บทที่ 3
กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการศึกษาการเคลื่อนไหวอนุกรมเวลาราคารายเดือนของสินค้าที่ใช้วิธีการพรรณนาอธิบายและใช้วิธีวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองอารีมา โดย Box-Jenkins พร้อมทั้งมีการอธิบายปัจจัยที่กำหนดการ

3.1.1 แนวคิดการพยากรณ์อนุกรมเวลา

การพยากรณ์อนุกรมเวลาสำหรับแบบจำลอง ARIMA (p, d, q) นั้นจะใช้ทฤษฎีของ Box-Jenkins เป็นเครื่องมือในการศึกษาครั้งนี้ โดยเบื้องต้นต้องพิจารณาว่าอนุกรมเวลามีคุณสมบัติของอนุกรมเวลาที่มีลักษณะนิ่ง (Stationary) เสียก่อน (Pindyck and Rubinfeld, 1997) การพิจารณาว่าจะเป็น stationary หรือไม่ พิจารณาได้จากค่าเฉลี่ยและค่าความแปรปรวนของข้อมูลดังนี้

$$E(y_t) = \mu_y \tag{3.1}$$

กล่าวคือ หากข้อมูลอนุกรมมีลักษณะนิ่ง (Stationary) แล้ว ณ ทุก ๆ ค่าที่เวลา t ใด ๆ จะมีค่าเฉลี่ย $E(y_t)$ คงที่ หรือเท่ากับ μ_y

$$E[(y_t - \mu_y)^2] = \sigma_y^2 \tag{3.2}$$

กล่าวคือ หากข้อมูลอนุกรมมีลักษณะนิ่ง (Stationary) นอกจากค่าเฉลี่ย $E(y_t)$ คงที่แล้ว ค่าความแปรปรวน (Variance) คงที่ สำหรับทุก ๆ ค่าของเวลาที่ t ใด ๆ

$$Cov(y_t, y_{t+k}) = \gamma_k \tag{3.3}$$

กล่าวคือ หากข้อมูลอนุกรมมีลักษณะนิ่ง (Stationary) แล้วเงื่อนไขหนึ่งที่มีความสำคัญ คือค่าความแปรปรวน (Covariance) จะต้องมีค่าคงที่ ณ เวลาที่ t ใด ๆ

ดังนั้นสามารถสรุปได้ว่า หากข้อมูลอนุกรมเวลานั้นมีลักษณะที่นิ่ง (Stationary) จะมีค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วม (จากสมการที่ 3.1, 3.2 และ 3.3) มีค่าที่คงที่ ณ ทุก ๆ เวลาที่เปลี่ยนแปลงไป ซึ่งสามารถทดสอบข้อมูลว่ามีลักษณะที่นิ่งหรือไม่จากการทดสอบ Unit Root

อนุกรมเวลา (Time series) คือค่าสังเกต (Observation) ชุดหนึ่งซึ่งถูกกำหนดขึ้น ณ เวลาต่างๆ ถ้าค่าสังเกตกระทำในเวลาที่ต่อเนื่องกัน จะเรียกว่า อนุกรมเวลาต่อเนื่อง แต่ถ้าค่าสังเกตกระทำ ณ จุดเวลาที่ไม่ต่อเนื่องกัน เรียกว่า อนุกรมเวลาไม่ต่อเนื่อง ดังนั้น การวิเคราะห์อนุกรมเวลาจึงเป็นการวิเคราะห์ค่าสังเกตที่มีการเปลี่ยนแปลงไปตามเวลาที่กระทำ ลักษณะของการเปลี่ยนแปลงอาจจะมีรูปแบบหรือไม่มีก็ได้ แต่ถ้าอนุกรมเวลามีลักษณะการเปลี่ยนแปลงที่มีรูปแบบก็จะทำให้สามารถที่จะพยากรณ์รูปแบบในอนาคตได้ (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542)

การศึกษาอนุกรมเวลาของราคาสินค้ารายเดือน ณ ตลาดต่างๆ โดยการกำหนดรูปแบบให้กับอนุกรมเวลา เพื่อพิจารณาความมีเสถียรภาพและลักษณะรูปแบบพฤติกรรมราคามีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้องกับสินค้าที่ทำการศึกษาหรือไม่ พร้อมทั้งนำรูปแบบของสมการที่ได้ทำการวิเคราะห์อนุกรมเวลาแล้วนำมาทำนายราคาสินค้าที่ศึกษาไปล่วงหน้า เพื่อนำค่าพยากรณ์ดังกล่าวมาใช้ ในการจัดสรรทรัพยากรที่เหมาะสมต่อไป

3.1.2 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit root test)

