บทที่ 3

แนวความคิดและระเบียบวิธีวิจัย

3.1 แนวความคิดที่ใช้ในการศึกษา

เทคนิคการพยากรณ์ในปัจจุบันได้รับการพัฒนาอย่างต่อเนื่อง ทั้งนี้อาจเป็นเพราะความต้องการเกี่ยวกับการพยากรณ์ในวงการธุรกิจในขณะนี้มีมาก ซึ่งเป็นผลสืบเนื่องมาจากการแข่งขันและความสลับซับซ้อนในการดำเนินธุรกิจที่มีมากขึ้นก็เป็นได้ และผลของการพยากรณ์ได้มีบทบาทสำคัญในขบวนการตัดสินใจอีกด้วย

3.1.1 แนวคิดการวิเคราะห์ข้อมูลอนุกรมเวลา (Time-Series Analysis)

อนุกรมเวลา (time-series) คือ ค่าสังเกต (observation) ชุดหนึ่งซึ่งถูกกำหนดขึ้น ณ เวลาต่าง ๆ ถ้าค่าสังเกตกระทำในเวลาที่ต่อเนื่องกันจะเรียกว่าอนุกรมเวลาต่อเนื่อง แต่ถ้าค่าสังเกตกระทำ ณ จุดเวลาที่ไม่ต่อเนื่องกันจะเรียกว่าอนุกรมเวลาไม่ต่อเนื่อง ดังนั้นการวิเคราะห์อนุกรมเวลาจึงเป็นการวิเคราะห์ค่าสังเกตที่มีการเปลี่ยนแปลงไปตามเวลาที่กระทำ ลักษณะของการเปลี่ยนแปลงอาจจะมีรูปแบบหรือไม่มีก็ได้ แต่ถ้าอนุกรมเวลามีลักษณะการเปลี่ยนแปลงที่มีรูปแบบก็จะทำให้สามารถที่จะพยากรณ์รูปแบบในอนาคตได้ (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542)

การเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่างๆได้แก่ แนวโน้ม (trend) ตัวแปรฤดูกาล (seasonal factor) ตัวแปรวัฏจักร (cyclical factor) และเหตุการณ์ที่ผิดปกติ (irregular movement) โดยวิธี Box – Jenkins จะสามารถแบ่งอนุกรมเวลาออกเป็น 2 ประเภทดังนี้

1) อนุกรมเวลาที่เป็น stationary series คืออนุกรมเวลา $\{Y_t\}$ ที่มีค่าเฉลี่ยและค่าความแปรปรวนของ Y_t คงที่ นั่นคือค่าเฉลี่ย $E(Y_t)$ และค่าความแปรปรวน $V(Y_t)$ มีค่าคงที่สำหรับอนุกรมแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้มและ/หรืออิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(Y_t)$ ไม่คงที่ และอนุกรมเวลาที่มีความแปรปรวนของ Y_t สูงจะเป็นลักษณะของอนุกรมเวลาที่ $V(Y_t)$ มีค่าไม่คงที่ ซึ่งจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่เป็น non stationary series นอกจากนั้นอนุกรมเวลาที่เป็น stationary series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ย และความแปรปรวนคงที่แล้วยังจะต้องมีค่าสัมประสิทธิ์ สหสัมพันธ์แบบออโตที่ lag k ขึ้นอยู่กับค่า k อย่าง

เดียว อนุกรมเวลาที่กำหนดรูปแบบ ARMA (p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น stationary series แล้ว

2) อนุกรมเวลาที่เป็น non stationary series เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น stationary series การจะหารูปแบบ ARMA (p,q) ให้กับอนุกรมเวลาดังกล่าวได้จะต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ stationary series เสียก่อน การแปลงอนุกรมเวลาที่เป็น non stationary series ให้เป็นอนุกรมเวลาที่เป็น stationary series อาจทำได้ด้วยวิธีการต่าง ๆ ดังนี้

2.1) การหาผลต่างปกติ (regular differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้ม นั่นคือ ถ้าอนุกรมเวลา $\{Y_t\}$ มีแนวโน้มอยู่ในอนุกรมเวลา จะแปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม $\{Z_t\}$ โดย $Z_t = \nabla^d Y_t$ โดย d เป็นลำดับของการหาผลต่าง และ ∇ คือผลต่างของตัวแปร เช่น เมื่อ d = 1 จะได้ $Z_t = \nabla Y_t = Y_t - Y_{t-1}$ 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 จำนวนครั้งที่หาผลต่าง จะขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น stationary series หรือไม่ ถ้ายังเป็น non stationary series ต้องหาผลต่างต่อไป โดยทั่วไปถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ d = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติกจะใช้ d=2

2.2) การหาผลต่างฤดูกาลของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรฤดูกาลเข้ามาเกี่ยวข้อง จะต้องแปลงอนุกรมเวลาเดิม $\{Y_t\}$ 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 $Z_t = Y_t - Y_{t-12}$ และเมื่อ D = 2 จะได้ $Z = \nabla^2_{12} Y_t$ หรือ $Z = \nabla (Y_t - Y_{t-12}) = Y_t - 2Y_{t-12} + Y_{t-24}$ เป็นต้น ผลต่างนี้จะทำกี่ครั้งขึ้นกับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น stationary series หรือไม่ ถ้ายังเป็น non stationary series ต้องหาผลต่างต่อไป

