บทที่ 3

กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการศึกษาพฤติกรรมการเคลื่อนไหวอนุกรมเวลารายเดือนของยอดขายสินค้าที่ศึกษา ใช้การเคลื่อนไหวของอนุกรมเวลาวิธีพรรณนาอธิบายพฤติกรรมและใช้วิธีวิเคราะห์เชิงปริมาณ โดยทำการกำหนดแบบจำลองอารีมาตามวิธีของ Box-Jenkins พร้อมทั้งมีการอธิบายปัจจัยที่กำหนดการเคลื่อนไหวและเมื่อพยากรณ์อนุกรมเวลาของสินค้า

3.1.1 การวิเคราะห์อนุกรมเวลา (Time series analysis)

การวิเคราะห์อนุกรมเวลา (time series analysis) การวิเคราะห์นี้ คือการแยกความเคลื่อนไหวต่าง ๆ ออกจากข้อมูลอนุกรมเวลา การวิเคราะห์การเปลี่ยนแปลงเคลื่อนไหวของอนุกรมเวลาต่าง ๆ พบว่าการเปลี่ยนแปลงเคลื่อนไหวในอนุกรมเวลามักจะประกอบไปด้วย 4 ความเคลื่อนไหว (วีนัส ฤาชัย, 2547)

1) ค่าแนวโน้ม (trend : T)

ค่าแนวโน้มเป็นการเคลื่อนไหวในระยะเวลาที่ค่อนข้างจะยาวนาน ค่าแนวโน้มปกติแสดงถึงทิศทางอนุกรมเวลาชุดนั้น ๆ มุ่งไปสู่ค่าแนวโน้มที่อาจมีลักษณะเป็นเส้นตรง เส้นโค้ง หรือลักษณะอื่นใดก็ได้

2) การเคลื่อนไหวตามฤดูกาล (seasonal movement : S)

การเคลื่อนไหวประเภทนี้ เคลื่อนไหวขึ้น ๆ ลง ๆ ซึ่งในเวลาเดียวกัน กล่าวคือ เคยสูงเคยต่ำในระยะเวลาใดก็มักจะสูงต่ำในระยะเวลานั้นต่อไป อิทธิพลของฤดูกาลนี้ โดยปกติจะเกิดขึ้นซ้ำ ๆ ในลักษณะคล้ายกันทุกปี การเคลื่อนไหวนี้มักจะแสดงในลักษณะสัมพันธ์ คือเป็นจำนวนเปอร์เซ็นต์ หรือเรียกว่าดัชนีฤดูกาล (seasonal index)

3) การเคลื่อนไหวตามวัฏจักร (cyclical movement : C)

เป็นการเคลื่อนไหวแบบขึ้นๆ ลงๆ ในระยะเวลานานกว่า 1 ปี การเปลี่ยนแปลงตามวัฏจักร โดยทั่วไปแบ่งออกเป็น 4 ระยะคือ

ระยะที่ 1 เป็นระยะฟื้นตัวหรือขยายตัว

ระยะที่ 2 เป็นระยะที่รุ่งเรือง

ระยะที่ 3 เป็นระยะหดตัว

ระยะที่ 4 เป็นระยะตกต่ำและในเวลาเดียวกันก็จะฟื้นตัวในระยะขยายตัว

4) การเคลื่อนไหวผิดปกติ (Irregular movement : I)

เป็นการเคลื่อนไหวเพิ่มขึ้นหรือลดลงโดยผิดปกติ เคลื่อนไหวอย่างมีลักษณะไม่แน่นอน โดยมีสาเหตุที่ทำให้เกิดขึ้น โดยไม่มีใครคาดหมาย หรือไม่อาจคาดการณ์ได้ล่วงหน้า เช่น ไฟไหม้ น้ำท่วม การนัดหยุดงาน สงครามโลก และการก่อวินาศกรรม เป็นต้น

3.1.2 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลา

การศึกษาอนุกรมเวลารายเดือนของสินค้า จะทำการกำหนดรูปแบบให้กับอนุกรมเวลา เพื่อพิจารณาความมีเสถียรภาพและลักษณะรูปแบบพฤติกรรมยอดขายของสินค้า ว่ามีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้องหรือไม่ พร้อมทั้งนำรูปแบบของสมการที่ได้ทำการวิเคราะห์อนุกรมเวลาแล้ว นำมาทำนายยอดขายสินค้าที่ศึกษาไปล่วงหน้า เพื่อนำค่าพยากรณ์ดังกล่าวมาใช้ประโยชน์ในการจัดสรรทรัพยากรที่เหมาะสมต่อไป

วิธีการวิเคราะห์อนุกรมเวลาของ Box–Jenkins เป็นวิธีการพยากรณ์ที่มีความถูกต้องและเหมาะสมกว่าวิธีอื่นในการพยากรณ์ระยะสั้นในช่วงเวลาตั้งแต่ 1 - 3 เดือน หากต้องการพยากรณ์ในช่วงเวลาที่ยาวนานกว่านี้ ควรนำข้อมูลที่ทันสมัยมาปรับค่าพยากรณ์ที่ได้ทำไว้แล้วเพื่อลดค่าความคลาดเคลื่อนจากการพยากรณ์ให้น้อยลง