แบบจำลองที่ใช้ในการศึกษาของข้อมูลอนุกรมเวลา (time-series data) สำหรับทดสอบว่าตัวแปรแต่ละตัวมีลักษณะ นิ่ง (stationary) หรือไม่ ซึ่งเป็นการทดสอบว่ามี unit root หรือไม่นั่นเองทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์ (2542) กล่าวว่า ข้อสมมุติ (assumptions) เบื้องหลังการประมาณค่าทางเศรษฐมิติโดยใช้ข้อมูลอนุกรมเวลานั้น คือข้อสมมุติเกี่ยวกับความนิ่ง (stationary) ของข้อมูล สมมุติว่าเรามีแบบจำลอง

$$Y_t = \alpha + \beta x_t + u_{1t} \tag{3.3}$$

และ

$$X_t = x_{t-1} + u_{2t} \quad ; \quad u_{2t} \sim iid(0, \sigma^2_{u_2}) \tag{3.4}$$

โดยที่ u_{2t} เป็นอนุกรมเวลาของตัวแปรสุ่ม (random variables) ที่แจกแจงแบบปกติที่เหมือนกันและเป็นอิสระต่อกัน โดยค่าเฉลี่ย (mean) เท่ากับศูนย์ และค่าความแปรปรวน (variance) คงที่

ซึ่งตัวแปร x นั้น ก็จะเป็นแนวเดินเชิงสุ่ม (random walk) และเป็น integrated of order one, I (1) เพราะฉะนั้นตัวแปร y ก็จะเป็น I (1) ด้วย ซึ่งโดยทฤษฎีเศรษฐมิติแล้วการถดถอยด้วยตัวแปรที่มีลักษณะไม่นิ่ง (nonstandard distribution) การใช้ตารางมาตรฐานที่เราใช้กันโดยทั่วไปสำหรับการทดสอบค่าสถิติต่าง ๆ ก็อาจนำไปสู่การลงความเห็นหรือข้อสรุปที่ผิดพลาดได้ซึ่งนำไปสู่ความเป็นไปได้ของการมีถดถอยที่ไม่ถูกต้อง (spurious regression) (Jonhston and Dinardo, 1997)

การทดสอบ unit root นั้นสามารถทดสอบได้โดยใช้การทดสอบ DF (Dick-Fuller(DF) test)(Dicky and Fuller, 1981) และการทดสอบ ADF (Augmented Dicky-Fuller test)(Said and Dickey 1984) สมมุติฐานว่าง (null hypothesis) ของการทดสอบ DF (DF test) คือ H_0 : $\rho = 1$ จากสมการ (3.5)

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{3.5}$$

ซึ่งเรียกว่าการทดสอบ unit root โดยถ้า $|\rho| < 1$ X_t จะลักษณะนิ่ง (stationary) ; และถ้า $\rho = 1$ X_t จะมีลักษณะไม่นิ่ง (nonstationary) อย่างไรก็ตามการทดสอบนี้สามารถทำได้อีกทางหนึ่งซึ่งเหมือนกับสมการ (3)

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \tag{3.6}$$

นั่นเอง โดยที่ $\rho = (1 + \theta)$ ถ้า θ ในสมการ (4) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3) จะมีค่าน้อยกว่า 1 ดังนั้นสามารถจะสรุปได้ว่า การปฏิเสธ $H_0 : \theta = 0$ ซึ่งเป็นการยอมรับ H_a : $\theta < 0$ หมายความว่า $\rho < 1$ และ X_t มี integration of order zero (Charemza and Deadman, 1992) นั่นคือ X_t มีลักษณะนิ่ง (stationary) และถ้าเราไม่สามารถปฏิเสธ $H_0 : \theta = 0$ ได้ ก็จะหมายความว่า X_t มีลักษณะไม่นิ่ง(non stationary) โดย X_t มีแนวคิดเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) เราเขียนแบบจำลองได้ดังนี้

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{3.7}$$

และมีแนวโน้มตามเวลาเชิงเส้น เราเขียนแบบจำลองได้ดังนี้

$$\Delta x_t = \alpha + \beta t + \theta x t_{t-1} + \varepsilon_t \tag{3.8}$$

โดยที่ t = แนวโน้มของเวลา

ถ้าหากสมการที่ (3.6) (3.7) และ (3.8) มีกระบวนการเชิงอัตถดถอยเข้ามาเกี่ยวข้อง จะเรียกว่า การทดสอบโดยวิธี Augmented Dickey – Fuller test (ADF test) ได้สมการเป็นดังนี้

$$\Delta x_t = \theta x t_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{3.9}$$

$$\Delta x_t = \alpha + \theta x t_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{3.10}$$

$$\Delta x_t = \alpha + \beta t + \theta x t_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{3.11}$$

