บทที่ 3
กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการวิเคราะห์อนุกรมเวลาแบบเดิมนั้น มีวิธีการพยากรณ์ที่ซับซ้อนเมื่อมีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้อง ส่วนวิธีการวิเคราะห์แบบ Box - Jenkins นั้นสามารถที่ขจัดปัญหาดังกล่าวได้ (วิชิต หล่อจีระชุณห์กุลและคณะ, 2539) ทำให้การพยากรณ์ข้อมูลที่เป็นอนุกรมเวลามีความแม่นยำและสามารถนำไปพยากรณ์ค่าของข้อมูลในอนาคตได้อย่างถูกต้องใกล้เคียงกับความเป็นจริง ดังนั้นในการค้นคว้าอิสระฉบับนี้จึงใช้วิธีการศึกษาการเคลื่อนไหวอนุกรมเวลาราคารายเดือนของทองคำ โดยใช้วิธีการพรรณนาอธิบาย วิธีวิเคราะห์เชิงปริมาณ และกำหนดแบบจำลองอารีมา โดย Box – Jenkins ดังมีทฤษฎีที่เกี่ยวข้อง ดังนี้

3.1.1 แนวคิดการพยากรณ์อนุกรมเวลา

อนุกรมเวลา (Time Series) คือค่าสังเกตหนึ่งชุดซึ่งถูกกำหนดขึ้น ณ เวลาต่างๆ โดยแบ่งออกเป็นอนุกรมเวลาต่อเนื่อง คือค่าสังเกตที่กระทำในเวลาที่ต่อเนื่องกัน และอนุกรมเวลาไม่ต่อเนื่อง คือค่าสังเกตที่กระทำ ณ จุดเวลาที่ไม่ต่อเนื่องกัน

ดังนั้น ในการวิเคราะห์อนุกรมเวลาจึงเป็นการวิเคราะห์ค่าสังเกตที่มีการเปลี่ยนแปลงไปตามเวลาที่กระทำ ลักษณะของการเปลี่ยนแปลงอาจมีรูปแบบหรือไม่มีรูปแบบก็ได้ ถ้าเป็นอนุกรมเวลามีลักษณะการเปลี่ยนแปลงที่มีรูปแบบ ก็สามารถที่จะพยากรณ์รูปแบบในอนาคตได้ (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542)

การศึกษาอนุกรมเวลาของราคาสินค้ารายเดือน ณ ตลาดต่างๆ จะทำการกำหนดรูปแบบให้กับอนุกรมเวลา เพื่อพิจารณาความมีเสถียรภาพและลักษณะของรูปแบบพฤติกรรมราคาว่ามีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้องกับสินค้าที่ทำการศึกษาหรือไม่ พร้อมทั้งนำรูปแบบของสมการที่ได้ทำการวิเคราะห์อนุกรมเวลามาทำนายราคาสินค้าที่ศึกษาไปล่วงหน้า เพื่อนำค่าพยากรณ์ดังกล่าวมาใช้ ในการจัดสรรทรัพยากรที่เหมาะสมต่อไป

3.1.2 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit Root Test)

ในการศึกษาที่อาศัยข้อมูลอนุกรมเวลานั้น เรามีข้อสมมุติว่าอนุกรมเวลานั้นจะต้องมีลักษณะ "นิ่ง (Stationary)" ปัญหาคือว่า "ความนิ่ง (Stationarity)" นั้นหมายถึงอะไรและทำไมเราต้องกังวลกับ "ความไม่นิ่ง (Nonstationarity)" ของข้อมูลอนุกรมเวลา สำหรับคำนิยามของคำว่า "นิ่ง (Stationary)" สามารถนิยาม ได้ดังนี้

กระบวนการเฟ้นสุ่ม (Stochastic Process) จะถูกเรียกว่า นิ่ง (Stationary) ก็ต่อเมื่อค่าเฉลี่ยและค่าความแปรปรวนของกระบวนการเฟ้นสุ่มมีค่าคงที่เมื่อเวลาเปลี่ยนแปลงไป รวมถึงค่าความแปรปรวนระหว่างสองคาบเวลาจะขึ้นอยู่กับระยะทาง (Distance) หรือความล้าหลัง (Lag) ระหว่างคาบเวลาทั้งสอง แต่จะไม่ขึ้นอยู่กับเวลาที่เกิดขึ้นจริงที่ความแปรปรวนร่วมได้ถูกคำนวณ (Gujarati, 2003)

คำนิยามของคำว่า "นิ่ง (Stationary)" ของกระบวนการเฟ้นสุ่มตามที่นิยามนี้เป็นที่รู้จักกันว่าเป็น Weakly Stationary Stochastic Process ซึ่งใช้กันมากในทางปฏิบัติ (Spanos,1986 and Gujarati, 2003) จากคำนิยามดังกล่าวเราสามารถเขียนคำนิยามนี้ในรูปของสัญลักษณ์ทางคณิตศาสตร์ ได้ดังนี้

กระบวนการเฟ้นสุ่ม (y_t) จะถูกรียกว่า "นิ่ง (Stationary)" ถ้า

$$\text{Mean} \quad : E(y_t) = \text{Constant} = \mu \tag{3.1}$$

$$\text{Variance} \quad : V(y_t) = \text{Constant} = \sigma^2 \tag{3.2}$$

$$\text{Covariance} \quad : Cov(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) = \sigma - \mu \tag{3.3}$$

การพิจารณาได้ว่าค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) มีค่าคงที่ เมื่อเวลาเปลี่ยนแปลงไป ในขณะที่ความแปรปรวนร่วม (Covariance) ระหว่างสองคาบเวลาจะขึ้นอยู่กับความกว้างระหว่างเวลา (Gap) เท่านั้น ไม่ได้ขึ้นอยู่กับเวลาที่เกิดขึ้นจริง ถ้าพบว่าไม่เป็นไปตามที่กล่าวมา กระบวนการเฟ้นสุ่มจะถูกเรียกว่ามีลักษณะ "ไม่นิ่ง" (Nonstationary) (Charemza and Deadman, 1992)

ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์ (2542) กล่าวว่า การนำอนุกรมเวลาไปใช้โดยไม่ได้ตรวจสอบ "ความนิ่ง (Stationarity)" ในทางทฤษฎีแล้วการถดถอยด้วยตัวแปรที่มีลักษณะไม่นิ่ง (Stationary) ค่าสถิติ t (t – statistics) จะมีการแจกแจงไม่มาตรฐาน (Nonstandard Distributions) ซึ่งผลที่ตามมาก็คือเมื่อเราใช้ตารางมาตรฐาน (Standard Tables) ต่างๆ อาจนำไปสู่การลงความเห็นหรือบทสรุปที่ผิดพลาดได้ และเป็นไปได้ที่จะนำไปสู่การถดถอยที่ไม่ถูกต้อง (Spurious

Regressions) (Johnston and Dinardo, 1997) เว้นแต่ว่าความสัมพันธ์ดังกล่าวมีลักษณะเป็นความสัมพันธ์แบบการร่วมกันไปด้วยกัน (Cointegrating Relationship) ซึ่งจะทำให้ค่าสถิติ t และ F ที่เราใช้กันตามปกติสามารถที่จะใช้ทดสอบได้ (Gujarati, 2003)

ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์ (2542) กล่าวว่า ข้อสมมุติเบื้องหลังการประมาณค่าทางเศรษฐมิติโดยใช้ข้อมูลอนุกรมเวลานั้น คือข้อสมมุติเกี่ยวกับความนิ่ง (Stationary) ของข้อมูล สมมุติว่าเรามีแบบจำลอง

$$Y_t = \alpha + \beta x_t + u_{1t} \qquad (3.4)$$

และ

$$X_t = x_{t-1} + u_{2t} \quad ; \quad u_{2t} \sim iid(0, \sigma^2_{u_2}) \qquad (3.5)$$

กำหนดให้ u_{2t} เป็นอนุกรมเวลาของตัวแปรสุ่ม (Random Variables) แจกแจงแบบปกติที่เหมือนกันและเป็นอิสระต่อกัน โดยค่าเฉลี่ย (Mean) จะต้องเท่ากับศูนย์ และค่าความแปรปรวน (Variance) คงที่ ซึ่งตัวแปร x นั้น เป็นแนวเดินเชิงสุ่ม (Random Walk) และเป็น Integrated of Order One, I(1) เพราะฉะนั้นตัวแปร y ก็จะเป็น I(1) ด้วย โดยทฤษฎีเศรษฐมิติแล้วการถดถอยด้วยตัวแปรที่มีลักษณะไม่นิ่ง ดังนั้นการใช้ตารางมาตรฐานที่เราใช้กันโดยทั่วไปสำหรับการทดสอบค่าสถิติต่างๆ ก็อาจนำไปสู่ข้อสรุปที่ผิดพลาดได้ทำให้นำไปสู่ความเป็นไปได้ของการถดถอยที่ไม่ถูกต้อง (Spurious Regression) (Jonhston And Dinardo, 1997) ดังได้กล่าวไปแล้วข้างต้น

ดังนั้นในการจะใช้ข้อมูลอนุกรมเวลาจำเป็นต้องทำการทดสอบว่า ตัวแปรแต่ละตัวมีลักษณะนิ่งหรือไม่ ซึ่งเป็นการทดสอบว่ามี Unit Root หรือไม่นั่นเอง

การทดสอบ Unit Root สามารถทดสอบโดยใช้การทดสอบ DF (Dickey - Fuller Test) (Dickey and Fuller, 1981) และการทดสอบ ADF (Augmented Dickey - Fuller Test) (Said and Dickey, 1984) สมมุติฐานว่าง (Null Hypothesis) ของการทดสอบ DF คือ $H_0 : \rho = 1$ จากสมการ (3.6)

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad (3.6)$$

เรียกว่าการทดสอบ Unit Root โดยถ้า $|\rho| < 1$ แล้ว X_t มีลักษณะนิ่ง (Stationary) ; แต่ถ้า $\rho = 1$ แล้ว X_t มีลักษณะไม่นิ่ง (Nonstationary) อย่างไรก็ตามการทดสอบนี้สามารถทำได้อีกทางหนึ่ง เหมือนกับสมการ (3.6)

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

กล่าวคือ $X_t = (1+\theta)X_{t-1} + \varepsilon_t$ คือสมการที่ (3.6) นั่นเอง โดยที่ $\rho = (1+\theta)$ ถ้า θ ในสมการ (3.7) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3.6) มีค่าน้อยกว่า 1 ดังนั้นสามารถสรุปได้ว่า การปฏิเสธ $H_0: \theta = 0$ เป็นการยอมรับ $H_a: \theta < 0$ 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 ก็หมายความว่า X_t มีลักษณะไม่นิ่ง (Nonstationary)

ถ้า X_t มีแนวคิดเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (Random Walk With Drift) เราสามารถเขียนแบบจำลอง ได้ดังนี้

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{3.8}$$

แต่ถ้า X_t มีแนวคิดเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไป (Random Walk With Drift) และมีแนวโน้มตามเวลาเชิงเส้น (Linear Time Trend) รวมอยู่ด้วยสามารถเขียนแบบจำลอง ได้ดังนี้

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t \tag{3.9}$$

กำหนดให้ t = เวลา

ซึ่งทำการทดสอบ $H_0: \theta = 0$ โดยมี $H_a: \theta < 0$ เช่นเดียวกับที่กล่าวมาข้างต้น

สรุปแล้ว Dickey – Fuller (1979) พิจารณาสมการถดถอย 3 รูปแบบ ที่แตกต่างกันในการทดสอบว่ามี Unit Root หรือไม่ ก็คือสมการที่ (3.7) (3.8) และ (3.9) โดยพารามิเตอร์ที่สนใจในทั้ง 3 สมการนี้คือ θ นั่นคือ ถ้า $\theta = 0$ แล้ว X_t มี Unit Root โดยการเปรียบเทียบค่าสถิติ t (t – statistic) ที่คำนวณได้กับค่าที่เหมาะสมที่อยู่ในตาราง Dickey – Fuller (Dickey – Fuller Tables) (Enders, 1995) หรือกับค่าวิกฤต McKinnon (McKinnon Critical Values) (Gujarati, 2003)

อย่างไรก็ตามค่าวิกฤต (Critical Values) จะไม่เปลี่ยนแปลงถ้าสมการที่ (3.7) (3.8) และ (3.9) ถูกแทนที่โดยกระบวนการอัตถดถอย (Autoregressive Processes) ดังสมการ

$$\Delta x_t = \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{3.10}$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{3.11}$$