วิธีการของ Box–Jenkins เป็นการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลาโดยการใช้ค่า Autocorrelation Function (ACF) และค่า Partial Autocorrelation Function (PACF) เป็นหลักในการพิจารณารูปแบบที่ใช้เลือกใช้จะอยู่ในกลุ่มของรูปแบบ Integrated Autoregressive-Moving Average order p and q อารีมา (p,d,q) ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากการสังเกตหรือการพยากรณ์ล่วงหน้าและความคลาดเคลื่อนของการพยากรณ์ล่วงหน้า โดยเป็นการรวมส่วนของรูปแบบ AR(p) และ MA(q) เข้าด้วยกัน โดยที่รูปแบบ AR(p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่า X_{t-1} , X_{t-2} , X_{t-3} ,......, X_{t-p} หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA(q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่าคลาดเคลื่อน e_{t-1} , e_{t-2} , e_{t-3} ,......, e_{t-q} หรือค่าความคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่า ซึ่งในที่นี้ ค่าสังเกต X_t ซึ่งรูปแบบ ARIMA (p,q) มีการกำหนดรูปแบบดังนี้

AR(p) คือ $X_t = \theta_0 + \phi_t X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t$ (3.1)

MA(q) คือ $X_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (3.2)

ARMA(p,q) คือ $X_t = \theta_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (3.3)

ARIMA(p,q) คือ $\Delta^d X_t = \theta_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (3.4)

ในกรณีของยอดขายประกันอุบัติเหตุหมู่ จะใช้การแทนด้วยสัญลักษณ์ $Sale_t$ ในลักษณะเช่นเดียวกันกับสมการข้างต้น อนุกรมเวลาที่จะนำมาศึกษาเพื่อใช้ในการพยากรณ์นั้น การเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่าง ๆ ได้แก่ แนวโน้ม (trend) ตัวแปรฤดูกาล (seasonal factor) ตัวแปรวัฏจักร (irregular movement) โดยวิธี Box-Jenkins จะสามารถแบ่งอนุกรมเวลาออกเป็น 2 ประเภท ดังนี้

1) อนุกรมเวลาที่เป็น stationary series

คืออนุกรมเวลา X_t ที่มีค่าเฉลี่ยและความแปรปรวนของ X_t คงที่ นั่นคือค่าเฉลี่ย $E(X_t)$ และค่าความแปรปรวน $V(X_t)$ มีค่าคงที่สำหรับแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้ม และ/หรือ อิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(X_t)$ ไม่คงที่ และอนุกรมเวลาที่มีความแปรปรวนของ X_t สูง จะเป็นลักษณะของอนุกรมเวลาที่ $V(X_t)$ มีค่าไม่คงที่ ซึ่งจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่เป็น stationary series นอกจากนี้อนุกรมเวลาที่เป็น stationary series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้ว ยังต้องมีค่าสัมประสิทธิ์สหสัมพันธ์แบบออโต (Autocorrelation) ที่ Lag k ขึ้นอยู่กับค่า k อย่างเดียว อนุกรมเวลาที่กำหนดรูปแบบ ARMA(p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น stationary series แล้ว

2) อนุกรมเวลาที่ไม่เป็น stationary series

เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น stationary series (ข้อ 1) ซึ่งการหารูปแบบ ARMA(p,q) ให้กับอนุกรมเวลาดังกล่าวได้ จะต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ stationary series เสียก่อน การแปลงอนุกรมเวลาที่ไม่เป็น stationary series ให้เป็นอนุกรมเวลาที่เป็น stationary series อาจทำได้ด้วยวิธีการต่าง ๆ ดังนี้

2.1) การหาผลต่างปกติ (regular differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้ม นั่นคือถ้าอนุกรมเวลา X_t ที่มีแนวโน้มอยู่ในอนุกรมเวลา จะต้องแปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม (Z_t) โดย $Z_t = \Delta^d X_t$ โดยที่ d เป็นลำดับของการหาผลต่างและ Δ คือผลต่างของตัวแปร เช่น เมื่อ d = 1 จะได้ $Z_t = \Delta X_t = X_t - X_{t-1}$ เมื่อ d = 2 จะได้ $Z_t = \Delta^2 X_t = \Delta(X_t - X_{t-1}) = \Delta X_t - \Delta X_{t-1} = X_t - 2\Delta X_{t-1} + X_{t-2}$ เป็นต้น จำนวนครั้งที่หาผลต่างขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่

เป็น stationary series หรือไม่ ถ้ายังไม่เป็น stationary series ต้องหาผลต่างต่อไป โดยทั่วไปถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ $d = 1$ อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติก (quadratic) จะใช้ $d = 2$

2.2) การหาผลต่างฤดูกาลของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรเข้ามาเกี่ยวข้อง จะต้องแปลงอนุกรมเวลาเดิม (X_t) ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล (Z_t) โดย $Z_t = \Delta_L^D X_t$ โดย D เป็นลำดับของการหาผลต่างฤดูกาล และ L เป็นจำนวนฤดูกาลต่อไป เช่นสำหรับอนุกรมเวลารายเดือน (L=12) เมื่อ D=1 จะได้ $Z_t = \Delta_{12} X_t$ หรือ $Z_t = X_t - X_{t-12}$ และเมื่อ D=2 จะได้ $Z = \Delta_{12}^2 X_t$ หรือ $Z = \Delta^2 (X_t - X_{t-12}) = X_t - 2X_{t-12} + X_{t-24}$ เป็นต้น ผลต่างนี้ต้องทำกี่ครั้งขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น stationary series หรือไม่ ถ้ายังไม่เป็น stationary series ต้องหาผลต่างต่อไป