ซึ่งการทดสอบแบบดังกล่าวมีการแจกแจงเหมือนกับการทดสอบ DF จึงสามารถใช้ค่าวิกฤตในการพิจารณาแบบเดียวกัน ส่วนการหา Lag Length ตามวิธีของ Ender (Ender, 1995: 227) นั้นให้เริ่มต้นที่ Lag Length ที่มากพอสมควรค่าหนึ่งแล้วค่อยๆ ลดค่า Lag Length ลงเรื่อยๆ เมื่อพบว่าค่าสถิติ t-test หรือ F-test ที่ใช้ในการทดสอบนั้นไม่มีนัยสำคัญ ณ ระดับที่กำหนด จนกระทั่งค่าสถิติมีนัยสำคัญ จึงจะถือว่าค่า Lag นั้นมีความเหมาะสม

3.1.3 แนวคิดการวิเคราะห์ข้อมูลอนุกรมเวลา Box-Jenkins

วิธีการของ Box – Jenkins นั้นเป็นวิธีการพยากรณ์ที่มีความถูกต้องกว่าวิธีอื่น และเหมาะสมกับการพยากรณ์ระยะสั้นในช่วงเวลา 1 เดือนถึง 3 เดือน หากต้องการจะพยากรณ์ในช่วงเวลาที่ยาวนานกว่านี้ควรนำข้อมูลที่ทันสมัยมาปรับค่าพยากรณ์ที่ได้ทำไว้แล้ว เพื่อให้ได้ค่าความคลาดเคลื่อนจากการพยากรณ์ลดลง

วิธีการของ Box-Jenkins เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา รูปแบบที่จะกำหนดให้กับอนุกรมเวลาจะเป็นรูปแบบในกลุ่มของรูปแบบ ARIMA (p,d,q) (Intergrated Autoregreesive-Moving Average order p and q) ซึ่งเป็นการ

รวมส่วนของรูปแบบ AR (p) และรูปแบบ MA (q) เข้าด้วนกัน ส่วนอันดับของ d คือจำนวนครั้งที่หาผลต่าง (Intergrated) รูปแบบ AR (p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่าของ $Y_{t-1}, ..., Y_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA (q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่าของ ความคลาดเคลื่อน $\varepsilon_{t-1}, ..., \varepsilon_{t-q}$ หรือความคลาดเคลื่อนที่อยู่ก่อนหน้า q ค่า ซึ่งรูปแบบ AR MA (p,q) มีการกำหนดรูปแบบดังนี้

แบบจำลองที่ใช้ในการพยากรณ์ คือแบบจำลองอารีมา ARIMA (p, d, q) ซึ่งประกอบไปด้วย 3 ส่วนดังนี้ การถดถอยด้วยตัวเอง (Autoregressive; AR: p) การมีอันดับ (Integrated; I: d) และการเคลื่อนที่ของความคลาดเคลื่อน (Moving Average; MA: q) สำหรับรูปแบบทั่วไปของอารีมา (ARIMA) สามารถเขียนเป็นความสัมพันธ์ได้ดังนี้

AR (p) คือ $Y_t = \theta_0 + \phi_t Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t$

MA (q) คือ $Y_t = \theta_0 + \varepsilon_t - \theta_t \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

AR MA (p,q) คือ $Y_t = \theta_0 + \phi_t Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

ARIMA (p,q) คือ $\Delta^d Y_t = \theta_0 + \phi_t Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

การกำหนดรูปแบบที่เหมาะสมให้กับอนุกรมเวลาจะใช้ค่าค่าสัมประสิทธิ์สหสัมพันธ์แบบออโต (Autocorrelation Function: ACF) และค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนแบบออโต (Partial Autocorrelation)

อนุกรมเวลาที่จะนำมาศึกษาเพื่อประโยชน์ในการพยากรณ์นั้นการเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่างๆ ได้แก่ แนวโน้ม (Trend) ตัวแปรฤดูกาล (Seasonal factor) ตัวแปรวัฏจักร (Cyclical factor) และเหตุการณ์ที่ผิดปกติ (Irregular movement) โดยวิธี Box-Jenkins จะสามารถแบ่งอนุกรมเวลาออกเป็น 2 ประเภทดังนี้

1. อนุกรมเวลาที่เป็น Satationary series เป็นอนุกรมเวลา $\{Y_t\}$ ที่มีค่าเฉลี่ยและความแปรปรวนของ Y_t คงที่ นั่นคือค่าเฉลี่ย $E(Y_t)$ และค่าความแปรปรวน $V(Y_t)$,มีค่าคงที่สำหรับแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้มและ/หรืออิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(Y_t)$ ไม่คงที่ และอนุกรมเวลาที่มีความแปรปรวนของ Y_t สูงจะเป็นลักษณะของอนุกรมเวลาที่ $V(Y_t)$ มีค่าไม่คงที่ ซึ่งจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่เป็น Stationary series นอกจากนั้นอนุกรมเวลาที่เป็น Sationary series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้วยังจะ