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{3.12}$$

ตามลำดับ ซึ่งค่าวิกฤตก็จะไม่เปลี่ยนแปลง ในการทดสอบ Unit Root (Enders, 1995 and Gujarati, 2003) จำนวนของ Lagged Difference Terms ที่จะนำเข้ามารวมในสมการนั้นต้องมีมากพอที่จะทำให้พจน์ค่าความคลาดเคลื่อน มีลักษณะเป็น Serially Independent และเมื่อนำเอาการทดสอบ DF มาใช้กับสมการ (3.10) (3.11) และ (3.12) เราเรียกการทดสอบนั้นว่า ADF (Augmented Dickey – Fuller Test) ค่าสถิติทดสอบ ADF (ADF Test Statistic) มีการแจกแจงเชิงเส้นกำกับ (Asymptotic Distribution) เหมือนกับสถิติ DF (DF Statistic) ดังนั้นสามารถหาค่าวิกฤตแบบเดียวกัน (Gujarati, 2003)

ในกรณีการหา Lag Length ที่เหมาะสมนั้น Enders (1995) ได้เสนอว่าให้เริ่มต้นที่ Lag Length ที่มากพอสมควรค่าหนึ่งแล้วค่อยๆ ลดค่าลงเรื่อยๆ โดยใช้ค่าสถิติทดสอบ t (t – test) หรือค่าสถิติทดสอบ F (F – test) เมื่อทดสอบแล้วพบว่าค่าสถิติ t – test หรือ F – test ที่ใช้ในการทดสอบนั้นไม่มีนัยสำคัญ ณ ค่าวิกฤตที่กำหนดให้ ต้องทำการทดสอบใหม่โดยทำการลดค่า Lag Length จนกระทั่งค่าสถิติมีนัยสำคัญจึงจะถือว่าค่า Lag นั้นมีความเหมาะสม สมมุติว่าเราใช้ Lag Length ที่ n* ถ้าค่าสถิติ t – test หรือ F – test ของ Lag n* ไม่มีนัยสำคัญ ณ ค่าวิกฤตที่กำหนดให้ เราต้องทำการประมาณค่าการถดถอยใหม่ โดยให้ Lag Length เท่ากับ n*-1 ทำอย่างนี้ไปเรื่อยๆ จนกระทั่ง Lag นั้นมีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญ

3.1.3 แนวคิดการวิเคราะห์ข้อมูลอนุกรมเวลา Box - Jenkins

วิธีการ Box – Jenkins นั้นเป็นวิธีการพยากรณ์ที่มีความถูกต้องและเหมาะสมกับการพยากรณ์ระยะสั้นในช่วงเวลาตั้งแต่ 1 เดือนถึง 3 เดือน หากต้องการพยากรณ์ในช่วงเวลาที่ยาวนานกว่านี้ ควรนำข้อมูลที่ทันสมัยมาทำปรับค่าพยากรณ์ที่ได้ทำไว้แล้ว เพื่อลดค่าความคลาดเคลื่อนจากการพยากรณ์ให้น้อยลง

วิธีการ Box - Jenkins นี้เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา รูปแบบที่จะกำหนดให้กับอนุกรมเวลาเป็นรูปแบบในกลุ่มของ ARIMA (p,d,q)

(Integrated Autoregressive - Moving Average Order p and q) เป็นการรวมส่วนของรูปแบบ AR (p) และรูปแบบ MA (q) เข้าด้วยกัน ส่วนอันดับของ d คือจำนวนครั้งที่หาผลต่าง (Integrated)

รูปแบบ AR (p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต Y_t ขึ้นอยู่กับค่าของ $Y_{t-1}, \ldots, Y_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า

รูปแบบ MA(q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต Y_t ขึ้นอยู่กับค่าของความคลาดเคลื่อน $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ หรือความคลาดเคลื่อนที่อยู่ก่อนหน้า q ค่า

ซึ่งรูปแบบ AR(p), MA(q) และ ARMA(p,q) มีการกำหนดรูปแบบ ดังนี้

AR(p) $\quad ; Y_t = \theta_0 + \phi_t Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \quad (3.13)$

MA(q) $\quad ; Y_t = \theta_0 + \varepsilon_t - \theta_t \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \quad (3.14)$

ARMA(p,q) $\quad ; Y_t = \theta_0 + \phi_t Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \quad (3.15)$

โดยแบบจำลองอารีมา ARIMA (p,d,q) ที่ใช้ในการพยากรณ์ ประกอบด้วย 3 ส่วน ดังนี้

1. การถดถอยด้วยตนเอง (Autoregressive; AR: p)
2. การมีอันดับ (Integrated; I: d)
3. การเคลื่อนที่ของความคลาดเคลื่อน (Moving Average; MA: q)

สำหรับรูปแบบทั่วไปของอารีมา (ARIMA) สามารถเขียนเป็นความสัมพันธ์ได้ ดังนี้

ARIMA(p,d,q)

$$; \Delta^d Y_t = \theta_0 + \phi_t Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \qquad (3.16)$$

โดยการกำหนดรูปแบบที่เหมาะสมให้กับอนุกรมเวลาจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบออโต (Autocorrelation Function: ACF) และค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนแบบออโต (Partial Autocorrelation Function: PACF)

อนุกรมเวลาที่นำมาศึกษาเพื่อประโยชน์ในการพยากรณ์นั้น การเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่างๆ ได้แก่ แนวโน้ม (Trend) ตัวแปรฤดูกาล (Seasonal Factor) ตัวแปรวัฏจักร (Cyclical Factor) และเหตุการณ์ที่ผิดปกติ (Irregular Movement) โดยวิธี Box - Jenkins แบ่งอนุกรมเวลาออกเป็น 2 ประเภทดังนี้

1. อนุกรมเวลาที่เป็น Stationary Series

เป็นอนุกรมเวลา $\{Y_t\}$ ที่มีค่าเฉลี่ยและความแปรปรวนของ Y_t คงที่ นั่นคือค่าเฉลี่ย $E(Y_t)$ และค่าความแปรปรวน $V(Y_t)$,มีค่าคงที่สำหรับแต่ละอนุกรมเวลา อนุกรมเวลาที่มีแนวโน้มและ/หรืออิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(Y_t)$ ไม่คงที่และอนุกรมเวลาที่มีความแปรปรวนของ Y_t สูงจะเป็นลักษณะของอนุกรมเวลาที่ $V(Y_t)$ มีค่าไม่คงที่โดยจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่เป็น Stationary Series นอกจากนี้อนุกรมเวลาที่เป็น Stationary Series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้ว ยังต้องมีค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตที่ Log K ขึ้นอยู่กับค่า K อย่างเดียว อนุกรมเวลาที่กำหนดรูปแบบ ARMA(p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น Stationary Series แล้ว