2.3) การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น stationary series นั้นทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาลควบคู่กันไป โดยค่า d เป็นลำดับของการหาผลต่างปกติ และ ค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และ D จะมีค่าเท่าใดนั้นขึ้นอยู่กับการหาผลต่างปกติและผลต่างฤดูกาลจนกว่าอนุกรมเวลาใหม่จะเป็น stationary series เช่น อนุกรมเวลารายเดือนที่มีทั้งแนวโน้มและฤดูกาล เมื่อ $d = 1$ และ $D = 1$ จะแปลงอนุกรมเวลาเดิม (X_t) ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล (Z_t) โดย $Z_t = \Delta\Delta_{12}X_t = \Delta X_t - X_{t-12} = X_t - X_{t-1} - X_{t-12} + X_{t-13}$ เป็นต้น

2.4) การหาลอการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือการแปลงอนุกรมเวลาเดิม (X_t) ให้เป็นอนุกรมเวลาใหม่ (Z_t) โดย $Z_t = \mathrm{In}(X_t)$ การแปลงอนุกรมเวลาลักษณะนี้จะทำเมื่อความแปรปรวน $V(X_t)$ ของอนุกรมเวลาไม่คงที่

3.1.3 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit root test)

แบบจำลองที่ใช้ในการศึกษาของข้อมูลอนุกรมเวลา สำหรับทดสอบว่าตัวแปรแต่ละตัวมีลักษณะนิ่ง (stationary) หรือไม่ คือการทดสอบ unit root ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์ (2542) กล่าวว่า การประมาณค่าทางเศรษฐมิติ โดยใช้ข้อมูลอนุกรมเวลานั้น มีข้อสมมุติเกี่ยวกับความนิ่งของข้อมูล สมมุติว่าแบบจำลองเป็นดังนี้

$$Y_t = \alpha + \beta X_t + u_t \qquad (3.5)$$

$$X_t = X_{t-1} + u_{21} \quad ; u_{2t} \sim \mathrm{iid}(0, \sigma_{u_2}^2) \qquad (3.6)$$

โดยที่ u_{21} เป็นอนุกรมเวลาของตัวแปรสุ่ม (random variables) ที่มีการแจกแจงปกติที่เหมือนกันและเป็นอิสระต่อกัน มีค่าเฉลี่ย (mean) เท่ากับศูนย์ และค่าความแปรปรวน (variance) คงที่ ซึ่งตัวแปร X เป็นแนวเดินเชิงสุ่ม และเป็น integrated of order one, I(1) เพราะฉะนั้นตัวแปร Y_t จะเป็น I(1) ด้วย ตามทฤษฎีเศรษฐมิตินั้น การถดถอยด้วยตัวแปรที่มีลักษณะไม่นิ่ง (nonstationary) ค่าสถิติ t (t-statistics) ที่ใช้โดยปกติจะมีการแจกแจงไม่มาตรฐาน (nonstandard distributions) ดังนั้น การใช้ตารางมาตรฐานสำหรับการทดสอบค่าสถิติแบบที่ใช้โดยทั่วไป อาจทำให้เกิดการสรุปที่ผิดพลาดและเกิดการถดถอยที่ไม่ถูกต้อง (spurious regressions) (Johnston and Dinardo, 1997)

การทดสอบ Unit Root นั้นทำให้หลายวิธี ได้แก่การทดสอบ ADF (Augmented Dickey–Fuller Test (ADF) test) (Said and Dickey, 1984) และ DF (Dickey–Fuller (DF) test) (Dickey and Fuller, 1981) โดยสมมุติฐานว่าง (null hypothesis) ของการทดสอบ DF (Dickey–Fuller (DF) test) คือ $H_0 : \rho = 1$ จากสมการ (3.7)

$$X_t \quad = \quad \rho X_{t-1} + \varepsilon_t \tag{3.7}$$

ถ้าการทดสอบพบว่า $|\rho| < 1$ แล้ว แสดงว่า X_t จะมีลักษณะนิ่ง แต่ถ้า $\rho = 1$ แล้ว X_t จะมีลักษณะไม่นิ่ง นอกจากนี้ยังสามารถทดสอบได้อีกทางหนึ่งซึ่งคล้ายกับสมการ (3.7)

