บทที่ 3
แนวความคิดและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

เทคนิคการพยากรณ์ในปัจจุบันได้รับการพัฒนาอย่างต่อเนื่อง ทั้งนี้อาจเป็นเพราะความต้องการเกี่ยวกับการพยากรณ์ในวงการธุรกิจในขณะนี้มีมาก ซึ่งเป็นผลสืบเนื่องมาจากการแข่งขันและความสลับซับซ้อนในการธุรกิจที่มีมากขึ้นก็เป็นได้ และผลของการพยากรณ์ได้มีบทบาทสำคัญในขบวนการตัดสินใจอีกด้วย ในการศึกษาการเคลื่อนไหวอนุกรมเวลามูลค่ารายเดือนของสินค้าที่ใช้วิธีการพรรณนาอธิบายและใช้วิธีวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองอารีมา โดยวิธี Box – Jenkins พร้อมทั้งมีการอธิบายปัจจัยที่กำหนดการ

3.1.1 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลา

การพยากรณ์ดัชนีราคาเหล็ก โดยการกำหนดรูปแบบให้กับอนุกรมเวลา เพื่อที่จะพิจารณาดัชนีราคาเหล็กล่วงหน้า และนำค่าดังกล่าวมาใช้ในการจัดสรรทรัพยากรให้เหมาะสม เพื่อนำมาตัดสินใจวางแผนการผลิตเกี่ยวกับเหล็กให้มีประสิทธภาพดียิ่งขึ้น และให้สอดคล้องกับความต้องการของตลาด

ในการศึกษาครั้งนี้ การพยากรณ์ดัชนีราคาเหล็กจะใช้ข้อมูลเป็นรายเดือน เป็นการใช้วิธีการพรรณนาอธิบายและวิธีการวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองอนุกรมเวลาในรูปแบบ ARIMA Model โดยวิธีของ Box – Jenkins โดยเบื้องต้นต้องพิจารณาว่าอนุกรมเวลามีคุณสมบัติของอนุกรมเวลาที่มีลักษณะนิ่ง (Stationary) หรือไม่ ซึ่งสามารถพิจารณาความนิ่งของข้อมูลได้จาก

ค่าเฉลี่ย : $E(X_t) = \text{Constant} = \mu$ (3.1)

ค่าความแปรปรวน : $V(X_t) = \text{Constant} = \sigma^2$ (3.2)

ค่าความแปรปรวนร่วม : $cov(X_t, X_{t+k}) = E(X_t - \mu)(X_{t+k} - \mu) = \sigma - \mu$ (3.3)

เพราะฉะนั้น ข้อมูลที่มีลักษณะนิ่ง จะต้องมีค่าเฉลี่ย ค่าความแปรปรวนของทุก ๆ ค่า ณ เวลา t ใด ๆ คงที่ ในขณะที่ความแปรปรวนร่วมระหว่างสองคาบเวลาเท่านั้น ไม่ขึ้นอยู่กับเวลาที่เปลี่ยนไป หากเงื่อนไขใดเงื่อนไขหนึ่งไม่เป็นดังที่กล่าวมา แสดงว่าข้อมูลมีลักษณะไม่นิ่ง (Nonstationary) (Charemza and Deadman, 1992)

ดังนั้น สามารถสรุปได้ว่า หากข้อมูลอนุกรมเวลานั้นมีลักษณะที่นิ่ง จะมีค่าเฉลี่ยค่าความแปรปรวน และค่าความแปรปรวนร่วม (จากสมการที่ 3.1, 3.2 และ 3.3) มีค่าที่คงที่ ณ ทุก ๆ เวลาที่เปลี่ยนแปลงไป ซึ่งสามารถทดสอบข้อมูลว่ามีลักษณะที่นิ่งหรือไม่จากการทดสอบ Unit Root

อนุกรมเวลา (Time Series) คือค่าสังเกต (Observation) ชุดหนึ่งซึ่งถูกกำหนดขึ้น ณ เวลาต่าง ๆ ถ้าค่าสังเกตกระทำในเวลาที่ต่อเนื่องกันจะเรียกว่า อนุกรมเวลาต่อเนื่อง แต่ถ้าค่าสังเกตกระทำ ณ จุดเวลาที่ไม่ต่อเนื่องกันเรียกว่า อนุกรมเวลาไม่ต่อเนื่อง ดังนั้น การวิเคราะห์อนุกรมเวลาจึงเป็นการวิเคราะห์ค่าสังเกตที่มีการเปลี่ยนแปลงไปตามเวลาที่กระทำ ลักษณะของการเปลี่ยนแปลงอาจจะมีรูปแบบหรือไม่มีก็ได้ แต่ถ้าอนุกรมเวลามีลักษณะการเปลี่ยนแปลงที่มีรูปแบบก็จะทำให้สามารถที่จะพยากรณ์รูปแบบในอนาคตได้ (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542) ซึ่งประโยชน์ของอนุกรมเวลาที่สำคัญคือ นำมาวิเคราะห์เพื่อพยากรณ์ค่าในอนาคต

การศึกษาอนุกรมเวลาของดัชนีราคาเหล็ก โดยการกำหนดรูปแบบให้กับอนุกรมเวลานั้น สามารถนำไปคาดคะเนความเป็นไปได้ของราคาเหล็กในอนาคต เพื่อใช้เป็นแนวทางในการวางแผนการผลิตให้สอดคล้องกับการจัดสรรทรัพยากรที่มีอยู่อย่างจำกัดให้เหมาะสมต่อไป