2.3) การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น stationary series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาล d และ D ควบคู่กันไป ซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และ D จะมีค่าเท่าไรนั้นขึ้นอยู่กับว่าเมื่อหาผลต่างและผลต่างฤดูกาลแล้วอนุกรมเวลาใหม่เป็น stationary series หรือไม่ ถ้ายังเป็น non stationary series ต้องหาผลต่างต่อไป เช่น อนุกรมเวลารายเดือนที่มีทั้งแนวโน้มและฤดูกาลเมื่อ d = 1 และ D = 1 จะแปลงอนุกรมเวลาเดิม $\{Y_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{Z_t\}$ โดย $Z_t = \nabla\nabla_{12} Y_t = \nabla(Y_t - Y_{t-12}) = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$ เป็นต้น

2.4) การหาลอการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือการแปลงอนุกรมเวลาเดิม $\{Y_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{Z_t\}$ โดย $Z_t = \ln(Y_t)$ การแปลงนี้จะทำเมื่อความแปรปรวนของ อนุกรมเวลาไม่คงที่ นั่นคือ $V(Y_t)$ สำหรับค่าเวลา t ต่าง ๆ มีค่าไม่คงที่

การพยากรณ์อนุกรมเวลา สำหรับแบบจำลอง ARIMA (p,d,q) นั้นจะอ้างอิงทฤษฎีของ Box-Jenkins (1976) เป็นเครื่องมือในการศึกษาครั้งนี้ โดยในเบื้องต้นต้องพิจารณาว่าอนุกรมเวลานั้นมีคุณสมบัติของอนุกรมเวลาที่มีลักษณะนิ่ง (stationary series) หรือไม่ (Pindyck and Rubinfeld, 1998) โดยสามารถพิจารณาได้จากค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วมของข้อมูลอนุกรมเวลาดังต่อไปนี้

ค่าเฉลี่ย : $E(Y_t) = \mu$ (3.1)

ค่าความแปรปรวน : $V(Y_t) = \sigma^2$ (3.2)

ค่าความแปรปรวนร่วม : $Cov(Y_t,Y_{t+k}) = \sigma - \mu$ (3.3)

เพราะฉะนั้น ข้อมูลอนุกรมเวลาที่มีลักษณะนิ่งจะต้องมีค่าเฉลี่ยและค่าความแปรปรวนของทุก ๆ ค่า ณ เวลา t ใด ๆ คงที่ ในขณะที่ค่าความแปรปรวนร่วมระหว่างสองคาบเวลาจะขึ้นอยู่กับช่องว่างระหว่างสองคาบเวลาเท่านั้น ไม่ขึ้นอยู่กับเวลาที่เปลี่ยนแปลงไป หากเงื่อนไขใดเงื่อนไขหนึ่งไม่เป็นดังที่กล่าวมา แสดงว่าข้อมูลมีลักษณะไม่นิ่ง (non stationary series) (Charemza and Deadman, 1992)

ดังนั้น สามารถสรุปได้ว่า ข้อมูลอนุกรมเวลาที่มีลักษณะนิ่งจะต้องมีค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วมมีค่าคงที่ ณ ทุก ๆ ช่วงเวลาที่เปลี่ยนแปลงไป ซึ่งสามารถทดสอบข้อมูลว่ามีลักษณะนิ่งหรือไม่จากการทดสอบ unit root

3.1.2 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit Root Test)

แนวคิดการทดสอบ unit root (Gujarati, 2003) ได้อธิบายว่าสามารถทดสอบได้โดยใช้ Dickey and Fuller test: DF-test (Dickey and Fuller, 1981) และ Augmented Dickey and Fuller test: ADF-test (Said and Dickey, 1984)

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad (3.4)$$

จากสมการ (3.4) สามารถทดสอบ DF-test โดยมีสมมติฐานว่าง (null hypothesis) คือ $H_0: \rho = 1$ โดยที่ถ้า $|\rho| < 1$ ข้อมูลอนุกรมเวลา X_t จะมีลักษณะนิ่ง (stationary) และถ้า $\rho = 1$ ข้อมูลอนุกรมเวลา X_t จะมีลักษณะไม่นิ่ง (non stationary) นอกจากนี้ยังสามารถทำการทดสอบได้อีกทางหนึ่งดังสมการ (3.5)

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \qquad (3.5)$$

หรือ $X_t = (1+\theta) X_{t-1} + \varepsilon_t$ ซึ่งก็คือสมการที่ (3.4) นั่นเอง โดยที่ $\rho = (1+\theta)$ ถ้า θ ในสมการ (3.5) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3.1) จะมีค่าน้อยกว่า 1 ดังนั้น สามารถสรุปได้ว่า 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 เป็น stationary series และถ้าเราไม่สามารถปฏิเสธ H_0: $\theta = 0$ ได้ ก็จะหมายความว่า X_t เป็น non stationary series

ถ้า X_t มีแนวคิดเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) เราสามารถจะเขียนแบบจำลองได้ดังนี้

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

และถ้า X_t เป็น random walk with drift และมีแนวโน้มตามเวลาเชิงเส้น (linear time trend) เราสามารถจะเขียนจำลองแบบได้ดังนี้