ต้องมีค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตที่ lag K ขึ้นอยู่กับค่า K อย่างเดียว อนุกรมเวลาที่กำหนดรูปแบบ AR MA (p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น Stationary series แล้ว

2. อนุกรมเวลาที่ไม่เป็น Stationary series เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น Stationary series การจะหารูปแบบ ARMA (p,q) ให้กับอนุกรมเวลาดังกล่าวได้จะต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ Stationary series เสียก่อน การแปลงอนุกรมเวลาที่ไม่เป็น Stationary series ให้เป็นอนุกรมเวลาที่เป็น Stationary series อาจทำได้ด้วยวิธีการต่างๆดังนี้

2.1 การหาผลต่างปกติ (Regular differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้มนั่นคือถ้าอนุกรมเวลา $\{Y_t\}$ มีแนวโน้มอยู่ในอนุกรมเวลาจะแปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม $\{Z_t\}$ โดย $Z_t = \nabla^d Y_t$ โดย d เป็นลำดับของการหาผลต่างและ ∇ คือผลต่างของตัวแปร เช่นเมื่อ d=1 จะได้ $Z_t = \nabla Y_t = Y_t$ เมื่อ d = 2 จะได้ $Z_t = \nabla^2 Y_t = \nabla(Y_t - Y_{t-1}) = \nabla Y_t - \nabla Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$ เป็นต้น จำนวนครั้งที่หาผลต่างจะขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม้เป็น stationaries series ต้องหาผลต่างต่อไปต่อไป โดยทั่วไป ถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ d = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติก (Quadratic) จะใช้ d = 2

2.2 การหาผลต่างฤดูกาล ของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรเข้ามาเกี่ยวข้อง จะต้องแปรงอนุกรมเวลาเดิม $\{Y_t\}$ ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล $\{Z_t\}$ โดย $Z_t = \nabla^d{}_L Y_t$ โดย D เป็นลำดับของการหาผลต่างฤดูกาล และ L เป็นจำนวนฤดูกาลต่อปี เช่น สำหรับอนุกรมเวลารายเดือน (L = 12) เมื่อ D = 1 จะได้ $Z_t = \nabla_{12} Y_t$ หรือ $Z_t = Y_t - Y_{12}$ หรือ และเมื่อ D = 2 จะได้ $Z_t = \nabla^2{}_{12} Y_t$ หรือ $Z_t = \nabla^2(Y_t - Y_{t-12})$ เป็นต้นผลต่างนี้จะทำกี่ครั้งขึ้นกับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม่เป็น Stationary series หรือไม่ถ้ายังไม่เป็น Stationary series ต้องหาผลต่างต่อไป

2.3 การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น Stationary series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาล โดยที่ค่า d และ D ควบคู่กันไปซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และD จะมีค่าเท่าใดนั้นขึ้นอยู่กับว่าเมื่อหาผลต่างปกติและผลต่างฤดูกาล แล้วอนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม่เป็น Stationary series ต้องหาผลต่างต่อไป เช่น อนุกรมเวลารายเดือน ที่มีทั้งแนวโน้มและฤดูกาล เมื่อ

d = 1 และ D = 1 จะแปลงอนุกรมเวลาเดิม $\{y_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{z_t\}$ ซึ่ง $Z_t = \nabla\nabla_{12} Y_t = \nabla(Y_t - Y_{t-12}) = \nabla Y_t - \nabla Y_{t-12} = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$ เป็นต้น

2.4 การหาลอกการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือ แปลงอนุกรมเวลาเดิม $\{y_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{z_t\}$ โดย $Z_t = \ln(Y_t)$ การแปลงนี้จะทำเมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ นั่นคือ $V(Y_t)$ สำหรับค่าเวลา t ต่างๆ

3.1.4 แบบจำลองการพยากรณ์ โดยวิธี Box-Jenkins

ในการพยากรณ์โดยวิธีของ Box-Jenkins ในรูปแบบ ARIMA นั้นต้องพิจารณาอนุกรมเวลา $\{Y_t\}$ ให้มีคุณสมบัติของอนุกรมเวลาที่เป็น Stationary series เสียก่อน การพิจารณาอนุกรมเวลาว่าเป็น Stationary series หรือไม่ จะพิจารณาได้จาก

(1) ค่าเฉลี่ย $E(Y_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆแล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อย ไม่แตกต่างกันมาก จะสรุปได้ว่า $E(Y_t)$ คงที่