3.1.2 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit root test)

การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit Root Test) เป็นการทดสอบว่าตัวแปรแต่ละตัวมีลักษณะนิ่ง หรือไม่นั้นเอง โดยข้อสมมุติ (Assumptions) เบื้องหลังการประมาณค่าทางเศรษฐมิติ โดยใช้ข้อมูลอนุกรมเวลานั้น คือข้อสมมุติเกี่ยวกับความนิ่งของข้อมูล สมมุติว่าเรามีแบบจำลองดังนี้ (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542)

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

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

โดย u_{2t} เป็นอนุกรมเวลาของตัวแปรสุ่ม (Random Variables) ที่มีการแจกแจงแบบปกติที่มีลักษณะเหมือนกันและเป็นอิสระต่อกัน โดยค่าเฉลี่ย (Mean) เท่ากับศูนย์ และค่าความแปรปรวน (Variance) นั้นคงที่ ซึ่งตัวแปร x นั้นเป็นแนวเดินเชิงสุ่ม (Random Walk) และเป็น Integrated of Order one, I (1) เพราะฉะนั้นตัวแปร y ก็จะเป็น I (1) ด้วย ซึ่งโดยทฤษฎีเศรษฐมิติ การถดถอยด้วยตัวแปรที่มีลักษณะไม่นิ่ง (Nonstandard Distribution) จะเป็นการใช้ตารางมาตรฐานที่ใช้กันทั่วไปในการทดสอบค่าสถิติต่าง ๆ และอาจนำไปสู่การลงความเห็นหรือข้อสรุปที่ผิดพลาดได้ ซึ่งนำไปสู่ความเป็นไปได้ของการมีการถดถอยที่ไม่ถูกต้อง (Spurious Regression) (Jonhson and Dinardo, 1997)

สำหรับการทดสอบ Unit Root สามารถทดสอบได้โดยใช้การทดสอบ DF (Dickey – Fuller (DF) Test) (Dickey and Fuller, 1981) และการทดสอบ ADF (Augmented Dickey – Fuller (ADF) Test) (Said and Dickey, 1984) Null Hypothesis ของ DF Test คือ

$H_0: \rho = 1$ จากสมการ (3.3)

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

ซึ่งเรียกว่า Unit Root Test โดยที่ถ้า $|\rho| < 1$ แล้ว X_t จะลักษณะนิ่ง ; และถ้า $\rho = 1$ X_t จะมีลักษณะไม่นิ่ง อย่างไรก็ตามการทดสอบนี้สามารถทำได้อีกทางหนึ่ง ซึ่งเหมือนกับสมการ (3.3) กล่าวคือ

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \qquad (3.7)$$

ซึ่งก็คือ $X_t = (1+\theta) X_{t-1} + \varepsilon_t$ ซึ่งคือสมการที่ (3.6) นั่นเอง โดยที่ $\rho = (1 + \theta)$ ถ้า θ ในสมการ (4) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3.6) มีค่าน้อยกว่า 1 ดังนั้น สามารถสรุปได้ว่า การปฏิเสธ $H_0: \rho = 0$ ซึ่งเป็นการยอมรับ $H_a: \theta < 0$ หมายความว่า $\rho < 1$ และ X_t มี Integration of Order Zero (Charemza and Deadman, 1992) นั่นคือ X_t มีลักษณะนิ่ง แต่ถ้าเราไม่สามารถปฏิเสธ $H_0: \theta = 1$ ได้ ก็จะหมายความว่า X_t มีลักษณะไม่นิ่ง ถ้า X_t มีแนวคิดเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (Random Walk With Drift) เราเขียนแบบจำลองได้ดังนี้

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

และถ้า X_t มีแนวโน้มตามเวลาเชิงเส้น เราเขียนแบบจำลองได้ดังนี้

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

โดยที่ t = เวลา ซึ่งจะทำการทดสอบ $H_0: \theta = 0$ โดยมี $H_a: \theta < 0$ เช่นเดียวกับที่กล่าวมาข้างต้น โดยสรุปแล้ว Dickey and Fuller ได้พิจารณาสมการทดถอย 3 รูปแบบที่แตกต่างกันในการทดสอบว่ามี Unit Root หรือไม่ ซึ่งสมการดังกล่าว ได้แก่

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \tag{3.10}$$

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{3.11}$$

$$\Delta X_t = \alpha + \beta t + \theta X t_{t-1} + \varepsilon_t \tag{3.12}$$

ตัวพารามิเตอร์ ที่อยู่ในความสนใจของทุกสมการ คือ θ นั่นคือ ถ้า $\theta = 0$; X_t จะมี Unit Root โดยการเปรียบเทียบ t – Statistic ที่คำนวณได้กับค่าที่เหมาะสมอยู่ใน Dickey and Fuller Table (Enders, 1995) หรือ กับ MacKinnon Critical Values (Gujarati, 1995)

อย่างไรก็ตาม Critical Values จะไม่เปลี่ยนแปลง ถ้าสมการ (3.6), (3.7) และ (3.8) ถูกแทนที่โดย Autoregressive Process (Enders, 1995 และ Gujarati, 1995)

$$\Delta x_t = \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{3.13}$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{3.14}$$

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{3.15}$$

จำนวนของ Lagged Difference Terms ที่นำเข้ามารวมในสมการนั้นต้องมีมากพอที่จะทำให้ Error Terms มีลักษณะเป็น Serially Independent และเมื่อนำเอา Diceky – Fuller (DF) Test มาใช้กับสมการ (3.13) (3.14) และ (3.15) เรียกว่า Augmented Dickey – Fuller