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

พิจารณาสมการถดถอยทั้ง 3 รูปแบบ คือสมการ (3.5), (3.6), (3.7) จะพบว่าตัวพารามิเตอร์ที่อยู่ในความสนใจของทุกสมการคือ θ เมื่อทำการทดสอบ DF-test ซึ่งมี H_0: $\theta = 0$ และ H_a: $\theta < 0$ โดยการเปรียบเทียบค่าสถิติ DF-test ที่คำนวณได้กับค่าวิกฤติที่เหมาะสมจาก Dickey Fuller tables (Enders, 1995) หรือกับค่าวิกฤติ MacKinnon (Gujarati, 2003) ถ้าไม่สามารถปฏิเสธสมมติฐานว่าง H_0: $\theta = 0$ ได้ ก็จะหมายความว่า X_t เป็น non stationary series หรือมี unit root แต่ถ้าสามารถปฏิเสธสมมติฐานว่างได้ ก็จะหมายความว่า X_t เป็น stationary series หรือไม่มี unit root นั่นเอง

อย่างไรก็ตามค่าวิกฤต (critical values) จะไม่เปลี่ยนแปลงถ้าสมการ (3.5), (3.6) และ (3.7) ถูกแทนที่โดยกระบวนการถดถอยในตัวเอง (Autoregressive processes)

$$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \quad (3.8)$$

$$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \quad (3.9)$$

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \quad (3.10)$$

จำนวนของ lagged difference terms ที่จะนำเข้ามารวมในสมการนั้นควรจะมีมากพอที่จะทำให้พจน์ค่าความคลาดเคลื่อน (error terms) มีลักษณะเป็น serially independent และเมื่อนำเอาการทดสอบ DF-test มาใช้กับสมการ (3.8), (3.9) และ (3.10) จะเรียกว่าการทดสอบ Augmented Dickey and Fuller test: ADF-test ค่าสถิติทดสอบ ADF-test มีการแจกแจงเชิงเส้นกำกับ (Asymptotic distribution) เหมือนกับสถิติ DF-test ดังนั้นก็สามารถใช้ค่าวิกฤตแบบเดียวกัน (Gujarati, 2003)

กรณีของการหา lag length ที่เหมาะสมนั้น Enders (1995) ได้เสนอแนะว่าวิธีหนึ่งในการหา lag length ก็คือ เริ่มต้นด้วยการให้มี lag length ที่ยาวมากพอ และก็ลดขนาดของ lag length ลงโดยใช้ค่าสถิติทดสอบ t-test หรือค่าสถิติทดสอบ F-test สมมติว่าเราใช้ lag length เท่ากับ n ถ้า t-statistic ของ lag n นั้น ๆ ไม่มีนัยสำคัญ ณ ค่าวิกฤติ ก็จะต้องทำการประมาณค่าการถดถอยใหม่โดยใช้ lag length n-1 ทำอย่างนี้เรื่อยไปจนกระทั่ง lag นั้นมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญ

เมื่อพิจารณาการตรวจสอบแล้วว่า อนุกรมเวลาที่ศึกษาเป็น non stationary จะต้องแปลงให้เป็น stationary เสียก่อน ภายหลังจากนั้นก็ทำตามวิธี Box-Jenkins ต่อไป

3.2.3 การพยากรณ์ข้อมูลอนุกรมเวลา โดยวิธี Box – Jenkins

การพยากรณ์อนุกรมเวลาโดยวิธี Box – Jenkins ในรูปแบบ ARIMA (p,d,q) ต้องพิจารณาอนุกรมเวลา $\{Y_t\}$ มีคุณสมบัติอนุกรมเวลาที่เป็น stationary เสียก่อน การพิจารณาว่าอนุกรมเวลาเป็น stationary หรือไม่ (Dickey and Fuller, 1979) จะพิจารณาจาก

1) ค่าเฉลี่ย $E(Y_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมากจะสรุปได้ว่า $E(Y_t)$ คงที่

2) ค่าความแปรปรวน $V(Y_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าความแปรปรวนแต่ละส่วนย่อยไม่แตกต่างกันมากนักจะสรุปได้ว่า $V(Y_t)$ คงที่

3) พิจารณาแนวโน้ม และ/หรือ ปัจจัยฤดูกาล ด้วยการวาดกราฟอนุกรมเวลา ในกรณีที่มีแนวโน้มและ/หรือปัจจัยฤดูกาล มักจะเห็นชัดเจนได้จากกราฟ

4) พิจารณาจาก correlogram ของค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ stationary ค่า correlogram ของ autocorrelation (r_k) จะมีค่าลดลงค่อนข้างเร็วเมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นถ้าค่า autocorrelation (r_k) มีค่าลดลงค่อนข้างช้าจะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่า autocorrelation (r_k) มีค่าลดลงค่อนข้างช้าและมีค่าค่อนข้างสูงที่ k = L , 2L, 3L จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลฤดูกาล และถ้าการเคลื่อนไหวของค่า correlogram ของ autocorrelation (r_k) มีลักษณะคล้ายลูกคลื่นโดยคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วว่า อนุกรมเวลาที่ศึกษาเป็น non stationary ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่เป็น non stationary จะต้องแปลงอนุกรมเวลาให้เป็น stationary เสียก่อน โดยการหาผลต่างสำหรับอนุกรมที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างและผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลาเดิมโดยการหาลอการิทึม ($Z_t = \ln Y_t$) จนกว่าจะได้อนุกรมเวลาใหม่ที่มีความแปรปรวนคงที่