$$\Delta X_t \quad = \quad \phi X_{t-1} + \varepsilon_t \tag{3.8}$$

สมมุติฐานว่าง (null hypothesis) ของการทดสอบ คือ $H_0 : \rho = 1$ และ $H_a : \theta < 0$ โดยที่ $\rho = (1+\theta)$ ซึ่งถ้า θ ในสมการ (3.8) มีค่าเป็นลบ จะได้ว่า $\rho < 1$ (ρ จากสมการ (3.7)) จึงสรุปได้ดังนี้ ถ้าปฏิเสธ $H_0 : \theta = 0$ หรือยอมรับ $H_a : \theta < 0$ หมายถึง X_t มีลักษณะนิ่ง หรือมี integration of order zero (Charemza and Deadman, 1992) ในทางตรงกันข้าม ถ้ายอมรับ H_0 จะหมายความว่า X_t มีลักษณะไม่นิ่ง (nonstationary) โดยการพิจารณาค่าสถิติ DF เปรียบเทียบกับค่าวิกฤต MacKinnon (McKinnon Critical Values) (Gujarati, 2003) หรือค่าวิกฤตจากตาราง Dickey–Fuller (Enders, 1995) โดยถ้า X มีแนวคิดเดินเชิงสุ่ม มีความโน้มเอียงทั่วไปรวมอยู่ด้วย สามารถเขียนแบบจำลองได้ดังสมการที่ (3.9) และถ้ามีแนวโน้มตามเวลาเชิงเส้นรวมอยู่ด้วย จะได้แบบจำลองดังสมการที่ (3.10) ตามลำดับ

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

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

โดยที่ t = เวลา

ถ้าหากสมการที่ (3.8) , (3.9) , (3.10) มีกระบวนการเชิงอัตถดถอยเข้ามาเกี่ยวข้องแล้ว จะเรียกว่า การทดสอบโดยวิธี Augmented Dickey Fuller test (ADF test) ได้สมการเป็นดังนี้

$$\Delta x_t = \theta x_{t-1} + \sum_{i=1}^{\rho} \varphi_i \Delta x_{t-1} + \varepsilon_t \qquad (3.11)$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \sum_{i=1}^{\rho} \varphi_i \Delta x_{t-1} + \varepsilon_t \qquad (3.12)$$

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{\rho} \varphi_i \Delta x_{t-1} + \varepsilon_t \qquad (3.13)$$

ซึ่งการทดสอบแบบดังกล่าวมีการแจกแจงเหมือนกับการทดสอบ DF จึงสามารถใช้ค่าวิกฤตในการพิจารณาแบบเดียวกัน และใช้ F-test ในการทดสอบว่าแบบจำลองที่เหมาะสมนั้น มีจุดตัดแกนและแนวโน้มเวลาหรือไม่ โดยคำนวณได้ดังนี้

$$F = \frac{(N-k)(SSR_R - SSR_{UR})}{r(SSR_{UR})} \qquad (3.14)$$

โดยที่ SSR_R = The sum of square of residual from the restricted model

SSR_{UR} = The sum of square of residual from the unrestricted model

N = Number of observation

k = Number of parameters estimated in the unrestricted model

r = Number of restrictions

และในการเลือก Lag Length จะใช้วิธี serial correlation LM test เพื่อหา Lag Length ที่มีค่า probability มากที่สุด จึงจะถือว่าค่า lag นั้นมีความเหมาะสม

3.1.4 แบบจำลองการพยากรณ์ โดยวิธี Box-Jenkins

การพยากรณ์อนุกรมเวลาโดยวิธีของ Box–Jenkins ในรูปแบบ ARIMA (p,d,q) จะต้องพิจารณาอนุกรมเวลาเป็น stationary Series หรือไม่ โดยพิจารณาจาก

1) ค่าเฉลี่ย $E(X_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วน ๆ แล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมากนัก สรุปได้ว่า $E(X_t)$ คงที่

2) ค่าความแปรปรวน $V(X_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ ทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วน ๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าความแปรปรวนแต่ละส่วนย่อยไม่แตกต่างกันมากนัก สรุปได้ว่า $V(X_t)$ คงที่

3) การพิจารณาแนวโน้ม และ/หรือปัจจัยฤดูกาล โดยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและ/หรือปัจจัยฤดูกาล จะเห็นได้ชัดเจนจากรูปที่เรียกว่าคอเรลโลแกรม (Correlogram)

4) พิจารณาจากคอเรลโลแกรม (Correlogram) ของสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ stationary ค่าคอเรลโลแกรมของสัมประสิทธิ์สหสัมพันธ์ (r_k) จะมีค่าลดลงค่อนข้างเร็วเมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นถ้าค่าสัมประสิทธิ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างจะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้าและมีค่าค่อนข้างสูง ที่ k = L, 2L, 3L จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลฤดูกาล และถ้าการเคลื่อนไหวของค่าคอเรลโลแกรมของสัมประสิทธิ์สหสัมพันธ์ (r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วว่า อนุกรมเวลาที่ศึกษาไม่เป็น stationary ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่เป็น stationary นั้น จะต้องแปลงอนุกรมเวลาให้เป็น stationary เสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างและผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลาเดิมโดยการหาลอการิทึม $Z_t = \ln(X_t)$ จนกว่าจะได้อนุกรมเวลาใหม่ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่ที่เป็น stationary series แล้วจะทำตามขั้นตอนของ Box–Jenkins

ขั้นตอนของ Box–Jenkins ประกอบด้วย 4 ขั้นตอน ดังนี้ ขั้นตอนที่หนึ่ง คือการกำหนดแบบจำลอง (identification) ขั้นตอนที่สอง คือ การประมาณค่าพารามิเตอร์ (parameter estimation) ขั้นตอนที่สาม คือ การตรวจสอบความถูกต้อง (diagnostic checking) และขั้นตอนที่สี่ คือ การพยากรณ์ (Forecasting)