(ADF) Test ซึ่ง ADF Test Statistic มีการแจกแจงแบบ Asymptotic Distribution เหมือนกับ DF Statistic ดังนั้น จึงสามารถใช้ Critical Values แบบเดียวกัน (Gujarati, 1995)

โดยวิธีการหา lag length ที่เหมาะสมนั้น Enders (1995) ได้เสนอว่าให้เริ่มต้นที่ lag length ที่มากพอสมควรค่าหนึ่งแล้วค่อย ๆ ลดค่าลงเรื่อย ๆ โดยใช้ค่าสถิติทดสอบ t (t – test) หรือค่าสถิติทดสอบ F (F – test) เมื่อทดสอบแล้วพบว่าค่าสถิติ t – test หรือ F – test ที่ใช้ในการทดสอบนั้นไม่มีนัยสำคัญ ณ ค่าวิกฤต (Critical Value) ที่กำหนดให้ ต้องทำการทดสอบใหม่โดยทำการลดค่า lag length จนกระทั่งค่าสถิติมีนัยสำคัญจึงจะถือว่าค่า Lag นั้นมีความเหมาะสม สมมุติว่าเราใช้ lag length ที่ n* ถ้าค่าสถิติ t – test หรือ F – test ของ lag n* ไม่มีนัยสำคัญ ณ ค่าวิกฤต (Critical Value) ที่กำหนดให้ เราต้องทำการประมาณค่าการถดถอยใหม่โดยให้ lag length เท่ากับ n*-1 ทำอย่างนี้ไปเรื่อย ๆ จนกระทั่งค่าสถิติมีนัยสำคัญ lag นั้นจึงถือว่ามีความเหมาะสม

3.1.3 แนวคิดการวิเคราะห์ข้อมูลอนุกรมเวลา Box-Jenkins

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

วิธีของ Box – Jenkins เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา แบบจำลองที่ใช้ในการพยากรณ์ คือ ตัวแบบ ARIMA (p,d,q) ซึ่งมีส่วนประกอบที่สำคัญ 3 ส่วน ได้แก่ Autoregreesive AR: (p), Intergrated (I) และ Moving Average MA: (q) สำหรับ AR (p) เป็นรูปแบบที่แสดงว่า ค่าสังเกต Y_t จะขึ้นอยู่กับค่าของ Y_{t-1} ,..., Y_{t-p} หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA (q) หมายถึง รูปแบบที่แสดงให้เห็นว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่าของความคลาดเคลื่อน ε_{t-1} ,..., ε_{t-q} หรือความคลาดเคลื่อนที่อยู่ก่อนหน้า q ค่า ส่วน Intergrated (I) เป็นการหาผลต่าง (Difference) ของอนุกรมเวลา และเหตุผลสำคัญที่ต้องหาผลต่างของอนุกรมเวลา เนื่องจากแบบจำลอง ARIMA จะใช้ได้กับตัวแปรหรืออนุกรมเวลาที่ไม่มีปัจจัยแนวโน้ม (Trend) หรือมีคุณสมบัติเป็น Stationary เท่านั้น ในกรณีที่ตัวแปรมีปัจจัยแนวโน้มรวมอยู่ด้วย ต้องขจัดปัจจัยแนวโน้มออกไปก่อน โดยการหาผลต่างระหว่างค่าของตัวแปรในช่วงเวลาติดกัน ซึ่งรูปแบบ AR MA (p,q) มีการกำหนดรูปแบบสมการดังนี้

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

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

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

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

การกำหนดรูปแบบที่เหมาะสมให้กับอนุกรมเวลา จะใช้ค่าสัมประสิทธิ์อัตตสหสัมพันธ์แบบออโต (Autocorrelation Function: ACF) และค่าสัมประสิทธิ์อัตตสหสัมพันธ์บางส่วนแบบออโต (Partial Autocorrelation)

การเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่าง ๆ ได้แก่ แนวโน้ม (Trend) ตัวแปรฤดูกาล (Seasonal Factor) ตัวแปรวัฏจักร (Cyclical Factor) และเหตุการณ์ที่ผิดปกติ (Irregular Movement) โดยวิธี Box - Jenkins จะสามารถแบ่งอนุกรมเวลาออกเป็น 2 ส่วนดังนี้

1. อนุกรมเวลาที่เป็น stationary series เป็นอนุกรมเวลา $\{Y_t\}$ ที่มีค่าเฉลี่ยและความแปรปรวนของ Y_t คงที่ นั่นคือมีคุณสมบัติทำให้ค่าเฉลี่ย $E(Y_t)$ และค่าความแปรปรวน $V(Y_t)$ มีค่าคงที่สำหรับแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้มและ/หรืออิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(Y_t)$ ไม่คงที่ และอนุกรมเวลาที่มีความแปรปรวนของ Y_t สูง จะเป็นลักษณะของอนุกรมเวลาที่ $V(Y_t)$ มีค่าไม่คงที่ โดยจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่เป็น Stationary Series นอกจากนั้นอนุกรมเวลาที่เป็น stationary series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้ว ยังจะต้องมีค่าสัมประสิทธิ์อัตตสหสัมพันธ์แบบออโตที่ lag K ขึ้นอยู่กับค่า K อย่างเดียว อนุกรมเวลาที่กำหนดรูปแบบ ARMA (p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น stationary series แล้ว