วิธีการวิเคราะห์อนุกรมเวลาของ Box-Jenkins เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา โดยใช้ค่า Autocorrelation Function: ACF และค่า Partial Autocorrelation Function: PACF เป็นหลักในการพิจารณา และรูปแบบที่เลือกใช้จะอยู่ในกลุ่มของรูปแบบ ARIMA (p,d,q) หรือเรียก Integrated Autoregressive - Moving Average order p and q ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากการสังเกตหรือค่าพยากรณ์ก่อนหน้าและความคลาดเคลื่อนของการพยากรณ์ก่อนหน้า โดยเป็นการรวมส่วนของรูปแบบ Autoregressive order p: AR(p) และรูปแบบ Moving Average order q: MA(q) เข้าด้วยกัน รูปแบบ AR(p) หมายถึง รูปแบบที่แสดงว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่า Y_{t-1}, Y_{t-2} .....,Y_{t-p} หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA(q) หมายถึง รูปแบบที่แสดงว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่าของความคลาดเคลื่อน ε_{t-1} , ε_{t-2} ,..., ε_{t-q} หรือค่าความคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่า ส่วน Integrated: I(d) เกิดจากการหาผลต่าง (differencing) ของอนุกรมเวลา ซึ่งรูปแบบ ARIMA มีการกำหนดรูปแบบสมการดังนี้

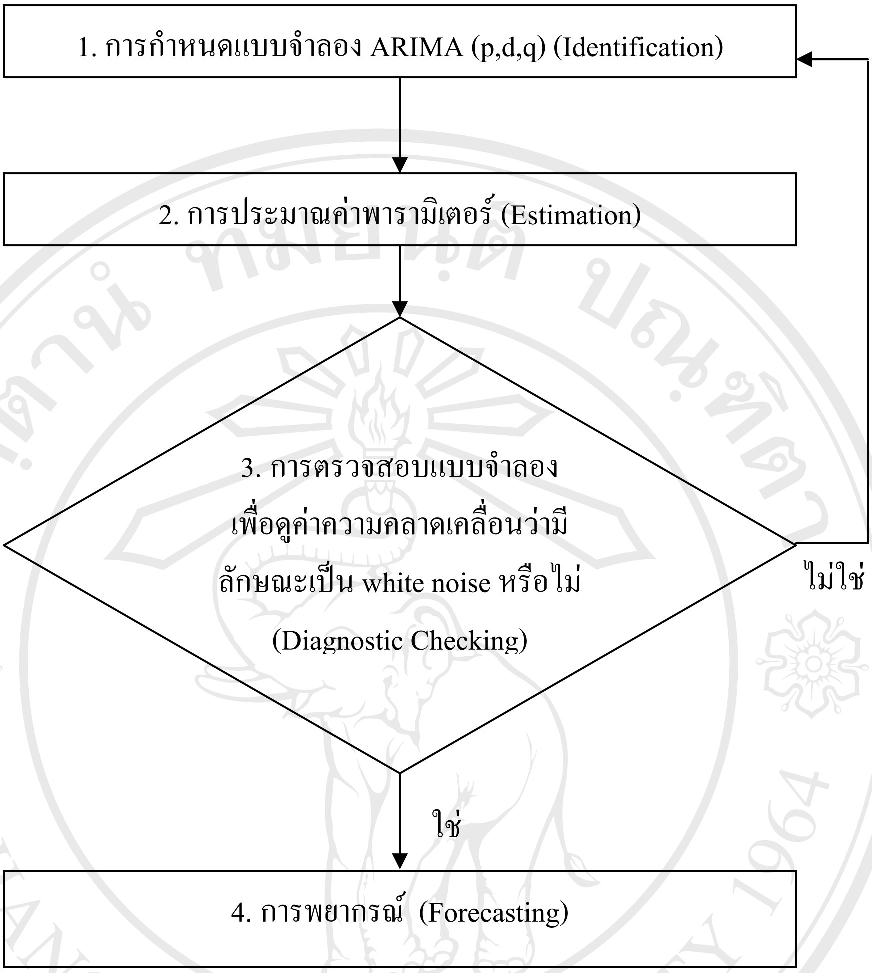
AR(p) $$Y_t = \theta_0 + \varnothing_1 Y_{t-1} + \ldots + \varnothing_p Y_{t-p} + \varepsilon_t \quad (3.11)$$

MA (q) $$Y_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \quad (3.12)$$

ARMA (p,q) $$Y_t = \theta_0 + \varnothing_1 Y_{t-1} + \ldots + \varnothing_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \quad (3.13)$$

ARIMA (p,d,q) $$\Delta^d Y_t = \theta_0 + \varnothing_1 \Delta^d Y_{t-1} + \ldots + \varnothing_p \Delta^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \quad (3.14)$$

การพยากรณ์โดยวิธีของ Box – Jenkins ประกอบด้วย 4 ขั้นตอนคือ การกำหนดแบบจำลอง (Identification) การประมาณค่าพารามิเตอร์ (Estimation) การตรวจสอบแบบจำลอง (Diagnostic Checking) และการพยากรณ์ (Forecasting) ตามลำดับ ดังจะพิจารณาจากรูปที่ 3.1



รูปที่ 3.1 ขั้นตอนการพยากรณ์โดยวิธี Box-Jenkins
ที่มา: Gujarati (2003)

1) การกำหนดแบบจำลอง (Identification) ให้กับอนุกรมเวลาที่เป็น stationary series เป็นการหารูปแบบ ARMA (p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลา โดยพิจารณาที่สหสัมพันธ์ (autocorrelation: ρ_k) โดยที่ ρ_k มีค่าอยู่ในช่วง [-1,1] คือการวัดความสัมพันธ์ของข้อมูลแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา สำหรับวิธีการหาค่า ρ_k สามารถหาได้จากความแปรปรวนร่วม (covariance) ของกลุ่มตัวอย่างที่ lag k (sample covariance) เทียบกับค่าความแปรปรวน (variance) ของกลุ่มตัวอย่าง (sample variance) ซึ่งพิจารณาได้จากสมการดังต่อไปนี้

$$\rho_k = \frac{\sum (y_t - \hat{y})(y_{t+k} - \hat{y})}{\sum (y_t - \bar{y})^2} \quad (3.15)$$