2. อนุกรมเวลาที่ไม่เป็น Stationary Series

เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น Stationary Series การหารูปแบบ ARMA(p,q) ให้กับอนุกรมเวลาดังกล่าวได้ ต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ Stationary Series เสียก่อน การแปลงอนุกรมเวลาที่ไม่เป็น Stationary Series ให้เป็นอนุกรมเวลาที่เป็น Stationary Series อาจทำได้ด้วยวิธีการต่างๆ ดังนี้

2.1 การหาผลต่างปกติ (Regular Differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้ม นั่นคือถ้าอนุกรมเวลา $\{Y_t\}$ มีแนวโน้มอยู่ในอนุกรมเวลาก็แปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม $\{Z_t\}$ โดย $Z_t = \nabla^d Y_t$ โดย d เป็นลำดับของการหาผลต่างและ ∇ คือผลต่างของตัวแปร เช่น เมื่อ 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 จำนวนครั้งที่หาผลต่างขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป โดยทั่วไปถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ d = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบ ควอดราติก (Quadratic) จะใช้ d = 2

2.2 การหาผลต่างฤดูกาลของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรเข้ามาเกี่ยวข้องต้องแปลงอนุกรมเวลาเดิม $\{Y_t\}$ ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล $\{Z_t\}$ โดย $Z_t = \nabla^d{}_L Y_t$ โดย D เป็นลำดับของการหาผลต่างฤดูกาล และ L เป็นจำนวนฤดูกาลต่อปี เช่น อนุกรมเวลารายเดือน (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 $Z_t = \nabla^2(Y_t - Y_{t-12})$ เป็นต้น ผลต่างนี้ต้องทำกี่ครั้งขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป

2.3 การหาผลต่างปกติและผลต่างฤดูกาล นั่นคืออนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น Stationary Series นั้นทำได้โดยการหาผลต่างปกติและ

ผลต่างฤดูกาล โดยค่า d และ D ควบคู่กันไปซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และ ค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และ D จะมีค่าเท่าใดขึ้นอยู่กับว่าเมื่อหาผลต่างปกติและผลต่างฤดูกาล 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 การแปลงอนุกรมเวลาเดิม $\{Y_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{Z_t\}$ โดย $Z_t = \ln(Y_t)$ การแปลงนี้จะทำเมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ นั่นคือ $V(Y_t)$ สำหรับค่าเวลา t ต่างๆ

3.1.4 แบบจำลองการพยากรณ์ โดยวิธี Box - Jenkins

ในการพยากรณ์วิธีของ Box - Jenkins ในรูปแบบ ARIMA ต้องพิจารณาอนุกรมเวลา $\{Y_t\}$ ให้มีคุณสมบัติของอนุกรมเวลาที่เป็น Stationary Series เสียก่อน การพิจารณาอนุกรมเวลาว่าเป็น Stationary Series หรือไม่นั้น สามารถพิจารณาได้จาก .

1. ค่าเฉลี่ย $E(Y_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆแล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่าง กันมากนัก สรุปได้ว่า $E(Y_t)$ คงที่

2. ค่าความแปรปรวน $V(Y_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าความแปรปรวนแต่ละส่วนย่อยไม่แตกต่างกันมากนัก สรุปได้ว่า $V(Y_t)$ คงที่

3. การพิจารณาแนวโน้มและ/หรือปัจจัยฤดูกาล โดยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและ/หรือปัจจัยฤดูกาล จะเห็นได้ชัดเจนจากรูปกราฟ

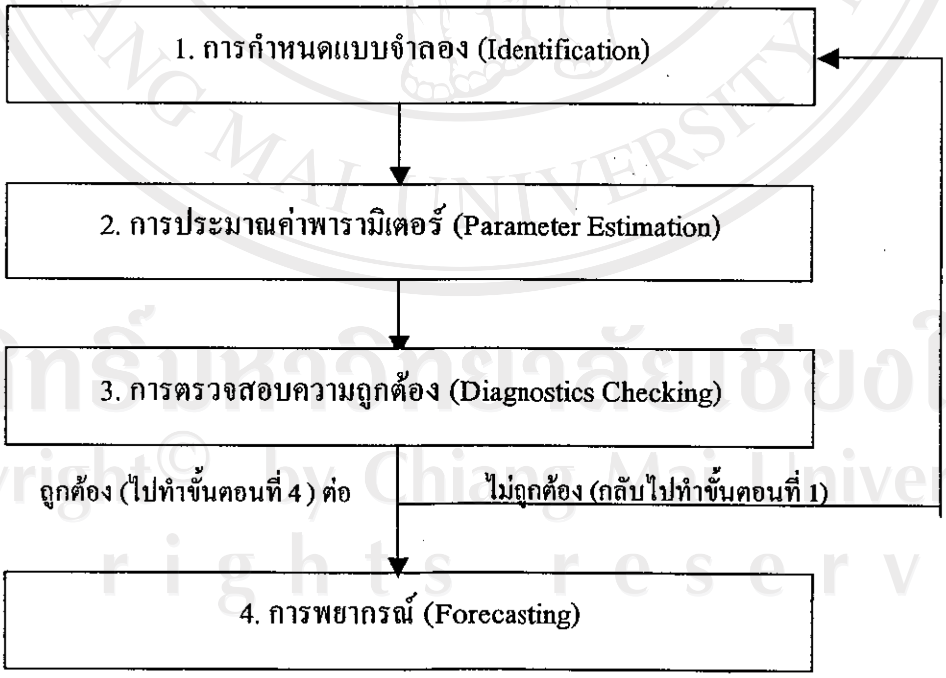
4. พิจารณาจาก Correlogram ของค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตของตัว อย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ Stationary ค่า Correlogram ของ Autocorrelation (r_k) จะมีค่าลดลงค่อนข้างเร็ว เมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นถ้าค่า Autocorrelation (r_k) มีค่าลดลงค่อนข้างช้า แสดงได้ว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่า Autocorrlation (r_k) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างสูง ที่ k = L ,2L ,3L แสดงได้ว่าอนุกรมชุดนี้มีแนวโน้มและอิทธิพลฤดูกาล ถ้าการเคลื่อนไหวของค่า Correlogram ของ Autocorrelation (r_k) มีลักษณะคล้ายลูกคลื่น โดยลูกคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาแล้วว่า อนุกรมเวลาที่ศึกษาไม่เป็น Stationary ก่อนที่จะกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่เป็น Stationary นั้น ต้องแปลงอนุกรมเวลาให้เป็น Stationary เสียก่อน หากเป็นอนุกรมเวลาที่มีแนวโน้มให้หาผลต่างจนได้อนุกรมเวลาที่เป็น Stationary ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างและผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลาเดิม โดยการหาลอการิทึม ($Z_t = \ln Y_t$) จนกว่าจะได้อนุกรมเวลาใหม่ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่ที่เป็น Stationary Series แล้วจะทำตามขั้นตอนของ Box – Jenkins ดังนี้