2. อนุกรมเวลาที่ไม่เป็น stationary series เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น Stationary Series การจะหารูปแบบ ARMA (p,q) ให้กับอนุกรมเวลาดังกล่าวได้จะต้องดำเนินการหาผลต่างของอนุกรมเวลาชุดนั้นแล้วนำไปทดสอบซ้ำ จนกระทั่งได้ข้อมูลที่มีคุณสมบัติ Stationary Series การแปลงอนุกรมเวลาที่ไม่เป็น stationary series ให้เป็นอนุกรมเวลาที่เป็น Stationary Series อาจทำได้ด้วยวิธีการต่าง ๆ ดังนี้

2.1 การหาผลต่างปกติ (Regular differencing) ของอนุกรมเวลา เป็นวิธีการกำจัดแนวโน้ม คือ ถ้าอนุกรมเวลา $\{Y_t\}$ มีแนวโน้มอยู่ในอนุกรมเวลา ก็จะแปลงให้เป็นอนุกรม

เวลาชุดใหม่ที่ไม่มีแนวโน้ม $\{Z_t\}$ โดย $Z_t = \nabla^d Y_t$ โดยที่ d เป็นลำดับของการหาผลต่าง และ ∇ 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 จำนวนครั้งที่หาผลต่างนั้น จะขึ้นอยู่กับว่าหาผลต่างแล้วอนุกรมเวลาใหม่เป็นแบบ stationary series หรือไม่ ถ้ายังไม่เป็นแบบ stationeries series ต้องหาผลต่างต่อไปเรื่อย ๆ โดยทั่วไปแล้ว ถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ d = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติก (Quadratic) จะใช้ d = 2

2.2 การหาผลต่างฤดูกาลของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรอื่นเข้ามาเกี่ยวข้อง 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 $Z_t = \nabla^2(Y_t - Y_{t-12})$ เป็นต้นผลต่างนี้จะทำกี่ครั้งขึ้นกับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็นแบบ stationary series หรือไม่ ถ้ายังไม่เป็นแบบ stationary series จะต้องหาผลต่างต่อไป

2.3 การหาผลต่างปกติและผลต่างฤดูกาล ในกรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็นแบบ stationary series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาล โดยจะพิจารณาค่า d และ D ควบคู่กันไป ซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และ D จะมีค่าเท่าใดนั้นขึ้นอยู่กับว่า เมื่อหาผลต่างปกติและผลต่างฤดูกาลแล้ว อนุกรมเวลาใหม่เป็นแบบ stationary series หรือไม่ ถ้ายังไม่เป็นแบบ 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 คือ การแปลงอนุกรมเวลาเดิม $\{Y_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{Z_t\}$ โดย $Z_t = \ln(Y_t)$ การแปลงครั้งนี้จะทำเมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ นั่นคือ $V(Y_t)$ สำหรับค่าเวลา t ต่าง ๆ

3.1.4 แบบจำลองการพยากรณ์ โดยวิธี Box – Jenkins

ในการพยากรณ์โดยวิธีของ Box – Jenkins ในรูปแบบ ARIMA นั้นต้องพิจารณาอนุกรมเวลา $\{Y_t\}$ ให้มีคุณสมบัติของอนุกรมเวลาที่เป็นแบบ stationary series เสียก่อน และการพิจารณาอนุกรมเวลาว่าเป็นแบบ stationary series หรือไม่ จะพิจารณาได้จาก

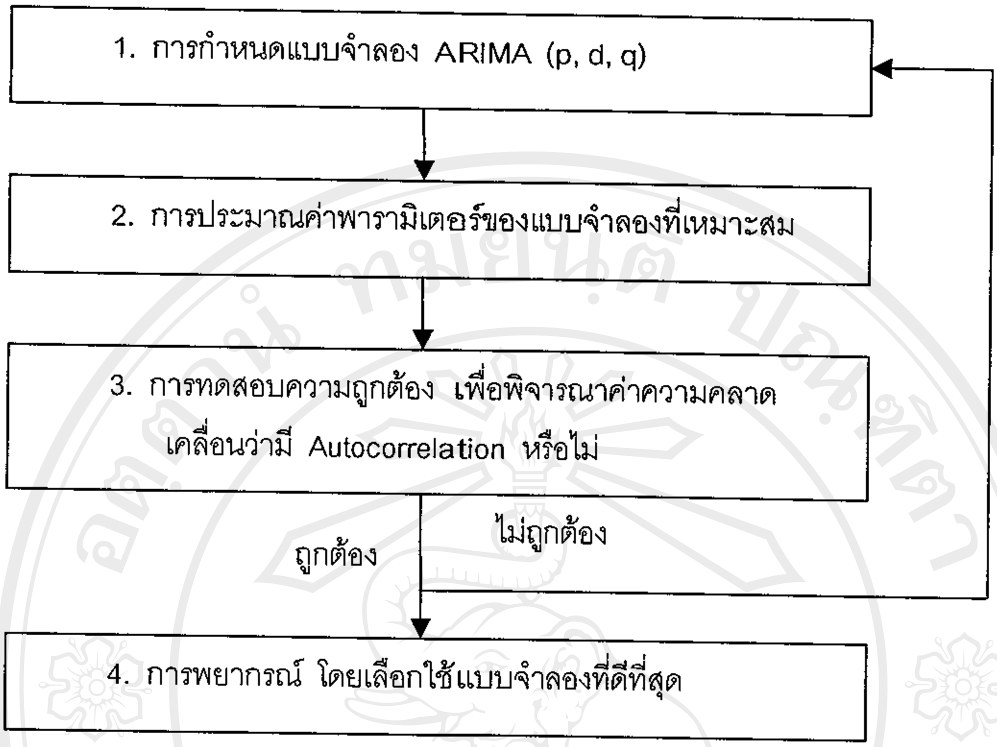
1. ค่าเฉลี่ย $E(Y_t)$ นั้นคงที่สำหรับทุกค่าของ t หรือไม่ จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วน ๆ แล้วหาค่าเฉลี่ยของอนุกรมเวลาในแต่ละส่วน ถ้าค่าเฉลี่ยในแต่ละส่วนย่อยไม่แตกต่างกันมาก จะสรุปได้ว่า $E(Y_t)$ คงที่