อย่างไรก็ตามเนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวแปรอิสระที่เป็นค่าความล่าช้า (lag) ของตัวแปรตาม (Autoregressive) และที่เป็นสหสัมพันธ์ของค่าความคลาดเคลื่อน (Moving Average) โดยทั้งนี้เนื่องจาก Autocorrelation Function: ACF จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อน แต่ไม่สามารถใช้อธิบายสหสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง Partial Autocorrelation Function: PACF จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule-Walker (Pindyck and Rubinfeld, 1998) ดังนี้

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

ถ้า k มากกว่า p จะได้

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

จากข้างต้นเมื่อทราบถึงคอเรลโลแกรมของ ACF และ PACF ก็นำมาหารูปแบบที่มีความเป็นไปได้ โดยสามารถพิจารณาความเป็นไปได้ของแบบจำลองจากตารางที่ 3.1 ดังต่อไปนี้

ตารางที่ 3.1 แสดงการพิจารณาคอเรลโลแกรมของ ACF และ PACF

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

ที่มา: Gujarati (2003)

จากตารางข้างต้น จะสามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมาให้นับเป็นค่าที่ p ของ AR(p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโรลแกรม เกิดขึ้น 1 แท่ง แปลได้ว่าแบบจำลองควรมีลักษณะเป็น AR(l) สำหรับ MA(q) นั้นก็จะมี ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไปใน

ขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สามารถสรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(2) และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่แบบจำลองควรจะเป็น ARMA(p,q) และเมื่อรวมกันกับการทดสอบความนิ่ง (stationary) แล้ว จะสามารถหาค่าผลต่าง (difference) ได้ ซึ่งผลจากการหาค่าของผลต่าง (differencing) จำนวน d ครั้งนั้น ก็จะได้แบบจำลอง ARIMA(p,d,q) แต่หากข้อมูลที่ทดสอบมีความนิ่งอยู่ก่อนแล้วนั้น แสดงว่าแบบจำลองควรจะเป็น ARMA(p,q) อย่างไรก็ตาม หลักการดังกล่าวก็เป็นเพียงเครื่องช่วยการพิจารณาในระดับหนึ่งเท่านั้น เนื่องจากวิธีดังกล่าวมีลักษณะเป็นศิลป์มากกว่าศาสตร์ ดังนั้นเพื่อประเมินแบบจำลองว่าแบบจำลองใดมีความเหมาะสมที่จะใช้เป็นตัวแทนกลุ่มข้อมูลจริง สามารถพิจารณาได้จากค่าสถิติดังต่อไปนี้เพื่อประกอบการตัดสินใจ

- ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (root mean square error: RMSE) โดยจะเป็นการวัดว่าค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองมีความแตกต่างกันมากน้อยเพียงใด ซึ่งหากค่า RMSE มีค่าเท่ากับศูนย์ จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงใดก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากเพียงนั้น (Pindyck and Rubinfeld, 1998) ซึ่งสามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

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

โดยที่ Y^s_t คือ ค่าที่ประมาณได้จากแบบจำลอง

Y^a_t คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

- ค่า Theil’s inequality coefficient: U หรือ TIC โดยในหลักการเบื้องต้น พบว่าสมการที่ใช้นี้ยังคงมีหลักการที่คล้ายกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือ ค่าสถิตินี้จะมีค่าอยู่ระหว่าง 0 และ 1 ทั้งนี้หากค่า U มีค่าเท่ากับศูนย์ จะหมายความว่าแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี แสดงถึงแบบจำลองที่ประมาณได้เป็นแบบจำลองที่เป็นตัวแทนข้อมูลจริงได้ดีที่สุด ในขณะที่ถ้า U มีค่าเท่ากับหนึ่ง แปลว่าแบบจำลองที่ประมาณได้เป็นแบบจำลองที่แย่ที่สุด (Pindyck and Rubinfeld, 1998) ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่าที่น้อย ๆ ดังจะพิจารณาได้จากสมการ

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^a)^2}} \qquad (3.19)$$

- Akaike information criterion: AIC คือ ค่าสถิติประยุกต์คล้ายกับ Adjusted R^2 แต่ใช้รูปแบบการใส่ค่าลอการิทึมฐานธรรมชาติ (natural logarithm) โดยหากค่าสถิตินี้มีค่าน้อยเพียงใด นั่นก็แปลว่า แบบจำลองที่ประมาณได้นั้นสามารถใช้เป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ ค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (lag length) ที่เหมาะสมอีกด้วย ดังแสดงในสมการต่อไปนี้

$$\ln AIC = (\frac{2k}{n}) + \ln(\frac{\sum \hat{u}_i^2}{n}) \qquad (3.20)$$

- Schwarz's Bayeian information criterion: SC, BIC หรือ SBC คือ วิธีการวัดปรับอย่างดี (goodess of fit) ของแบบจำลองได้ดีกว่า AIC เป็นวิธีที่ประยุกต์คล้ายกับ AIC สามารถเขียนสมการได้ดังนี้

$$SBC = \log\left[\frac{\sum \hat{u}_i^2}{n}\right] + \frac{2k \log n}{n} \qquad (3.21)$$