1) การกำหนดแบบจำลอง (Identification)

คือการกำหนดรูปแบบให้กับอนุกรมเวลาที่เป็น stationary series กล่าวคือเป็นการหารูปแบบ ARMA(p,q) ที่คาดว่าเหมาะสมให้กับอนุกรมเวลาโดยที่ Autocorrelation : ρ_k คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลัง k หน่วยเวลาโดยที่ ρ_k มีค่าเท่ากับ

$-1 \leq \rho_k \leq 1$ โดยพิจารณาเปรียบเทียบค่า autocorrelation(r_k) ของอนุกรมเวลาตัวอย่าง กับค่า Autocorrelation (ρ_k) ของอนุกรมเวลาประชากรที่มีช่วงเวลาย้อนหลัง k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$\rho_k = \frac{\sum_{t=a}^{n-k}(X_{t-q})(X_{t+k+q})}{\sum_{t=a}^{n}(X_{t-q})^2} \tag{3.15}$$

โดยที่ $X_t = \sum_{t=a}^{n}(X_t)$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตาม เนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวแปรอิสระที่เป็นค่าความล่าช้า (lag) ของตัวแปรตาม (autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (moving average) เนื่องจาก Autocorrelation Function (ACF) ใช้ในการอธิบายสัน สัมพันธ์ของค่าความคลาดเคลื่อน แต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง Partial Autocorrelation Function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule–Walker (Pindyck and Rubinfeld, 1998) ดังนี้

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

ถ้า k > p จะได้

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

การกำหนดลำดับขั้น p , q ในแบบจำลอง (identifying 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

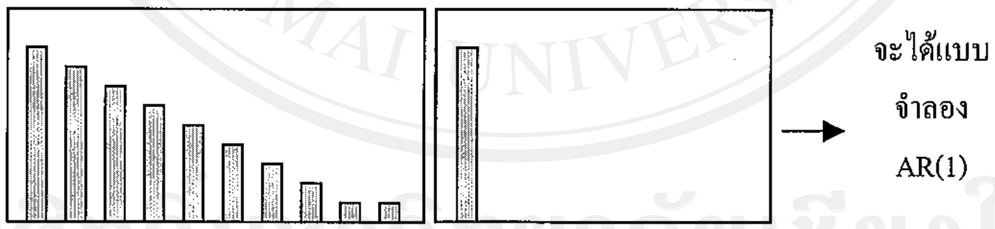
ตารางที่ 3.1 ตารางแสดงการพิจารณา ACF และ PACF

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 ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วหายไป จำนวนของแท่งของค่าที่เกิดขึ้นให้นับเป็นค่าที่ p ของ AR(p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่ลู่โค้งเข้าหาแกนระนาบ และ PACF มีแท่งคอเรลโลแกรมเกิด ขึ้น 1 แท่ง นั่นคือ แบบจำลองควรมีลักษณะเป็น AR(1) ดังรูป 3.1

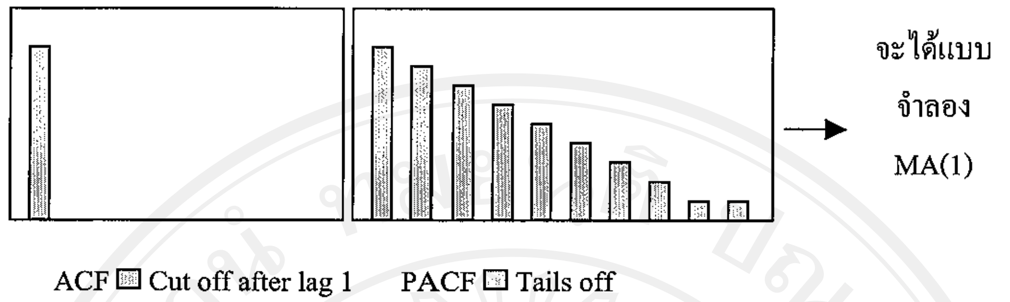
รูปที่ 3.1 ตัวอย่างการเกิดแบบจำลอง AR(p)



ACF ▤ Tails off PACF ▤ Cut off after lag 1

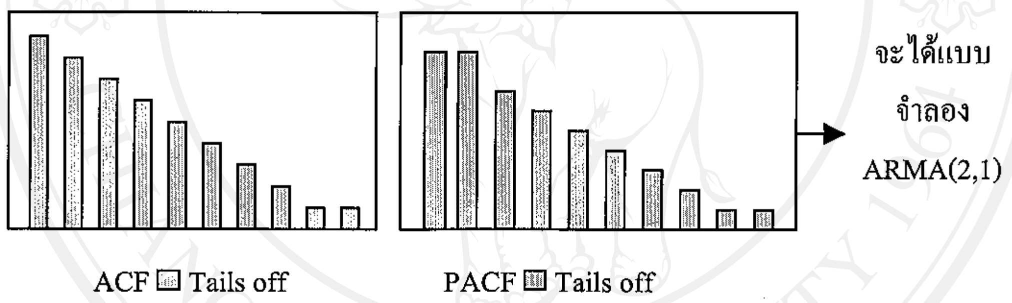
สำหรับ MA(q) นั้นก็จะมี ACF เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF ลู่โค้งเข้าหาแกนระนาบนั้น ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 1 แท่ง แล้วหายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(1) ดังแสดงตามรูป 3.2