2. ค่าความแปรปรวน $V(Y_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่ จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วน ๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาในแต่ละส่วน ถ้าค่าความแปรปรวนในแต่ละส่วนย่อยไม่แตกต่างกันมากนัก จะสรุปได้ว่า $V(Y_t)$ คงที่

3. พิจารณาแนวโน้มและ/หรือปัจจัยฤดูกาล ด้วยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและ/หรือปัจจัยฤดูกาล มักจะเห็นชัดเจนได้จากรูปกราฟ

4. พิจารณาจากคอรีโลแกรม (Correlogram) ของค่าสัมประสิทธิ์อัตตสหสัมพันธ์แบบออโตของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ stationary ค่าคอรีโลแกรม (correlogram) ของ autocorrelation (r_k) จะมีค่าลดลงค่อนข้างเร็ว เมื่อ k มีค่าเพิ่มมากขึ้น ดังนั้น ถ้าค่า autocorrelation (r_k) มีค่าลดลงค่อนข้างช้า จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่า autocorrlation (r_k) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างสูง ที่ k = L, 2L, 3L จะเป็นข้อสังเกตว่าอนุกรมชุดนี้มีแนวโน้มและอิทธิพลฤดูกาล และถ้าการเคลื่อนไหวของค่า correlogram ของ autocorrelation (r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้ว พบว่า อนุกรมเวลาที่ศึกษาไม่เป็นแบบ stationary ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่เป็นแบบ stationary จะต้องแปลงอนุกรมเวลาให้เป็นแบบ stationary เสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็นแบบ stationary แต่ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาล ให้หาผลต่างและผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็นแบบ stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ ให้แปลงอนุกรมเวลาเดิมโดยการหาลอการิทึม ($Z_t = \ln Y_t$) จนกว่าจะได้อนุกรมเวลาอันใหม่ ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่ที่เป็นแบบ stationary series แล้วจะทำตามขั้นตอนของ Box – Jenkins ดังรูปที่ 3.1



รูปที่ 3.1 แสดงขั้นตอนการพยากรณ์โดยวิธี Box – Jenkins
ที่มา: Gujarati (2003)

ขั้นตอนการพยากรณ์โดยวิธีของ Box - Jenkins มี 4 ขั้นตอนได้แก่

1. การกำหนดแบบรูปแบบ (Indentification) ให้กับอนุกรมเวลาที่เป็นแบบ 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) ของอนุกรมเวลาตัวอย่างกับค่า autocorrelation (ρ_k) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

โดยที่ $\gamma_k = \mathrm{cov}(Z_t, Z_{t-k}) = E[(Z_t - \mu)(Z_{t-k} - \mu)]$

$$\gamma_0 = \mathrm{cov}(Z_t, Z_{t-0}) = E[(Z_t - \mu)^2]$$

แต่ ρ_k เป็นการประมาณค่าของประชากร เป็นการประมาณค่าที่จะต้องสุ่มมาจากตัวอย่างของประชากร ซึ่งจะทำได้ง่ายและประหยัด ดังนั้นจึงได้กำหนด r_k เป็น autocorrelation ที่มาจากตัวอย่างโดยมีสูตรดังนี้ (วิชิต หล่อจีระชุณห์กุล, 2539)

$$r_k = \frac{c_k}{c_0}$$

โดยที่ $c_k = \mathrm{cov}(Z_t, Z_{t-k}) = E[(Z_t - \bar{Z})(Z_{t-k} - \bar{Z})]$

$c_0 = \mathrm{cov}(Z_t, Z_{t-0}) = E[(Z_t - \bar{Z})^2]$

ซึ่งสามารถพิสูจน์ เป็นสมการได้ดังนี้

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

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตาม เนื่องจากอนุกรมเวลาต้องเผชิญกับปัญหาสหสัมพันธ์ ทั้งที่เกิดจากตัวตามแปรอิสระที่เป็นค่าความล่าช้า (lag) ของตัวแปรตาม (autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (moving average) เนื่องจาก autocorrelation function (ACF) ซึ่งจะใช้ในการอธิบายอัตตสหสัมพันธ์ของค่าความคลาดเคลื่อน แต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตามได้ ซึ่ง partial autocorrelation function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ซึ่งสามารถพิจารณาได้จากสมการ Yule – walker (Pindyck and Rubinfeld, 1996) ดังนี้

$$\rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \ldots + \phi_p \qquad (3.16)$$