- R^2 คือ การวัดว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใด หากค่านี้เท่ากับ 1 ก็หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% ในทางกลับกันหากค่านี้มีค่าเท่ากับ 0 แปลความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่อย่างไรก็ตามพบว่า หากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมาก ๆ ก็จะทำให้ค่า R^2 มากขึ้นด้วย ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้ดังนี้

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \qquad (3.22)$$

โดยที่ $\sum \hat{u}_i^2$ คือผลรวมของค่าความคลาดเคลื่อน

- Adjusted R^2 มีการชั่งน้ำหนักกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการต่อไปนี้

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

โดยที่ n คือ ค่าสังเกตทั้งหมด

- Durbin-Watson statistic: DW- statistic ใช้ทดสอบ autocorrelation แสดงความมีสหสัมพันธ์ในตัวเองของค่าความคลาดเคลื่อน (error term) โดยค่า DW เข้าใกล้ 2 ถือว่าไม่มีสหสัมพันธ์ในตัวเอง แสดงว่าแบบจำลองที่ใช้อยู่ในระดับที่น่าเชื่อถือได้ เขียนสมการได้ดังนี้

$$DW = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} \tag{3.24}$$

จากค่าสถิติข้างต้นทั้งหมด จะนำมาใช้ประกอบในการพิจารณาเลือกแบบจำลอง ARIMA (p,d,q) ที่เหมาะสมที่สุด โดยจะคัดเลือกแบบจำลองในขั้นตอนนี้ไว้หลายแบบจำลอง เพื่อทำการเลือกอีกครั้งในขั้นตอนการพยากรณ์ เพื่อที่จะเปรียบเทียบว่าแบบจำลองใดจะมีความสามารถในการพยากรณ์ได้ดีที่สุด

2) การประมาณค่าพารามิเตอร์ในรูปแบบ (Estimation) จะทำได้โดยการหาค่าประมาณแบบง่าย หรือค่าประมาณที่ได้จากการวิเคราะห์ตัวเลข (numerical analysis) สำหรับค่าประมาณแบบง่าย จะทำโดยการสร้างสมการที่มาจากความสัมพันธ์ระหว่าง autocorrelation (ρ_k) และพารามิเตอร์ โดยสมการที่สร้างขึ้นจะมีจำนวนเท่ากับพารามิเตอร์ที่ต้องการประมาณ ส่วนค่าประมาณที่ได้จากการวิเคราะห์ตัวเลขจะได้จากการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด ขั้นตอนของการวิเคราะห์ตัวเลขจะต้องมีการกำหนดค่าประมาณเริ่มต้น ซึ่งส่วนใหญ่จะใช้การประมาณแบบง่ายเป็นค่าประมาณเริ่มต้น เมื่อการวิเคราะห์สิ้นสุดจะได้ค่าประมาณสุดท้ายที่จะนำไปใช้ประโยชน์ในการสร้างสมการพยากรณ์

3) การตรวจสอบแบบจำลอง (Diagnostic Checking) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลองแล้ว จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธีได้แก่ การพิจารณาคอเรลโลแกรมของ autocorrelation ของกลุ่มตัวอย่างหรือของค่าคลาดเคลื่อน การทดสอบค่าพารามิเตอร์ในแบบจำลองด้วยการทดสอบแบบ t และการทดสอบความเหมาะสมของแบบจำลองโดยการทดสอบของ Box - Ljung หรือการทดสอบของ Box – Pierce ซึ่งจะพิจารณาจาก Q–statistic (Gujarati, 2003) ดังสมการ

$$Q = n\sum_{k=1}^{m} \hat{\rho}_k^2 \tag{3.25}$$

โดยที่ n คือ จำนวนของข้อมูล

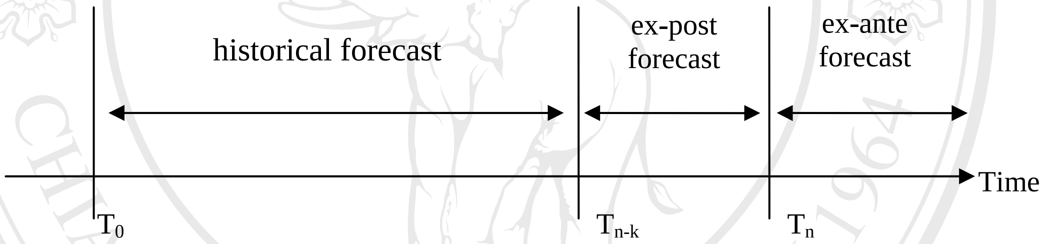
m คือ ค่า lag length

โดยมีการกำหนดค่า Q–statistic เพื่อเป็นการทดสอบว่าสหสัมพันธ์ในตัวเองของค่าความคลาดเคลื่อนจากการประมาณ (estimated residuals) ทุกช่วงเวลาที่ห่างกัน k มีความเป็นอิสระหรือไม่ จากสมมติฐานดังต่อไปนี้

$$H_0 : \rho_1\left(\hat{\varepsilon}_t\right) = \rho_2\left(\hat{\varepsilon}_t\right) = \ldots = \rho_k\left(\hat{\varepsilon}_t\right) = 0$$
$$H_a : \rho_1\left(\hat{\varepsilon}_t\right) \neq \rho_2\left(\hat{\varepsilon}_t\right) \neq \ldots \neq \rho_k\left(\hat{\varepsilon}_t\right) \neq 0$$