ขั้นตอนของ Box – Jenkins ประกอบด้วย 4 ขั้นตอน ดังนี้

1. การกำหนดแบบจำลอง (Identification)
2. การประมาณค่าพารามิเตอร์ (Parameter Estimation)
3. การตรวจสอบความถูกต้อง (Diagnostics Checking)
4. การพยากรณ์ (Forecasting)

ดังจะพิจารณาถึงขั้นตอนต่างๆ ได้จากรูป 3.1



รูป 3.1 การแสดงขั้นตอนของ Box - Jenkins

ที่มา: Gujarati (2003)

1) การกำหนดแบบจำลอง (Identification)

คือการกำหนดรูปแบบให้กับอนุกรมเวลาที่เป็น Stationary Series กล่าวคือเป็นการหารูปแบบ ARMA (p,q) ที่เหมาะสมให้กับอนุกรมเวลาโดยที่

Autocorrelation: ρ_k คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลัง k หน่วยเวลา โดยที่ ρ_k มีค่าเท่ากับ $-1 \leq \rho_k \leq 1$ โดยพิจารณาเปรียบเทียบค่า Autocorrelation (r_k) ของอนุกรมเวลาตัวอย่างกับค่า Autocorrelation (ρ_k) ของอนุกรมเวลาประชากร ที่มีช่วงเวลาย้อนหลัง k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$\rho_k = \frac{\sum_{t=a}^{n-k}\left(Y_{t-q}\right)\left(Y_{t+k-q}\right)}{\sum_{t=a}^{n}\left(Y_{t-q}\right)^2} \qquad (3.17)$$

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

q คือจำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตามอนุกรมเวลามักจะมีปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวแปรตามที่เป็นค่าระยะห่างของตัวแปรตาม (Autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (Moving Average) เนื่องจาก Autocorrelation Function (ACF) จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อน แต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง Patial Autocorrelation Function (PACF) จะใช้วัดความสัมพันธ์ของข้อมูลที่เกิดจากการถดถอยด้วยตัวเอง ดังจะสามารถพิจารณาได้จากสมการ Yule - Walker (Pindyck and Rubinfeld, 1996) ดังนี้

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p \qquad (3.18)$$

ถ้า $k > p$ จะได้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \qquad (3.19)$$

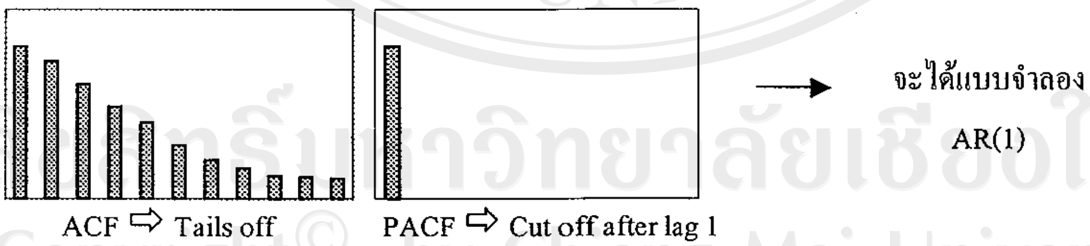
การกำหนดลำดับขั้น p, q ในแบบจำลอง (Identiftying the Dependence Order of Model) ขั้นตอนนี้เป็นการระบุว่าแบบจำลองนี้ควรจะมี Autoregressive (p) เท่าใด Differencing (d) ที่ลำดับเท่าใด และ Moving Average (q) เท่าใด โดยพิจารณาจาก ACF และ PACF ซึ่งอาจจะใช้ตาราง 3.1 ดังต่อไปนี้พิจารณาร่วม

ตาราง 3.1 ตารางแสดงการพิจารณา ACF และ PACF

ชนิดแบบจำลอง	รูปแบบ ACF	รูปแบบ PACF
AR(p)	ลู่โค้งเข้าหาแกน (Tails Off)	เกิดค่าที่ชัดเจนเพียง p ค่าแล้วหายไป (Cut Off After Lag p)
MA(q)	เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป (Cut Off After Lag q)	ลู่โค้งเข้าหาแกน (Tails Off)
ARMA(p,q)	ลู่โค้งเข้าหาแกน (Tails Off)	ลู่โค้งเข้าหาแกน (Tails Off)

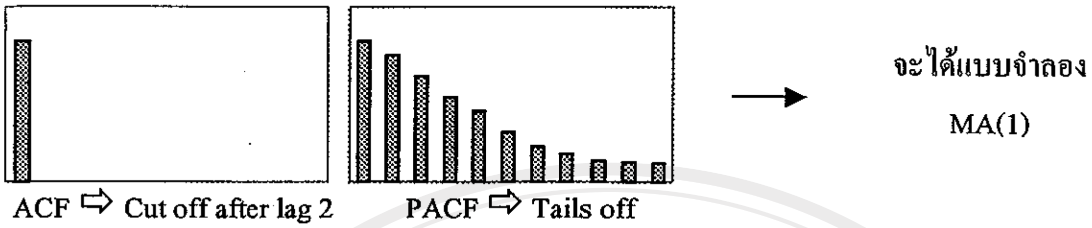
ที่มา: Gujarati (2003)

