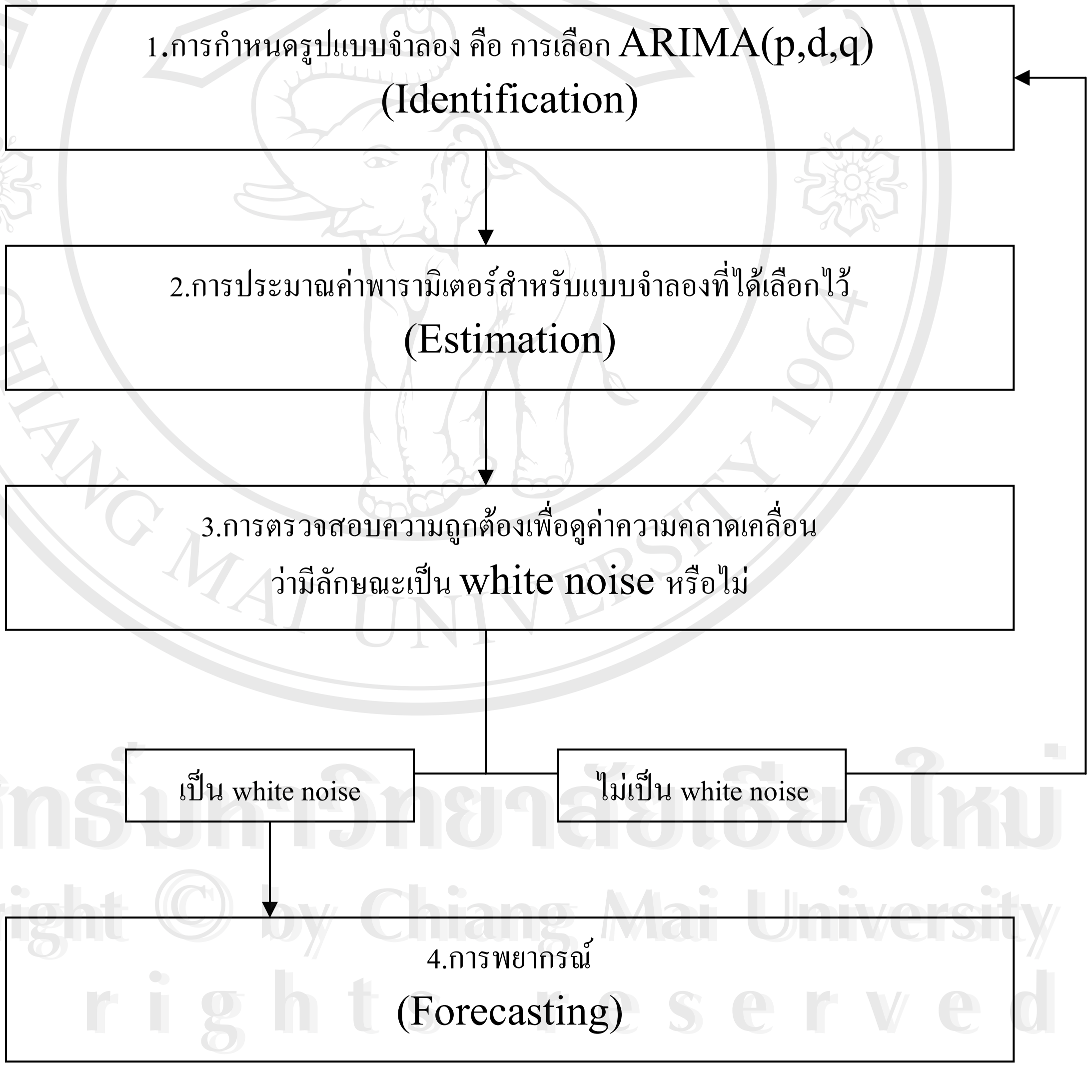
โดยให้ X_t^s = ค่าที่ประมาณจากแบบจำลอง

X_t^a = ค่าข้อมูลจริง

T = จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

ช่วงที่ 2 Ex-post forecast Period คือช่วง T_{n-4} ถึงช่วง T_n เป็นการพยากรณ์เพื่อเปรียบเทียบกับค่าข้อมูลจริงโดยใช้รูปแบบจากช่วง Estimation Period ในการเลือกรูปแบบที่มีความเหมาะสมจะพิจารณาจากค่า ค่า root mean squared error ที่ดีที่สุด และ Theil's inequality coefficient ที่มีค่าต่ำสุด

ช่วงที่ 3 Ex-ante forecast Period คือช่วงที่พยากรณ์อนาคตโดยใช้รูปแบบจากช่วง Ex-post forecast Period ที่มีความเหมาะสมที่สุดในการพยากรณ์



รูปที่ 3.2 ขั้นตอนการพยากรณ์โดยวิธี Box and Jenkins

3.3 วิธีการวิจัย

การพยากรณ์ราคาผลปาล์มในครั้งนี้จะใช้วิธีการวิเคราะห์ โดยกำหนดแบบจำลองให้กับอนุกรมเวลาในรูปแบบ ARIMA Model โดยวิธีของ Box Jenkins สามารถสรุปขั้นตอนการสร้างแบบจำลองได้ดังนี้คือ

1. การนำข้อมูลมาพิจารณาแนวโน้มว่าข้อมูลมีความนิ่งหรือไม่ (stationary or non-stationary) โดยการทดสอบ unit root

2. การกำหนดลำดับขั้น p,q ในว่าแบบจำลองนี้ควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใด และ moving average , q เท่าใด เพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

3. การประมาณค่าพารามิเตอร์ ต้องพิจารณาค่า adjusted R^2 ด้วยคือดูว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้มากน้อยแค่ไหน และพิจารณาค่า AIC ว่า สมการที่ได้ มีค่าความคลาดเคลื่อนของผลการพยากรณ์มากน้อยแค่ไหน

4. การตรวจสอบความถูกต้อง (diagnostics) ว่าสมการแบบจำลองสามารถแทนข้อมูลเบื้องต้นได้ดีที่สุดและมีความเหมาะสมที่จะใช้ในการพยากรณ์ต่อไป แต่หากพบว่าแบบจำลองที่ได้ไม่เหมาะสมจะต้องทำตามขึ้นตอนที่ 1 ใหม่ โดยจะพิจารณาคุณสมบัติความเป็นเชิงสุ่ม (White Noise) ของค่าประมาณความคลาดเคลื่อนโดยพิจารณาจากค่า Q-statistic ว่ามีคุณสมบัติความเป็นเชิงสุ่ม(White Noise) หรือไม่

5. การพยากรณ์ (forecasting) จะแบ่งการพยากรณ์ออกเป็น3 ช่วงได้แก่ ช่วง Historical forecast ช่วง Ex-post forecast และ ช่วง Ex-ante forecast เมื่อได้พิจารณาแบบจำลองโดยเลือกเอาแบบจำลองที่ดีที่สุดที่สามารถเป็นตัวแทนข้อมูลเบื้องต้นได้แล้วก็นำแบบจำลองที่ได้ไปใช้ในการประมาณค่า ณ เวลาต่อไปได้