รูปที่ 3.2 ตัวอย่างการเกิดแบบจำลอง MA(q)



และหาก ACF และ PACF ลู่โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรเป็น ARMA(p,q) ตัวอย่างการเกิดแบบจำลอง ARMA(p.q) ดังรูป 3.3

รูปที่ 3.3 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 มีค่าต่ำแสดงว่าสมการที่ได้จากแบบจำลองมีความคลาดเคลื่อนน้อย ซึ่งเป็นการประมาณสมการที่ดีและสามารถจะอธิบายได้ว่าค่าที่พยากรณ์ได้มีความโน้มเอียงต่ำ ดังนั้นหากว่าค่า RMSE มีค่าต่ำมากเท่าใดก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าที่แท้จริงได้ดีเท่านั้น สามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ได้ดังนี้

$$\text{RMSE} = \sqrt{(1/T)\sum_{t=1}^{T}(X_t^s - X_t^a)} \tag{3.18}$$

กำหนดให้ X_t^s = ค่าที่ประมาณจากแบบจำลอง

X_t^a = ค่าข้อมูลจริง

T = จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

ข) Theil's Inequality Coefficient (TIC : U)

เป็นค่าสถิติที่สามารถบอกถึงมาตรวัดสัมพัทธ์ ของค่าความคลาดเคลื่อนกำลังสอง (Mean Square Error : MSE) ว่าเกิดมาจากส่วนใด การพิจารณา Theil's Inequality Coefficient (TIC) จะทำให้รู้ถึงสาเหตุที่มีผลต่อค่า MSE ของตัวแบบการพยากรณ์ (วิชิต หล่อจีระชุณห์กุล และคณะ, 2539) โดยในหลักการเบื้องต้นพบว่าสมการ Theil's Inequality Coefficient (TIC) ที่ใช้นี้ยังคงมีหลักการที่คล้ายกันกับ RMSE แต่มีสิ่งที่แตกต่างออกไปก็คือ ค่า Theil's Inequality Coefficient (TIC) จะมีค่าอยู่ระหว่าง 0 ถึง 1 ทั้งนี้หากค่า TIC ที่ได้มีค่าเท่ากับศูนย์ นั่นหมายความว่าค่าที่ประมาณมีค่าเท่ากันกับข้อมูลจริงพอดี แสดงว่าแบบจำลองที่ประมาณขึ้นมาเป็นแบบจำลองที่จะเป็นตัวแทนของข้อมูลที่แท้จริงได้อย่างสมบูรณ์ หากค่า TIC ที่ได้มีค่าเท่ากับหนึ่ง ก็หมายความว่าแบบจำลองที่ประมาณขึ้นมาเป็นแบบจำลองที่ไม่ดี ไม่สามารถจะเป็นตัวแทนของข้อมูลจริงได้ ค่า TIC สามารถได้จากสมการ ดังต่อไปนี้

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

กำหนดให้ X_t^s = ค่าที่ประมาณจากแบบจำลอง

X_t^a = ค่าข้อมูลจริง

T = จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

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

คือ การวัดค่าของตัวแปรอิสระว่า สามารถอธิบายตัวแปรตามได้มากน้อยเพียงใด หากค่า R^2 เท่ากับหนึ่ง นั่นหมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้อย่างสมบูรณ์ ในทางตรงข้าม หากค่านี้มีค่าเท่ากับศูนย์ ก็ความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้

แต่อย่างไรก็ตาม หากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมากก็จะทำให้ค่า R^2 มากขึ้นด้วย ซึ่งถือว่าเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการ (3.20)

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

ดังนั้นเพื่อปรับปรุงข้อจำกัดข้างต้น จึงเกิดค่าสถิติใหม่ คือค่า Adjusted R^2 ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการ (3.21)

$$\overline{R}^2 = 1 - \frac{\sum \hat{u}_i^2 /(n-k)}{\sum X_i^2 /(n-1)} \tag{3.21}$$

นอกจากนี้ Greene (2002) ยังได้กล่าวด้วยว่าการใช้ Adjusted R^2 ($\overline{R}^2$) นั้นมีคำถามว่าการใช้ Adjusted R^2 ($\overline{R}^2$) เป็นการลงโทษที่ทำให้มีการสูญเสียระดับขั้นความเสรี (degree of freedom) นั้น เพียงพอหรือยังที่จะประกันว่าเกณฑ์นี้ได้นำไปสู่การชี้แบบจำลองที่ถูกต้อง เมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะฉะนั้นจึงมีการเสนอทางเลือกอื่นที่ใช้เกณฑ์เกี่ยวกับการปรับได้อย่างดี (goodness of fit) อีกทั้ง (Gujarati; 2003) กล่าวอ้างว่า นอกจาก Adjusted R^2 ($\overline{R}^2$) ที่จะใช้วัดการปรับได้อย่างดีแล้ว ยังมีวิธีวัดแบบ Mallow's Cp, Amemiya's Pc, Hocking's Sp, Akaike's AIC, Schwarz Criterion, Hannan – Quinn Criterion และ Shibata Criterion