จากตาราง 3.1 สามารถกำหนดรูปแบบของแบบจำลองได้ดังนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วหายไป จำนวนของแท่งของค่าที่เกิดขึ้น ให้นับเป็นค่าที่ p ของ AR(p) เช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF มีแท่งคอเรลโลแกรม เกิดขึ้น 1 แท่ง นั่นคือแบบจำลองควรมีลักษณะเป็น AR(1) ดังรูป 3.2



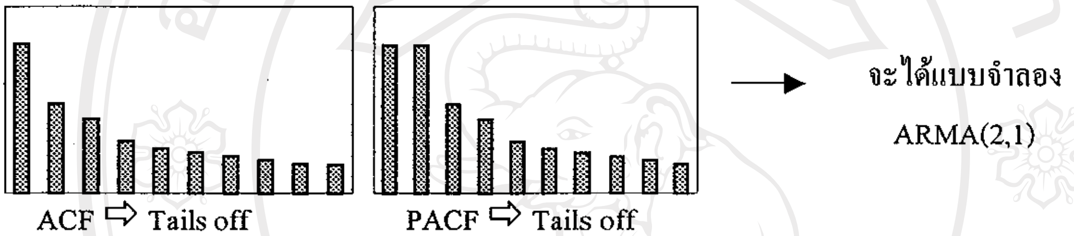
รูป 3.2 ตัวอย่างการเกิดแบบจำลอง AR (p)

สำหรับ MA(q) นั้นก็จะมี ACF เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบนั้น เช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 1 แท่ง แล้วหายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(1) ดังรูป 3.3



รูป 3.3 ตัวอย่างการเกิดแบบจำลอง MA (q)

และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรเป็น ARMA (p,q) ตัวอย่างการเกิดแบบจำลอง ARMA (p,q) ดังรูป 3.4



รูป 3.4 ตัวอย่างการเกิดแบบจำลอง ARMA (p,q)

กรณีที่ ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ หากรวมกับการทดสอบความนิ่ง (Stationary) ในขั้นตอนที่ 1 ก็สามารถหาค่าของ Difference ได้ ซึ่งผลจากการ Difference จำนวน d ครั้ง จะได้แบบจำลอง ARIMA(p,d,q)

อย่างไรก็ตาม หลักการดังกล่าวเป็นเพียงเครื่องมือช่วยในการพิจารณาเพียงระดับหนึ่งเท่านั้น ดังนั้นในการประเมินแบบจำลองว่าแบบจำลองใดมีความเหมาะสมที่จะเป็นตัวแทนของกลุ่มข้อมูลจริงนั้น เราสามารถพิจารณาได้จากค่าสถิติเพื่อประกอบการตัดสินใจ ดังต่อไปนี้

ก. ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root Mean Squared Error: RMSE)

คือการวัดค่าความคลาดเคลื่อนระหว่างค่าที่แท้จริงเปรียบเทียบกับค่าที่ประมาณขึ้นจากแบบจำลองว่ามีความใกล้เคียงมากน้อยเพียงใด กล่าวคือหากค่า RMSE มีค่าต่ำแสดงว่า สมการที่ได้จากแบบจำลองมีความคลาดเคลื่อนน้อย ซึ่งเป็นการประมาณสมการที่ดีและสามารถจะอธิบายได้ว่าค่าที่พยากรณ์ได้มีความโน้มเอียงต่ำ ดังนั้นหากว่าค่า RMSE มีค่าต่ำมากเท่าใดก็แสดงว่าแบบ

จำลองนั้นสามารถเป็นตัวแทนค่าที่แท้จริงได้ดีเท่านั้น สามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง ได้ดังนี้

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)} \tag{3.20}$$

กำหนดให้ Y_t^s คือค่าที่ประมาณจากแบบจำลอง

Y_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

ข. Theil's Inequality Coefficient (U)

เป็นค่าสถิติที่สามารถบอกถึงมาตราวัดสัมพัทธ์ของค่าความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) ว่าเกิดมาจากส่วนใด การพิจารณา Theil's Inequality Coefficient (U) จะทำให้รู้ถึงสาเหตุที่มีผลต่อค่า MSE ของตัวแบบการพยากรณ์ (วิชิต หล่อจีระชุณห์กุล และคณะ, 2539) โดยในหลักการเบื้องต้นพบว่าสมการ Theil's Inequality Coefficient (U) ที่ใช้นี้ยังคงมีหลักการที่คล้ายกันกับ RMSE แต่มีสิ่งที่แตกต่างออกไปก็คือ ค่า Theil's Inequality Coefficient (U) จะมีค่าอยู่ระหว่าง 0 ถึง 1 ทั้งนี้ หากค่า U ที่ได้มีค่าเท่ากับศูนย์ นั่นหมายความว่าค่าที่ประมาณมีค่าเท่ากันกับข้อมูลจริงพอดี แสดงว่าแบบจำลองที่ประมาณขึ้นมาเป็นแบบจำลองที่จะเป็นตัวแทนของข้อมูลที่แท้จริงได้อย่างสมบูรณ์ หากค่า U ที่ได้มีค่าเท่ากับหนึ่ง ก็หมายความว่าแบบจำลองที่ประมาณขึ้นมาเป็นแบบจำลองที่ไม่ดี ไม่สามารถจะเป็นตัวแทนของข้อมูลจริงได้ ค่า U สามารถได้จากสมการดังต่อไปนี้

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^s\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^a\right)^2}} \tag{3.21}$$

กำหนดให้ Y_t^s คือค่าที่ประมาณจากแบบจำลอง

Y_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

ค. ค่า R^2 และ Adjusted R^2 ($\overline{R}^2$)

คือการวัดค่าของตัวแปรอิสระว่าสามารถอธิบายตัวแปรตามได้มากน้อยเพียงใด หากค่า R^2 เท่ากับหนึ่ง นั่นหมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้อย่างสมบูรณ์ ในทางตรงข้าม หากค่านี้มีค่าเท่ากับศูนย์ ก็ความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้

แต่อย่างไรก็ตาม หากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมากก็จะทำให้ค่า R^2 มากขึ้นด้วย ซึ่งถือว่าเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการ (3.22)

$$R^2 = 1 - \frac{\Sigma \hat{u}_i}{\Sigma y_i^2} \tag{3.22}$$

ดังนั้นเพื่อปรับปรุงข้อจำกัดข้างต้น จึงเกิดค่าสถิติใหม่ คือค่า Adjusted R^2 ($\overline{R}^2$) ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการ (3.23)

$$\overline{R}^2 = 1 - \frac{\Sigma \hat{u}_i^2 /(n-k)}{\Sigma y_i^2 /(n-1)} \tag{3.23}$$