ทั้งนี้ค่า Q นั้นจะพบว่ามีการแจกแจงแบบ chi-square ที่มีดีกรีเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมติฐานว่า สมมติฐานว่างคือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น white noise นั่นแปลว่าแบบจำลองมีการกระจายแบบปกติ (Normal distribution) มีลักษณะปราศจากสหสัมพันธ์ (Autocorrelation) และไม่มีความแปรปรวนแตกต่างกัน (Hetaroscedasticity) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นมีลักษณะเป็น white noise แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบของแบบจำลองใหม่

4) การพยากรณ์ (Forecasting) ใช้สมการพยากรณ์ที่สร้างจากรูปแบบการพยากรณ์ที่กำหนดและผ่านการตรวจสอบในขั้นตอนที่ผ่านมาแล้ว แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องเป็นแบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลอง โดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง historical forecast อันเป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา ($T_0 – T_{n-k}$) การพยากรณ์ช่วง ex-post forecast คือการพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์เปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า root mean squared error: RMSE ค่า Theil's inequality coefficient: U และค่า Akaike information criterion: AIC จะพิจารณาค่าสถิติทั้ง 3 ค่าที่มีค่าต่ำที่สุด ซึ่งได้จากการทำการพยากรณ์เมื่อเลือกแบบจำลองที่ดีที่สุดได้แล้ว จึงนำแบบจำลองนั้นมาทำการพยากรณ์แบบ ex-ante forecast ซึ่งเป็นการพยากรณ์ข้อมูลไปข้างหน้า ดังรูป



รูปที่ 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา: Pindyck and Rubinfeld (1998)

3.2 ระเบียบวิธีวิจัย

ในการศึกษาครั้งนี้ จะทำการพยากรณ์มูลค่าการส่งออกรายเดือนของปลาทูน่ากระป๋อง ตั้งแต่เดือนมกราคม พ.ศ. 2534 ถึงเดือนพฤษภาคม พ.ศ. 2549 จำนวนทั้งหมด 185 เดือน ซึ่งจะใช้วิธีการพรรณนาอธิบาย และวิธีการวิเคราะห์เชิงปริมาณ โดยกำหนดรูปแบบจำลองให้กับอนุกรมเวลาในรูปแบบ ARIMA โดยวิธีของ Box – Jenkins เนื่องจากเป็นวิธีที่จะให้ค่าพยากรณ์ที่มีความถูกต้องสูงโดยเฉพาะในการพยากรณ์ระยะสั้น โดยมีขั้นตอนดังต่อไปนี้

1) การแปลงข้อมูลให้อยู่ในรูปลอการิทึมฐานธรรมชาติ (natural logarithm) เนื่องจากข้อมูลอนุกรมเวลามูลค่าการส่งออกปลาทูน่ากระป๋องมีลักษณะแปรปรวนมาก เพื่อให้ค่าความแปรปรวนของอนุกรมเวลาคงที่สำหรับค่า t ต่าง ๆ ดังนั้นจึงต้องแปลงข้อมูลอนุกรมเวลาเดิมให้เป็นอนุกรมเวลาใหม่ซึ่งอยู่ในรูปลอการิทึมฐานธรรมชาติ ทั้งนี้ในการศึกษาจะใช้ตัวแปร lnCTN แทนมูลค่าการส่งออกปลาทูน่ากระป๋องในรูปลอการิทึมฐานธรรมชาติ

2) การทดสอบความนิ่งของข้อมูล เป็นการพิจารณาว่าข้อมูลอนุกรมเวลามีลักษณะนิ่ง (stationary) หรือไม่ ด้วยการทดสอบ unit root โดยใช้ ADF-test ดังแสดงในสมการต่อไปนี้

$$\Delta \ln CTN_t = \theta \ln CTN_{t-1} + \sum_{i=1}^{4} \phi_i \Delta \ln CTN_{t-i} + \varepsilon_t \qquad (3.26)$$

$$\Delta \ln CTN_t = \alpha + \theta \ln CTN_{t-1} + \sum_{i=1}^{4} \phi_i \Delta \ln CTN_{t-i} + \varepsilon_t \qquad (3.27)$$

$$\Delta \ln CTN_t = \alpha + \beta t + \theta \ln CTN_{t-1} + \sum_{i=1}^{4} \phi_i \Delta \ln CTN_{t-i} + \varepsilon_t \qquad (3.28)$$

ทั้งนี้การหา lag length จะเริ่มที่ความยาว lag ที่ 4 แล้วลดขนาดลงเรื่อย ๆ หากค่า t-statistic ไม่มีนัยสำคัญทางสถิติ จนกระทั่งค่า t-statistic มีนัยสำคัญทางสถิติ แล้วจึงทำการเปรียบเทียบค่าสถิติ ADF-test ที่คำนวณได้กับค่าวิกฤติ MacKinnon (Gujarati, 2003) โดยมีสมมติฐานดังนี้

H_0: $\theta = 0$ คือ ข้อมูลมี unit root หรือมีลักษณะที่ไม่นิ่ง (non stationary) ซึ่งต้องทำการหาผลต่าง (differencing) ในอันดับต่อไป

H_a: $\theta < 0$ คือ ข้อมูลปราศจาก unit root หรือมีลักษณะที่นิ่ง (stationary) ณ อันดับนั้นที่ทำการทดสอบ