ถ้า k มากกว่า q จะได้

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p} \qquad (3.17)$$

การกำหนดลำดับขั้น p,q ในแบบจำลอง (Identiftying the Dependence Order of Model) ขั้นตอนนี้คือ การระบุว่าแบบจำลองควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใด และ moving average, q เท่าใด โดยพิจารณาจากค่า ACF และ PACF ซึ่งแสดงดังตารางที่ 3.1 ดังต่อไปนี้

ตารางที่ 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)

จากตารางที่ 3.1 จะสามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีค่าลดลงอย่างช้าแทนที่จะเป็นศูนย์เมื่อช่วงเวลาห่างกัน q เป็นต้นไปหรือมีค่าลดลงเข้าสู่ศูนย์อย่างรวดเร็ว ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็น ค่าที่ p ของ AR(p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรม เกิดขึ้น 1 แท่ง หมายความได้ว่าแบบจำลองควรมีลักษณะเป็น AR(1)

สำหรับ MA(q) นั้นก็จะมี ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF มีค่าลดลงอย่างช้าแทนที่จะเป็นศูนย์เมื่อช่วงเวลาห่างกัน q เป็นต้นไปหรือมีค่าลดลงเข้าสู่ศูนย์อย่างรวดเร็ว ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สามารถสรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(2)

หาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรจะเป็น ARMA(p,q) และเมื่อรวมกันกับการทดสอบความนิ่ง ในขั้นตอนที่ 1 แล้ว จะสามารถหาค่าของ difference ได้ ซึ่งผลจากการ difference จำนวน d ครั้งนั้น ก็จะได้แบบจำลอง ARIMA(p,d,q) แต่อย่างไรก็ตาม หลักการดังกล่าวก็เป็นเพียงเครื่องช่วยการพิจารณาในระดับหนึ่งเท่านั้น ดังนั้นเพื่อประเมินแบบ

จำลอง ว่าแบบจำลองใดมีความเหมาะสมที่จะใช้เป็นตัวประมาณค่าของกลุ่มข้อมูลจริง สามารถพิจารณาได้จากค่าสถิติดังต่อไปนี้เพื่อประกอบในการตัดสินใจ

ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMS) โดยจะเป็นการวัดค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองว่ามีความแตกต่างกันมากน้อยเพียงใด ซึ่งหากค่า RMSE มีค่าเท่ากับศูนย์ จะหมายความว่าแบบจำลองที่ประมาณได้นั้น มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงไร ก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากน้อยเท่านั้น ซึ่งสามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)} \tag{3.18}$$

กำหนดให้ Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

นอกจากจะใช้ค่า RMSE ประเมินความถูกต้องของแบบจำลองแล้ว ยังสามารถพิจารณาค่าสถิติตัวอื่น ๆ เช่น Theil's Inequality Coefficient โดยในหลักการเบื้องต้นพบว่า สมการที่ใช้นี้มีหลักการที่คล้ายกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือ ค่าสถิติจะมีค่าอยู่ระหว่าง 0 และ 1 ทั้งนี้หากค่า TIC มีค่าเท่ากับศูนย์ หมายความว่า ค่าที่ได้จากการประมาณจะมีค่าเท่ากันพอดีกับค่าที่เป็นข้อมูลจริง แสดงถึงแบบจำลองที่ประมาณได้ว่า เป็นแบบจำลองที่เป็นตัวแทนข้อมูลจริงได้อย่างดีที่สุด ในขณะที่ถ้า TIC มีค่าเท่ากับหนึ่ง แปลว่า แบบจำลองที่ประมาณได้เป็นแบบจำลองที่แย่ที่สุด ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่า TIC ที่น้อย ๆ ดังจะพิจารณาได้จากสมการที่ (3.19)

$$TIC = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^w - Y_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^s\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^a\right)^2}} \tag{3.19}$$

กำหนดให้ Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

อย่างไรก็ตาม ยังมีค่าสถิติอีกหลายอย่างที่สามารถนำมาพิจารณาประกอบร่วมกับ RMSE และ Theil's Inequality Coefficient เพื่อใช้ในการเลือกแบบจำลองที่เหมาะสมสำหรับการวิเคราะห์ที่ดีที่สุด อาทิ เช่น R^2 Adjusted R^2 และ Akaike Information Criterion (AIC) ซึ่งสามารถอธิบายได้ดังต่อไปนี้

R^2 คือ การวัดค่าตัวแปรอิสระว่าสามารถอธิบายตัวแปรตามได้ดีเพียงใด หากค่านี้เท่ากับ 1 หมายความว่า ตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% ในทางกลับกัน หากค่านี้มีค่าเท่ากับ 0 แปลความหมายได้ว่า ตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่อย่างไรก็ตาม พบว่า หากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมาก ๆ นั้นจะทำให้ค่า R^2 มากขึ้นด้วย ทำให้เกิดความคลาดเคลื่อน ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการที่ (3.20) ดังนั้นเพื่อปรับปรุงข้อจำกัดข้างต้น จึงมีการคิดค่าสถิติใหม่คือ ค่า Adjusted R^2 ($\bar{R}^2$) ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการที่ (3.21)

$$R^2 = 1 - \frac{\Sigma \hat{u}_i}{\Sigma y_i^2} \tag{3.20}$$

$$\bar{R}^2 = 1 - \frac{\Sigma \hat{u}_i^2 /(n-k)}{\Sigma y_i^2 /(n-1)} \tag{3.21}$$