นอกจากนี้ Greene (2002) ยังได้กล่าวด้วยว่า การใช้ Adjusted R^2 ($\overline{R}^2$) นั้นมีคำถามว่า การใช้ Adjusted R^2 ($\overline{R}^2$) เป็นการลงโทษที่ทำให้มีการสูญเสียระดับขั้นความเสรี (Degree of Freedom) นั้น เพียงพอหรือยังที่จะประกันว่าเกณฑ์นี้ได้นำไปสู่การชี้แบบจำลองที่ถูกต้อง เมื่อขนาดตัวอย่างเพิ่มขึ้นเพราะฉะนั้นจึงมีการเสนอทางเลือกอื่นที่ใช้เกณฑ์เกี่ยวกับการปรับได้อย่างดี (Goodness of Fit) อีกทั้ง (Gujarati, 2003) กล่าวอ้างว่า นอกจาก Adjusted R^2 ($\overline{R}^2$) ที่จะใช้วัดการปรับได้อย่างดีแล้ว ยังมีวิธีวัดแบบ Mallows' Cp, Amemiya's Pc, Hocking's Sp, Akaike's AIC, Schwarz Criterion, Hannan – Quinn Criterion และ Shibata Criterion

แต่สำหรับการค้นคว้าอิสระฉบับนี้จะกล่าวถึงเฉพาะการวัดหรือเกณฑ์ที่นิยมใช้กันมากเท่านั้นซึ่ง ได้แก่ Akaike's AIC และ Schwarz Criterion เนื่องจากเป็นวิธีวัดที่เกี่ยวเนื่องกับโปรแกรมที่ใช้ในการพยากรณ์

ง. Akaike Information Criterion (AIC)

Maddala (1992) ได้กล่าวว่า AIC นี้นิยมใช้กันทั่วไปในแบบจำลองที่ไม่เป็นเชิงเส้น (Nonlinear Models) เป็นค่าสถิติที่ประยุกต์คล้ายกับ Adjusted R^2 ($\overline{R}^2$) แต่ใช้รูปแบบการใส่ค่าลอกาลิทึมธรรมชาติ (Natural Logarithm) ค่าสถิตินี้ยังสามารถที่จะนำไปใช้หาค่าย้อนหลัง (Lag Length) ที่เหมาะสมได้อีกด้วย สามารถเขียนเป็นสมการ AIC ดังสมการ (3.24)

$$AIC = \ln\left(\frac{\underline{e}'\underline{e}}{n}\right) + \left(\frac{2k}{n}\right) \tag{3.24}$$

กำหนดให้ $\underline{e'e} = \sum_{i=1}^{n} e_i^2$

$\underline{e'e}$ คือผลบวกของกำลังสองของส่วนตกค้างหรือส่วนที่เหลือ (Residual) จากแบบจำลองใดๆ

จ. Schwarz Criterion (SC)

คือ วิธีการวัดปรับได้อย่างดี (Goodness of Fit) เป็นวิธีที่ประยุกต์คล้ายกับ Akaike Information Criterion (AIC) เนื่องจากเป็นวิธีหนึ่งใน Information Criteria เหมือนกัน

โดย Information Criteria จะประกอบด้วย Schwarz Criterion (SC), the Hannan – Quinn Criterion (HQ), the Final Prediction Error (FPF) และ Akaike Information Criterion (AIC) แต่ SC จะให้ค่าในการประเมินแบบจำลองที่ดีกว่า AIC สามารถเขียนสมการ SC ได้ดังสมการ (3.25)

$$SC = \log\left(\frac{\underline{e'e}}{n}\right) + \frac{k \log n}{n} \tag{3.25}$$

หากค่าสถิติ Akaike Information Criterion (AIC) และ Schwarz Criterion (SC) มีค่าน้อยเพียงใด ก็แสดงว่าแบบจำลองที่ประมาณขึ้นสามารถเป็นตัวแทนของข้อมูลจริงได้ดีเท่านั้น

จากค่าสถิติที่กล่าวมาข้างต้นจะถูกนำมาใช้ประกอบการพิจารณาเลือกแบบจำลอง ARIMA(p,d,q) ที่มีความเหมาะสมที่สุด โดยจะทำการคัดเลือกแบบจำลองในขั้นตอนนี้ 4 - 5 แบบจำลอง เพื่อทำการเปรียบเทียบว่าแบบจำลองใดมีความสามารถในการพยากรณ์มากที่สุด

2) การประมาณค่าพารามิเตอร์ (Parameter Estimation)

คือการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (Autoregressive; AR: p) และรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (Moving Average; MA: q) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (Simple Least Square) และวิธีการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะนำไปใช้ในการพยากรณ์ หากรูปแบบของความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุดแล้ว

3) การตรวจสอบแบบจำลอง (Diagnostic Checking)

เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์แบบจำลองแล้วต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดขึ้นนั้น มีความเหมาะสมหรือไม่ การตรวจสอบทำได้หลายวิธี ได้แก่ การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) การทดสอบค่าพารามิเตอร์ในแบบจำลองด้วยการทดสอบแบบ t และการทดสอบความเหมาะสมของแบบจำลองโดยการทดสอบของ Box และ Pierce หรือการทดสอบของ Box และ Ljung เป็นต้น

รวมทั้ง Gujarati (2003) ก็ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลองโดยใช้การทดสอบของ Box และ Pierce ซึ่งจะแสดงได้โดยใช้ Q - Statistic ดังสมการ (3.26)

$$Q = n\sum_{k=1}^{m} \rho_k^2 \tag{3.26}$$

กำหนดให้ n คือจำนวนของข้อมูล

m คือค่า Lag Length

จากสมการ (3.26) ค่า Q นั้นมีการแจกแจงแบบ Chi - Square ที่มีดีกรีเท่ากับ m อยู่ภายใต้ข้อสมมติฐานที่ว่า สมมติฐานว่าง (Null Hypothesis) คือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น White Noise กล่าวคือแบบจำลองมีลักษณะไม่มีอัตสหสัมพันธ์ (Autocorrelation) ดังนั้นเราสามารถใช้แบบจำลองนั้นในการพยากรณ์ต่อไปได้ แต่ถ้าแบบจำลองนั้นไม่เหมาะสมต้องย้อนกลับไปทำขั้นตอนที่ 1 คือทำการกำหนดรูปแบบจำลองใหม่