(2) ค่าความแปรปรวน $V(Y_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่ จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆแล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าความแปรปรวนแต่ละส่วนย่อยไม่แตกต่างกันมากนัก จะสรุปได้ว่า $V(Y_t)$ คงที่

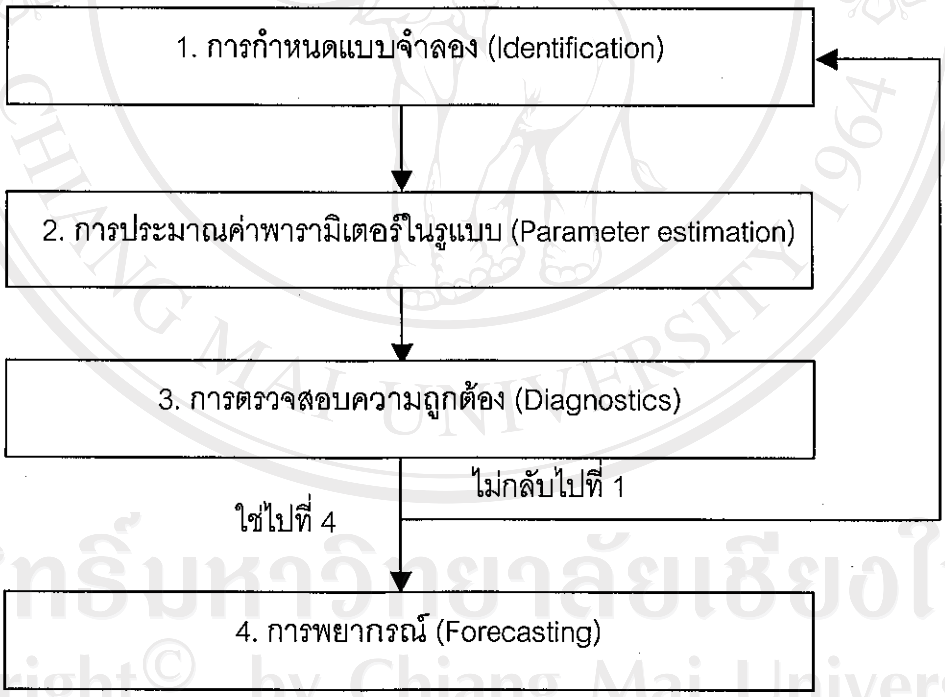
(3) พิจารณาแนวโน้มและ/หรือปัจจัยฤดูกาล ด้วยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและ/หรือปัจจัยฤดูกาล มักจะเห็นชัดเจนได้จากรูป

(4) พิจารณาจาก Correlogram ของค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ Stationary ค่า Correlogram ของ Autocorrelation (r_k) จะมีค่าลดลงค่อนข้างเร็วเมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นถ้าค่า Autocorrelation (r_k) มีค่าลดลงค่อนข้างช้าจะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่า Autocorrlation (r_k) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างสูง ที่ k = L , 2L , 3L จะเป็นข้อสังเกตว่าอนุกรมชุดนี้มีแนวโน้มและอิทธิพลฤดูกาล และถ้าการเคลื่อนไหวของค่า Correlogram ของ Autocorrelation (r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วว่า อนุกรมเวลาที่ศึกษาไม่เป็น Stationary ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่เป็น Stationary จะต้องแปลงอนุกรมเวลาที่เป็น

Stationary เสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างและผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ ให้แปลงอนุกรมเวลาเดิมโดยการหาลอการิทึม (Zt = ln Yt) จนกว่าจะได้อนุกรมเวลาใหม่ ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่ที่เป็น Stationary Series แล้วจะทำตามขั้นตอนของ Box – Jenkins ดังนี้

ขั้นตอนของ Box-Jenkins ประกอบด้วย 4 ขั้นตอน ดังนี้ ขั้นตอนที่หนึ่ง คือการกำหนดรูปแบบจำลอง (Identification) ขั้นตอนที่สอง คือการประมาณค่า (Estimation) ขั้นตอนที่สาม คือวิเคราะห์ความถูกต้อง (Diagnostic Checking) และขั้นตอนที่สี่ คือการพยากรณ์ (Forecasting) ตามลำดับ ดังจะพิจารณาจากรูป 3.1



รูป 3.1 แสดงขั้นตอนของ Box and Jenkins
ที่มา: Gujarati (2003)

1) การกำหนดแบบจำลอง (Identification) ให้กับอนุกรมเวลาที่เป็น Stationary Series เป็นการหารูปแบบ ARMA (p.q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยที่