แต่สำหรับการค้นคว้าอิสระฉบับนี้จะกล่าวถึงเฉพาะการวัดหรือเกณฑ์ที่นิยมใช้กันมากเท่านั้น ซึ่งได้แก่ Akaike Information Criterion (AIC) และ Schwarz Criterion (SC) 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 (3.22)

$$\text{AIC} = \left[\frac{2k}{n}\right] + \log\left[\frac{\sum \hat{u}_i^2}{n}\right] \qquad (3.22)$$

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

N = ค่าสังเกตทั้งหมด

จ) Schwarz Criterion (SC)

คือ วิธีการวัดปรับได้อย่างดี (goodness of fit) ของแบบจำลองได้ดีกว่า Akaike Information Criterion (AIC) เป็นวิธีที่ประยุกต์คล้ายกับ AIC สามารถเขียนสมการ SC ได้ดังสมการ (3.23)

$$\text{SC} = \log\left[\frac{\sum \hat{u}_i^2}{n}\right] + \frac{2k \log n}{n}$$

(3.23)

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

จากค่าสถิติที่กล่าวมาข้างต้นทั้งหมด จะถูกนำมาใช้ประกอบการพิจารณาเลือกแบบจำลอง ARIMA(p,d,q) ที่มีความเหมาะสมที่สุด โดยจะทำการคัดเลือกแบบจำลองในขั้นตอนนี้ 4-5 แบบ จำลอง เพื่อทำการเปรียบเทียบว่าแบบจำลองใดมีความสามารถในการพยากรณ์มากที่สุด

2) การประมาณค่าพารามิเตอร์ (Parameter Estimation)

คือการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (autoregressive ; AR : p) และรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (moving average ; MA : q) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (simple least square) และ วิธีการถดถอยแบบไม่เป็นเชิงเส้น (nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะนำไปใช้ในการพยากรณ์ หากรูปแบบของความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุดแล้ว

3) การตรวจสอบความถูกต้อง (Diagnostic Checking)

เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์แบบจำลองแล้วต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดขึ้นนั้นมีความเหมาะสมหรือไม่ ซึ่งการตรวจสอบทำได้หลายวิธี ยกตัวอย่างเช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลอง โดยใช้การทดสอบของ Box และ Pierce ซึ่งจะแสดงได้โดยใช้ Q-Statistic ดังในสมการ (3.24)

$$Q = n\sum_{k=1}^{m}\rho_k^2 \tag{3.24}$$

กำหนดให้ n = จำนวนของข้อมูล

m = ค่า Lag length

จากสมการ (3.24) ค่า Q นั้น จะพบว่ามีการแจกแจงแบบ Chi – Square ที่มีดีกรีเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมติฐานที่ว่าสมมติฐานว่าง (null hypothesis) คือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น White Noise กล่าวคือ แบบจำลองมีลักษณะไม่มีอัตสหสัมพันธ์ (Autocorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นปราศจากอัตสหสัมพันธ์แล้วจะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสม ต้องย้อนกลับไปทำขั้นตอนที่ 1 คือทำการกำหนดรูปแบบจำลองใหม่

4) การพยากรณ์ (Forecasting)

นำแบบจำลองที่ผ่านการตรวจสอบความถูกต้องตามขั้นตอนแล้วมาใช้ในการพยากรณ์ แต่การพยากรณ์ข้อมูลไปข้างหน้านั้น ต้องใช้แบบจำลองที่สามารถให้ค่าพยากรณ์ได้แม่นยำที่สุด ดังนั้นจึงต้องมีการทดสอบแบบจำลอง โดยทำการแบ่งการพยากรณ์ออกเป็น 3 ช่วง ดังนี้

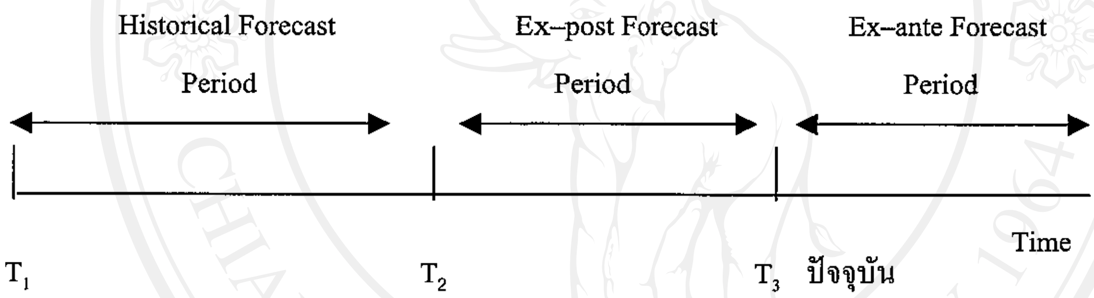
1. Historical Forecast คือการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา (T_2)