Akaike Information Criterion (AIC) คือ ค่าสถิติที่ประยุกต์คล้ายกับ $\bar{R}^2$ แต่ใช้ในรูปแบบของค่าลอกกาลิทึมฐานธรรมชาติ (natural logarithm) โดยหากค่าสถิตินี้มีค่าน้อยเพียงใด นั่นหมายความว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น อีกทั้งค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (lag length) ที่เหมาะสมที่สุดได้อีกด้วย ดังแสดงในสมการที่ (3.20)

$$\ln AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{\Sigma \hat{u}_i^2}{n}\right) \tag{3.22}$$

กำหนดให้ $\Sigma \hat{u}_i^2$ คือ ผลรวมของค่าความคลาดเคลื่อน

n คือ ค่าสังเกตทั้งหมด

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

2. การประมาณค่าพารามิเตอร์ในรูปแบบ (Estimation) ทำได้โดยการหาค่าประมาณแบบง่ายหรือค่าประมาณที่ได้จากการวิเคราะห์ตัวเลข (numerical analysis) สำหรับค่าประมาณแบบง่ายจะทำโดยการสร้างสมการที่มาจากความสัมพันธ์ระหว่าง ρ_k และ ตัวพารามิเตอร์ โดยสมการที่สร้างขึ้นจะมีจำนวนเท่ากับจำนวนพารามิเตอร์ที่ต้องการประมาณค่า ส่วนค่าประมาณที่ได้จากการวิเคราะห์ตัวเลขจะได้จากการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด (Ordinary Least – Squares: OLS) ขั้นตอนของการวิเคราะห์ตัวเลขจะต้องมีการกำหนดค่าประมาณเริ่มต้น ซึ่งส่วนใหญ่จะใช้การประมาณแบบง่ายเป็นค่าประมาณเริ่มต้น เมื่อการวิเคราะห์สิ้นสุดจะได้ค่าประมาณสุดท้ายที่จะทำไปใช้ประโยชน์ในการสร้างสมการพยากรณ์ต่อไป

3. การตรวจสอบแบบจำลอง (Diagnostics) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลองแล้ว จะต้องทำการตรวจสอบทุกครั้งว่า รูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธี ยกตัวอย่าง เช่น การพิจารณาคอรีโลแกรมของอัตตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลอง โดยใช้การทดสอบของ Box และ Pierce ซึ่งจะแสดงได้โดยใช้ Q-statistic ดังในสมการที่ (3.21)

$$Q = n\sum_{k=1}^{m}\rho_k^2 \tag{3.23}$$

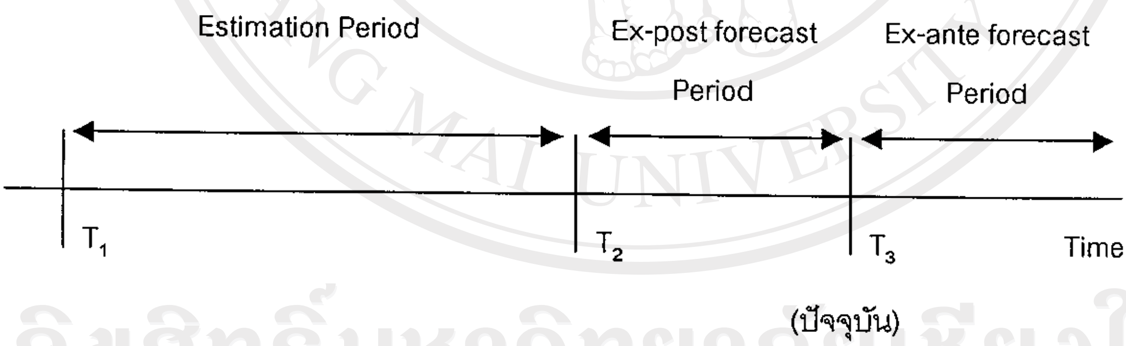
กำหนดให้ n คือ จำนวนของข้อมูล

m คือ ค่า lag length

จากสมการที่ (3.17) ค่า Q นั้นจะพบว่า มีการแจกแจงแบบ Chi – square ที่มีองศาอิสระ (degree of freedom) เท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมติฐานที่ว่า สมมติฐานว่างคือ

ค่าความคลาดเคลื่อนที่ได้จากการประมาณที่มีลักษณะเป็น white noise หมายความว่า แบบจำลองไม่มีอัตตสหสัมพันธ์ (autocorrelation) ดังนั้นหากตรวจสอบพบว่า แบบจำลองนั้นไม่มีอัตตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

4. การการพยากรณ์ (Forecasting) เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากการวิเคราะห์ความถูกต้องแล้ว จะสามารถนำแบบจำลองมาใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด ดังนั้นการพยากรณ์จึงจำเป็นต้องมีการทดสอบแบบจำลองโดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง Historical Forecast อันเป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจาณา (T_2) โดยการพยากรณ์ในช่วง Ex – Post Forecast คือ การพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์ จากนั้นเปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า Root Mean Square Error (RMSE) และ Theil's Inequality Coefficient (TIC) และค่า Akaike Information Criterion (AIC) โดยจะพิจารณาค่าสถิติทั้งสามค่า แล้วเลือกค่าสถิติที่มีค่าน้อยที่สุด ซึ่งได้จากการทำการพยากรณ์ เมื่อเลือกรูปแบบจำลองที่ดีที่สุดแล้วจึงนำแบบจำลองนั้นมาทำการพยากรณ์แบบ Ex – Ante Forecast ซึ่งเป็นการพยากรณ์ข้อมูลไปข้างหน้า ดังแสดงในรูปที่ 3.2