Autocorrelation : ρ_k คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยที่ ρ_k มีค่าเท่ากับ $-1 \le \rho_k \le 1$ โดยพิจารณาเปรียบเทียบค่า Autocorrelation (r_k) ของอนุกรมเวลาตัวอย่าง กับค่า Autocorrelation (ρ_k) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$\rho_k = \frac{\sum_{t=a}^{n-k}\left(Y_{t-q}\right)\left(Y_{t+k-q}\right)}{\sum_{t=a}^{n}\left(Y_{t-q}\right)^2} \qquad (3.12)$$

โดยที่ $$Y_t = \sum_{t=a}^{n}\left(Y_t\right)$$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตามเนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดากตัวตามแปรอิสระที่เป็นค่าความล่าช้า (Lag) ของตัวแปรตาม (Autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (Moving average) เนื่องจาก Autocorrelation Function (ACF) จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อนแต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง Patial Autocorrelation Function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule-walker (Pindyck and Rubinfeld, 1996) ดังนี้

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p \qquad (3.13)$$

ถ้า k มากกว่า p จะได้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p} \qquad (3.14)$$

การกำหนดลำดับขั้น p,q ในแบบจำลอง (Identiftying the dependence order of model) ขั้นตอนนี้คือการระบุว่าแบบจำลองนี้ควรจะมี Autoregressive,p เท่าใด Differencing, d

ที่ลำดับเท่าใด และ Moving average ,q เท่าใด โดยพิจารณาจาก ACF และ PACF ซึ่งอาจจะใช้ตาราง 3.1 ดังต่อไปนี้พิจารณาร่วม

ตาราง 3.1 ตารางแสดงการพิจารณา 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 4 จะสามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็น ค่าที่ p ของ AR(p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรม เกิดขึ้น 1 แท่ง แปลได้ว่าแบบจำลองควรมีลักษณะเป็น AR(1) สำหรับ MA(q) นั้นก็จะมี ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สามารถสรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(2) และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรจะเป็น ARMA(p,q) และเมื่อรวมกันกับการทดสอบความนิ่ง (Stationary) ในขั้นตอนที่ 1 แล้ว จะสามารถหาค่าของ Difference ได้ ซึ่งผลจากการ Difference จำนวน d ครั้งนั้น ก็จะได้แบบจำลอง ARIMA(p,d,q) แต่อย่างไรก็ตามหลักการดังกล่าวก็เป็นเพียงเครื่องช่วยการพิจารณาในระดับหนึ่งเท่านั้น เนื่องจากวิธีการดังกล่าวมีลักษณะที่เป็นศิลป์ (Art) มากกว่าศาสตร์ (Science) ดังนั้นเพื่อประเมินแบบจำลอง ว่าแบบจำลองใดมีความเหมาะสมที่จะใช้เป็นตัวแทนกลุ่มข้อมูลจริง สามารถพิจารณาได้จากค่าสถิติดังต่อไปนี้เพื่อประกอบในการตัดสินใจ

ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root mean square Error: RMS) โดยจะเป็นการวัดค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองมี

ความแตกต่างกันมากน้อยเพียงใด ซึ่งหากค่า RMSE มีค่าเท่ากับศูนย์ จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงไรก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากเพียงนั้น สามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)} \tag{3.15}$$

กำหนดให้ Y_t^s คือค่าที่ประมาณจากแบบจำลอง

Y_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

นอกจากจะใช้ค่า RMSE ในการประเมินความถูกต้องของแบบจำลองแล้ว ยังสามารถพิจารณาค่าสถิติตัวอื่น เช่น Theil's inequality coefficient โดยในหลักการเบื้องต้น พบว่าสมการที่ใช้นี้ยังคงมีหลักการที่คล้ายกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือค่าสถิตินี้จะมีค่าอยู่ระหว่าง 0 และ 1 ทั้งนี้หากค่า U มีค่าเท่ากับศูนย์ นั่นก็หมายความว่าค่าที่ได้จากการประมาณมีค่าเท่ากันพอดีกับค่าที่เป็นข้อมูลจริง แสดงถึงแบบจำลองที่ประมาณได้เป็นแบบจำลองที่เป็นตัวแทนข้อมูลจริงได้อย่างดีที่สุด ในขณะที่ถ้า U มีค่าเท่ากันกับหนึ่ง แปลว่าแบบจำลองที่ประมาณได้เป็นแบบจำลองที่แย่ที่สุด ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่า U ที่น้อย ๆ ดังจะพิจารณาได้จากสมการที่ (3.16)

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^w - Y_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^s\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^a\right)^2}} \tag{3.16}$$

กำหนดให้ Y_t^s คือค่าที่ประมาณจากแบบจำลอง

Y_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

อย่างไรก็ตาม ยังมีค่าสถิติอีกหลายอย่างที่สามารถนำมาพิจารณาประกอบร่วมกันกับ RMSE และ Theil's inequality coefficient เพื่อใช้ในการเลือกแบบจำลองที่เหมาะสมที่สุด อาทิ เช่น R^2 Adjusted R^2 และ Akaike Information Criterion (AIC) ซึ่งสามารถอธิบายได้ดังต่อไปนี้