4) การพยากรณ์ (Forecasting)

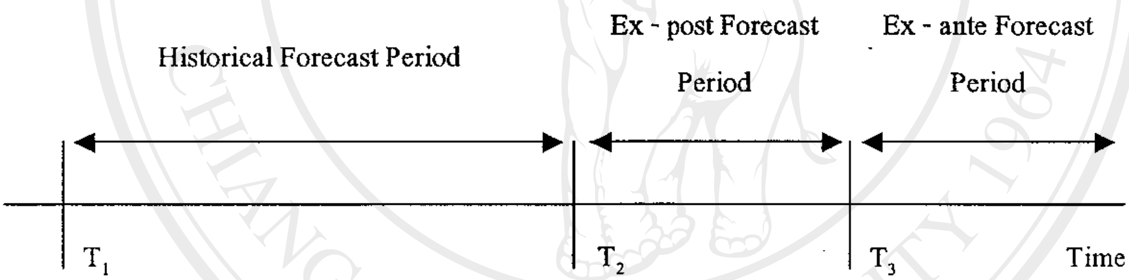
นำแบบจำลองที่ผ่านการตรวจสอบความถูกต้องแล้วมาใช้ในการพยากรณ์ แต่การพยากรณ์ข้อมูลไปข้างหน้านั้นต้องใช้แบบจำลองที่สามารถให้ค่าพยากรณ์ได้แม่นยำที่สุด ดังนั้นจึงต้องมีการทดสอบแบบจำลอง โดยทำการแบ่งการพยากรณ์ออกเป็น 3 ช่วง ดังนี้

1. Historical Forecast คือการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่จะเริ่มพิจารณา (T_2)

2. Ex - post Forecast คือการพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์เปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า Root Mean Squared Error, ค่า Theil's Inequality Coefficient และค่า Akaike Information Criterion เลือกแบบจำลองที่ได้ค่าทั้งสามค่าข้างต้นที่น้อยที่สุด เป็นแบบจำลองที่ดีที่สุด

3. Ex - ante Forecast คือการพยากรณ์ข้อมูลไปข้างหน้า โดยนำแบบจำลองที่ดีที่สุดที่หาได้ในช่วงที่ 2 (Ex-post Forecast) นำแบบจำลองนั้นมาทำการพยากรณ์

การพยากรณ์ทั้งสามช่วงเวลาสามารถแสดงได้ดังรูป 3.5



รูป 3.5 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา: Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ราคาทองคำในการศึกษาครั้งนี้ ใช้วิธีการพรรณนาอธิบายและวิธีการวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองให้กับอนุกรมเวลาในรูปแบบ ARIMA Model โดยวิธีของ Box – Jenkins มีขั้นตอนดังต่อไปนี้

1) การแจกแจงข้อมูลโดยการวาดกราฟระหว่าง ราคาขายทองคำ (y_t) กับเวลา (t) เพื่อพิจารณาแนวโน้มว่าข้อมูลมีเสถียรภาพหรือไม่มีเสถียรภาพ

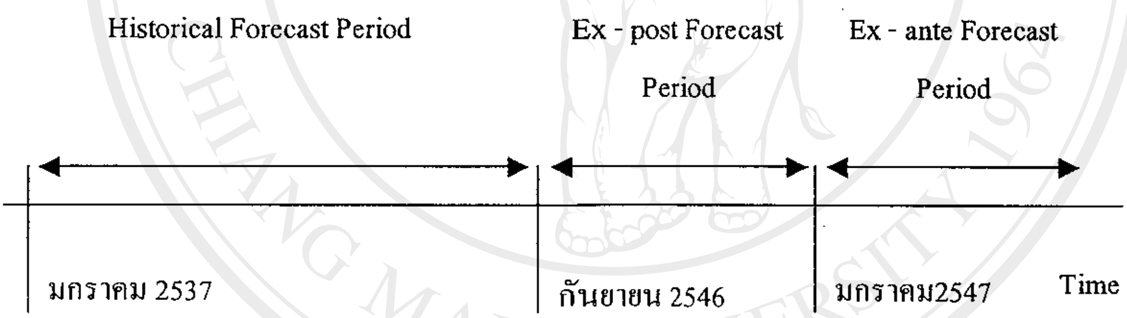
2) เปลี่ยนข้อมูลอนุกรมเวลาราคาขายทองคำ (P_t) ให้อยู่ในรูปลอกาลิทึมธรรมชาติ (Natural Logarithm) ดังนั้นตัวแปรสำหรับการทดสอบ Unit Root จึงเป็น $\ln P_t$

3) การทดสอบ Unit Root โดยวิธี Augmented Dickey – Fuller ดังสมการ (3.10) (3.11) และ (3.12) โดยในการทดสอบนั้นตัวแปรจะอยู่ในรูปลอกาลิทึมธรรมชาติเป็นราคาขายแบ่งเป็นทองแท่งและทองรูปพรรณหน่วยเป็นบาทต่อบาททองคำข้อมูลเป็นรายเดือน (lnP_t)

4) การกำหนดแบบจำลอง ARIMA(p,d,q) โดยพิจารณาคอเรลโลแกรม Autocorrelation Function (ACF) และค่า Partial Autocorrelation Function (PACF) เพื่อระบุว่าแบบจำลองนี้ควรมี Autoregressive, p เท่าใด และ Moving Average, q เท่าใด โดยเลือกสร้างไว้ 4 – 5 แบบจำลองเพื่อหาแบบจำลองที่มีความเหมาะสมที่สุด

5) การตรวจสอบความถูกต้อง โดยการพิจารณา Q – statistic ว่าแตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ เพื่อเข้าสู่ขั้นตอนการพยากรณ์ต่อไป

6) การพยากรณ์ ทำการพยากรณ์ราคาทองแท่งและทองรูปพรรณ โดยทำการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง Historical Forecast, Ex-post Forecast และ Ex-ante Forecast แสดงได้รูป 3.6



รูป 3.6 แสดงช่วงเวลาที่ใช้ในการพยากรณ์จริง