รูปที่ 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา: Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ดัชนีราคาเหล็กในครั้งนี้ จะใช้วิธีการพรรณนาอธิบายและวิธีการวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองอนุกรมเวลาในรูปแบบ ARIMA Model โดยวิธีของ Box – Jenkins สามารถสรุปการสร้างแบบจำลองด้วย ARIMA (Gujarati, 2003) ได้ ดังนี้คือ

1. ค่าลอกกาลิทึมฐานธรรมชาติ (Natural logarithm: ln) ข้อมูลในอนุกรมเวลาของดัชนีราคาเหล็กและ X_t ได้ถูกเปลี่ยนเป็น ST_t ดังนั้น ตัวแปรที่จะใช้ในสมการต่าง ๆ จึงถูกเปลี่ยนเป็น $\ln ST_t$ ซึ่งหมายถึงดัชนีราคาเหล็ก

2. การทดสอบความนิ่งของข้อมูล (Unit Root test) เป็นการพิจารณาว่าข้อมูลอนุกรมมีลักษณะนิ่งหรือไม่ โดยการทดสอบ Unit Root ดังสมการต่อไปนี้

$$\Delta x_t = \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \qquad (3.24)$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \qquad (3.25)$$

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \qquad (3.26)$$

หากข้อมูลที่วาดกราฟนั้นมีค่าเพิ่มขึ้นเรื่อย ๆ ตามเวลาเปลี่ยนแปลงเพิ่มมากขึ้นหรือการทดสอบ Unit Root ของข้อมูล ถ้าข้อมูลนั้นพบว่ามี Unit Root แสดงว่าข้อมูลมีลักษณะไม่นิ่ง จะต้องทำการเปลี่ยนแปลงรูปแบบโดยการทำ Difference แต่ถ้าหากตรวจสอบแล้ว พบว่ามีลักษณะนิ่ง สามารถทำในขั้นตอนถัดไป

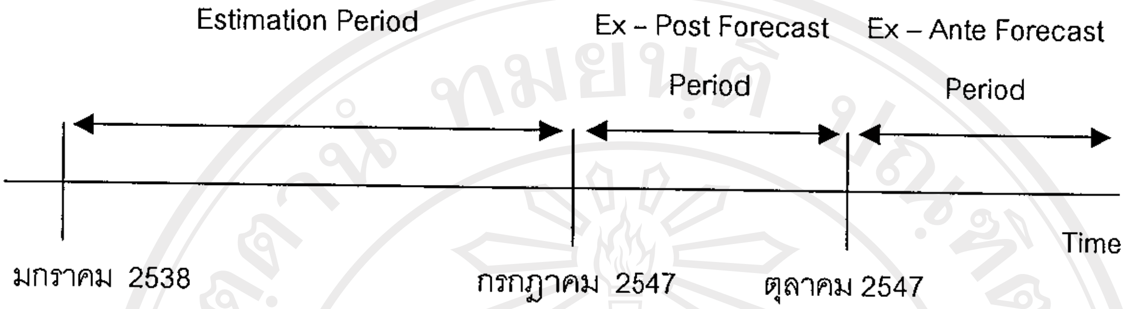
3. การกำหนดลำดับขั้น p,q ในแบบจำลอง (Identiftying the Dependence Order of Model) ขั้นตอนนี้คือ การระบุว่าแบบจำลองนี้ควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใด และ moving average, q เท่าใด โดยพิจารณาจาก ACF และ PACF โดยควรสร้างหลาย ๆ แบบจำลอง เพื่อเปรียบเทียบและเลือกหาแบบจำลองที่ดีและเหมาะสมที่สุดในการพยากรณ์

4. การประมาณค่าพารามิเตอร์ เพื่อนำค่าพารามิเตอร์ที่ได้นั้น ไปทำการพยากรณ์ราคาต่อไป

5. ตรวจสอบรูปแบบ (Diagnostic checking) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในรูปแบบแล้ว ต้องตรวจสอบว่ารูปแบบที่ได้มีความเหมาะสมจริงหรือไม่ โดยทดสอบค่าพารามิเตอร์จากค่า t – test แล้วดูค่า Root Mean Squared Error (RMSE) และ ค่า Theil's Inequality Coefficient (TIC) เลือกรูปแบบที่ให้ค่าสถิติเหล่านี้ที่ดีที่สุด (มีค่าเข้าใกล้ศูนย์)

6. การพยากรณ์ โดยใช้สมการพยากรณ์ที่สร้างจากรูปแบบการพยากรณ์ที่กำหนดและผ่านการตรวจสอบตามขั้นตอนแล้วมาพยากรณ์ผลที่คาดว่าจะเกิดขึ้นในช่วงเวลาต่อ ๆ ไป ซึ่งการพยากรณ์โดยวิธีการของบ๊อกซ์และเจนกินส์ จะให้ค่าพยากรณ์ได้ดีในช่วงเวลาสั้น ๆ โดยนำค่าที่เกิดขึ้นจริงในปีล่าสุดมาใส่ในสมการและพยากรณ์ผลในช่วงเวลาถัดไป ซึ่งจะทำการแบ่ง

การพยากรณ์ออกเป็น 3 ช่วง ได้แก่ช่วง Historical Forecast ช่วง Ex – Post Forecast และช่วง Ex – Ante Forecast ดังแสดงในรูปที่ 3.3



รูปที่ 3.3 แสดงช่วงเวลาที่ใช้ในการพยากรณ์จริง