R^2 คือการวัดค่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใด หากค่านี้เท่ากับ 1 ก็หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100 % ในทางกลับกัน หากค่านี้มีค่าเท่ากับ 0 แปลความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่อย่างไรก็ตาม พบว่าหากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมาก ๆ ก็จะทำให้ค่า R^2 มากขึ้นด้วย ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการที่ (3.17) ดังนั้นเพื่อปรับปรุงข้อจำกัดข้างต้น จึงเกิดค่าสถิติใหม่ คือค่า Adjusted R^2 ($\overline{R}^2$) ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการที่ (3.18)

$$R^2 = 1 - \frac{\Sigma \hat{u}_i}{\Sigma y_i^2} \tag{3.17}$$

$$\overline{R}^2 = 1 - \frac{\Sigma \hat{u}_i^2 /(n-k)}{\Sigma y_i^2 /(n-1)} \tag{3.18}$$

Akaike Information Criterion (AIC) คือค่าสถิติที่ประยุกต์คล้ายกับ $\overline{R}^2$ แต่ใช้รูปแบบการใส่ค่าลอกกาลิทึมฐานธรรมชาติ (Natural logarithm) โดยหากค่าสถิตินี้มีค่าน้อยเพียงใด นั่นก็แปลว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (Lag Length) ที่เหมาะสมอีกด้วย ซึ่งแสดงในสมการที่ (3.17)

$$\ln AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{\Sigma \hat{u}_i^2}{n}\right) \tag{3.19}$$

กำหนดให้ $\Sigma \hat{u}_i^2$ คือผลรวมของค่าความคลาดเคลื่อน

n คือค่าสังเกตทั้งหมด

จากค่าสถิติข้างต้นทั้งหมดจะนำมาใช้ประกอบในการพิจารณาเลือกแบบจำลอง ARIMA (p, d, q) ที่เหมาะสมที่สุด โดยจะคัดเลือกแบบจำลองในขั้นตอนนี้ไว้ 3 – 4 แบบจำลอง เพื่อทำการเลือกอีกครั้งเพื่อที่จะเปรียบเทียบว่าแบบจำลองใดจะมีความสามารถในการพยากรณ์มากที่สุด

2) การประมาณค่าพารามิเตอร์ในรูปแบบ (Estimation) คือการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (AR) และรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (MA) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (Simple least square) แต่สามารถที่จะใช้วีการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์ได้ หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3) การตรวจสอบแบบจำลอง (Diagnostics) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลอง จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธี ยกตัวอย่าง เช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลอง โดยใช้การทดสอบของ Box และ Pierce ซึ่งจะแสดงได้โดยใช้ Q statistic ดังในสมการที่ (3.18)

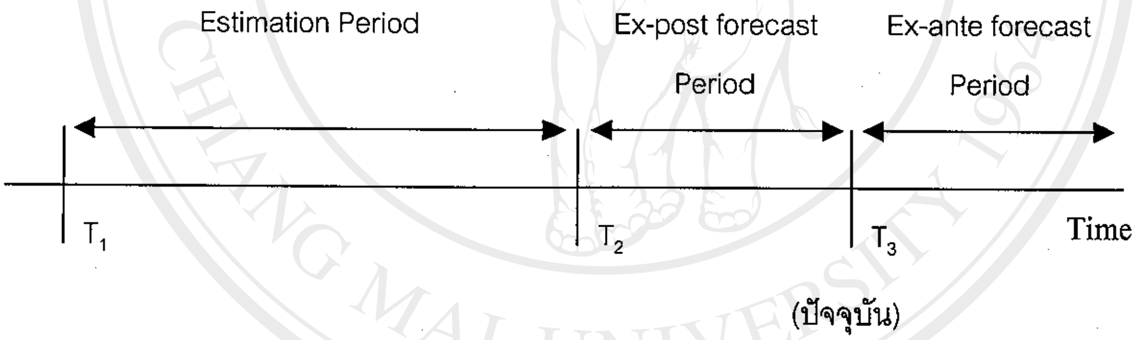
$$Q = n\sum_{k=1}^{m}\rho_k^2 \qquad (3.20)$$

กำหนดให้ n คือจำนวนของข้อมูล

m คือค่า Lag Length

จากสมการที่ (14) ค่า Q นั้นจะพบว่ามีการแจกแจงเป็นแบบ Chi-square ที่มีดีกรีเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมติฐานว่า สมมติฐานว่าง คือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น White Noise นั่นก็แปลว่าแบบจำลองมีลักษณะปราศจากอัตสหสัมพันธ์ (Autocorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นปราศจากอัตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