เมื่อพบว่าข้อมูลมีลักษณะนิ่งแล้ว ก็จะสามารถนำข้อมูลอนุกรมเวลานั้นมาเพื่อหารูปแบบ ARIMA(p,d,q) ด้วยวิธีของ Box – Jenkins ต่อไป

3) การกำหนดรูปแบบของแบบจำลอง คือ การระบุว่ารูปแบบจำลองควรจะมีค่า Autoregressive, p เท่าใด Differencing, d ที่ลำดับเท่าใด และ Moving Average, q เท่าใด โดยพิจารณาจาก correlogram ของ Autocorrelation Function: ACF และ Partial Autocorrelation Function: PACF โดยสร้างแบบจำลองไว้หลาย ๆ แบบ เพื่อใช้เปรียบเทียบและเลือกแบบจำลองที่ดีและเหมาะสมที่สุดในการพยากรณ์ ซึ่งมีการกำหนดรูปแบบของ ARIMA(p,d,q) ไว้ดังนี้

$$\Delta^{d}\ln CTN_{t} = \theta_{0} + \varnothing_{1}\Delta^{d}\ln CTN_{t-1} + \ldots + \varnothing_{p}\Delta^{d}\ln CTN_{t-p} + \varepsilon_{t} - \theta_{1}\varepsilon_{t-1} - \ldots - \theta_{q}\varepsilon_{t-q} \quad (3.29)$$

ในการศึกษาครั้งนี้จะพิจารณารูปแบบ ARIMA(p,d,q) โดยใช้ค่าสถิติเพื่อประกอบการตัดสินใจดังนี้ Akaike information criterion: AIC, Schwarz's Bayeian information criterion: SBC, Adjusted R^2, Durbin-Wason statistic: DW, และ F-statistic

4) การประมาณค่าพารามิเตอร์ คือ การนำเอารูปแบบ ARIMA(p,d,q) ที่เลือกจากขั้นตอนการหารูปแบบที่เหมาะสมมาประมาณค่าพารามิเตอร์ โดยพิจารณาความมีนัยสำคัญทางสถิติด้วย t-statistic

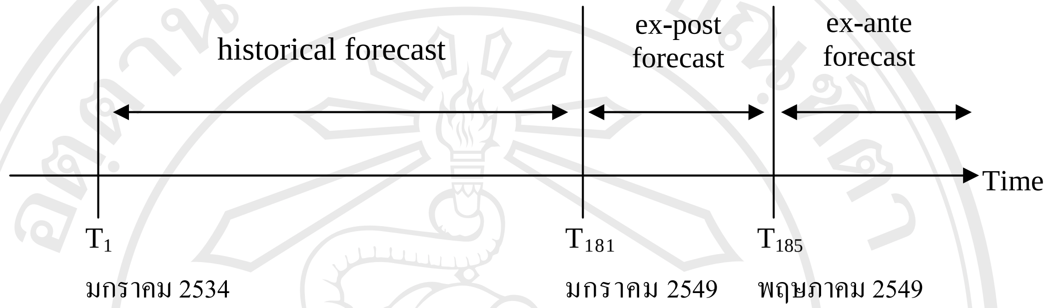
5) การตรวจสอบความถูกต้อง คือ การตรวจสอบสหสัมพันธ์ (Autocorrelation) จากค่าความคลาดเคลื่อนที่ประมาณได้(estimated residual: ε_t) ว่ามีลักษณะเป็น white noise หรือไม่ โดยพิจารณาจากค่า Q – statistic

6) การพยากรณ์ คือ การพยากรณ์ข้อมูลโดยใช้สมการที่สร้างจากรูปแบบจำลองและผ่านการตรวจสอบตามขั้นตอนแล้ว มาพยากรณ์ผลที่คาดว่าจะเกิดขึ้นในช่วงเวลาต่อ ๆ ไป โดยในการศึกษาครั้งนี้จะแบ่งช่วงการพยากรณ์เป็น 3 ช่วง ดังนี้

- Historical forecast เป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา คือ ตั้งแต่ช่วงเวลาที่ 1 ถึง 181 โดยการถดถอยข้อมูลใหม่และพยากรณ์ พิจารณาค่าสถิติ root mean square error: RMSE และ Theil's inequality coefficient: U ที่ต่ำที่สุด

- Ex-post forecast คือ การกำหนดการพยากรณ์ในช่วงเวลาสั้น ๆ เพื่อเปรียบเทียบว่าแบบจำลองใดจะมีความสามารถในการพยากรณ์ดีที่สุด โดยการลดจำนวนข้อมูลลง 4 ค่า จาก 185 ค่าสังเกต เหลือ 181 ค่าสังเกต แล้วทำการถดถอยข้อมูลใหม่และพยากรณ์ 4 คาบเวลาถัดไป คือพยากรณ์ช่วงเวลาที่ 182 ถึง 185 เพื่อเปรียบเทียบกับค่าจริงของข้อมูลที่มีอยู่ โดยพิจารณาค่าสถิติ RMSE และ U ที่ต่ำที่สุด

- Ex-ante forecast เมื่อทราบรูปแบบจำลองที่สามารถพยากรณ์ได้ดีที่สุดจากช่วง Ex-post forecast แล้ว จึงนำแบบจำลองนั้นไปพยากรณ์ช่วงเวลาถัดไปอีก 4 คาบเวลา กล่าวคือ การพยากรณ์ ณ ช่วงเวลาที่ 186 ถึง 189



รูปที่ 3.3 แสดงช่วงเวลาที่ใช้ในการพยากรณ์จริง