4) การการพยากรณ์ (Forecasting) เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากการวิเคราะห์ความถูกต้องแล้ว ก็สามารถนำแบบจำลองใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลองโดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง Historical forecast อันเป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจาณา (T_2) การพยากรณ์ช่วง Ex-post forecast คือการพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์แล้วเปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า Root Mean Square Error (RMSE) และ Theil's Inequality Coefficient (TIC) และค่า Akaike Information Criterion (AIC) จะพิจารณาค่าสถิติทั้งสามค่าที่มีค่าน้อยที่สุด ซึ่งได้จากการทำการพยากรณ์ เมื่อเลือกรูปแบบจำลองที่ดีที่สุดแล้ว จึงนำแบบจำลองนั้นมาทำการพยากรณ์แบบ Ex-ante forecast ซึ่งเป็นการพยากรณ์ข้อมูลไปข้างหน้า ดังรูป 3.2



รูป 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา: Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ราคายางพารา โดยวิธี ARIMA เป็นการนำเอาข้อมูลอนุกรมเวลามาหารูปแบบแบบจำลองที่เหมาะสมที่สุด ซึ่งมีขั้นตอนดังต่อไปนี้

1) การลอกกาลิทึมฐานธรรมชาติ (Natural logarithm : ln) ข้อมูลในอนุกรมเวลาราคายางพารา และ X_t ได้ถูกเปลี่ยนเป็น P_t ดังนั้น ตัวแปรที่จะใช้ในสมการต่างๆจึงถูกเปลี่ยนเป็น $\ln P_t$

2) การทดสอบความนิ่งของข้อมูล (Unit root test) เป็นการพิจารณาว่าข้อมูลอนุกรมมีลักษณะนิ่งหรือไม่โดยการทดสอบ unit root ดังสมการต่อไปนี้

$$\Delta x_t = \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{3.21}$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{3.22}$$

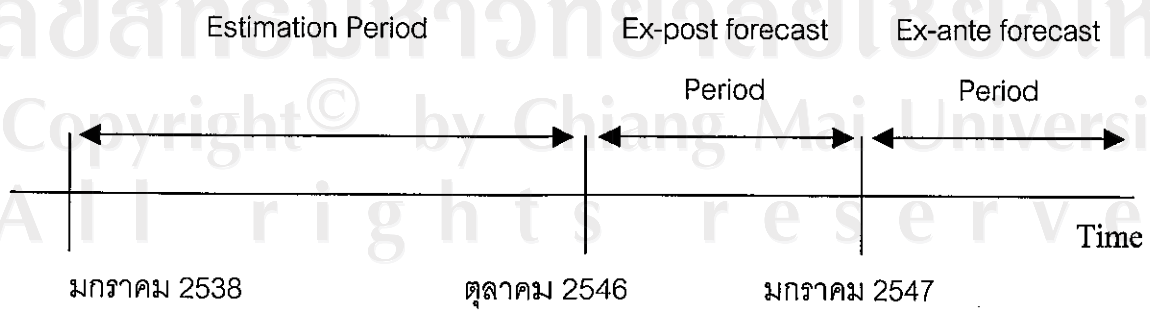
$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{3.23}$$

3) การกำหนดแบบจำลอง ARIMA (pdq) โดยพิจารณา คอเรลโลแกรม Autocorrelation Function (ADF) และ Partial Autocorrelation Function (PADF) เพิ่มที่จะสามารถระบุว่าแบบจำลองควรจะมี Autoregressive (p) เท่าใด และ Moving average (q) เท่าใด โดยเลือกสร้างไว้หลายแบบจำลองเพื่อหาแบบจำลองที่เหมาะสมที่สุด

4) การประมาณค่าพารามิเตอร์ เพื่อนำค่าพารามิเตอร์ที่ได้นั้นนำไปทำการพยากรณ์ราคาต่อไป

5) การตรวจสอบความถูกต้อง เมื่อทำการหาแบบจำลองที่เหมาะสมและประมาณค่าพารามิเตอร์แล้ว จึงทำการทดสอบแบบจำลองโดยพิจารณาจากค่า Q-statistic จากคอเรลโลแกรมของอัตตสหสัมพันธ์ของกลุ่มตัวอย่าง

6) การพยากรณ์ จะทำการพยากรณ์ดัชนีราคายางพารา ซึ่งจะทำการแบ่งการพยากรณ์ออกเป็น 3 ช่วง ได้แก่ ช่วง Historical forecast ช่วง Ex-post forecast และ ช่วง Ex-ante forecast ดังรูป 3.3



รูป 3.3 แสดงช่วงเวลาที่ใช้ในการพยากรณ์จริง