2. Ex–post Forecast คือการพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่ง (T_2-T_3) แล้วทำการพยากรณ์เปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า Root Mean Squared Error, ค่า Theil's Inequality Coefficient และค่า Akaike Information Criterion โดยพิจารณาค่าสถิติทั้งสามค่าที่มีค่าน้อยที่สุด เลือกเป็นแบบจำลองที่ดีที่สุด

3. Ex–ante Forecast คือการพยากรณ์ข้อมูลไปข้างหน้า (T_3 เป็นต้นไป) โดยนำแบบจำลองที่ดีที่สุดที่หาได้ในช่วงที่ 2 (Ex–post Forecast) นำแบบจำลองนั้นมาทำการพยากรณ์ข้อมูลไปข้างหน้า

ซึ่งการพยากรณ์ทั้งสามช่วงเวลานั้น สามารถแสดงได้ดังรูป 3.4

รูปที่ 3.4 แสดงช่วงเวลาที่ใช้ในการพยากรณ์



ที่มา : Pindyck and Rubinfeld (1998)

โดยคาบเวลา T_1 ถึง T_2 (Historical Forecast Period) แสดงถึงช่วงระยะเวลาของข้อมูลที่เรานำมาใช้ในการสร้างแบบจำลอง หรืออีกนัยหนึ่งก็คือช่วงระยะเวลาที่แบบจำลองของเราสามารถทำการจำลองหรือลอกเลียนแบบข้อมูล (simulation) หรือค่าของตัวแปรอิสระ (endogenous variable) ออกมาได้หลังจากการทำ regression แล้ว

และโดยที่ T_3 คือช่วงเวลา ณ ปัจจุบัน ซึ่งเราสามารถหาข้อมูลจากความเป็นจริงได้ของตัวแปรต่าง ๆ ในแบบจำลอง

ดังนั้นช่วงเวลา T_2 ถึง T_3 คือช่วงเวลาที่เราสามารถใช้แบบจำลองที่เราสร้างขึ้นมาดำเนินการประมาณการและทดสอบผล โดยทำการเปรียบเทียบกับค่าของข้อมูลที่มีอยู่และเกิดขึ้นจริง (Ex–post Forecast)

ส่วนช่วงเวลา T_3 ขึ้นไปนั้นเป็นช่วงเวลาในอนาคตที่ยังไม่เกิดขึ้นจริง กล่าวคือเป็นช่วงเวลาที่เราใช้แบบจำลองในการคาดคะเนหรือประมาณการค่าของตัวแปรอิสระ (endogenous variable) ในอนาคตที่ยังไม่เกิดขึ้น (Ex–ante Forecast)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ยอดขายสินค้าประกันอุบัติเหตุหมู่ในการศึกษาครั้งนี้ ใช้วิธีการพรรณนาอธิบายและวิธีการวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองให้กับอนุกรมเวลาในรูปแบบของ ARIMA Model โดยวิธีของ Box–Jenkins มีขั้นตอนดังต่อไปนี้

3.2.1) การทดสอบความนิ่งของข้อมูล (Unit root test) เป็นการพิจารณาว่าข้อมูลอนุกรมเวลามีลักษณะนิ่งหรือไม่ ทดสอบโดยวิธี Augmented Dickey–Fuller ดังสมการ (3.25) (3.26) และ (3.27) โดยในการทดสอบนั้นตัวแปรของยอดขายสินค้าประกันอุบัติเหตุหมู่ คือ $Sale_t$

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

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

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{\rho} \varphi_i \Delta x_{t-1} + \varepsilon_t \qquad (3.27)$$

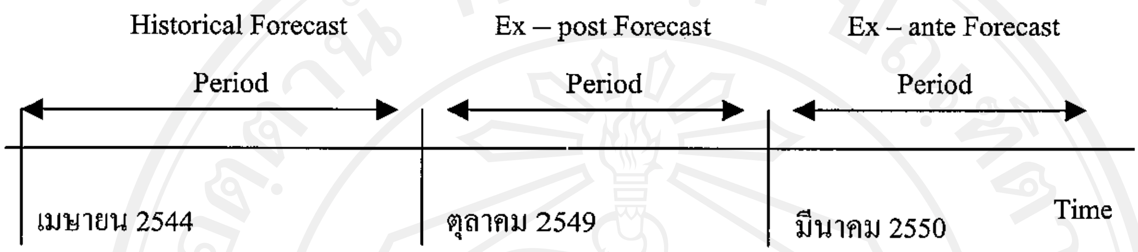
3.2.2) การกำหนดแบบจำลอง ARIMA (p,d,q) โดยพิจารณาคอเรลโลแกรม Autocorrelation Function (ACF) และค่า Partial Autocorrelation Function (PACF) เพื่อระบุว่าแบบจำลองนี้ควรมี Autoregressive, p เท่าใด และ Moving Average, q เท่าใด โดยเลือกสร้างไว้ 4-5 แบบจำลองเพื่อหาแบบจำลองที่มีความเหมาะสมที่สุด

3.2.3) การประมาณค่าพารามิเตอร์ เพื่อนำค่าพารามิเตอร์ที่ได้นั้นนำไปทำการพยากรณ์ข้อมูลต่อไป

3.2.4) การตรวจสอบความถูกต้อง เมื่อทำการหาแบบจำลองที่เหมาะสมและประมาณค่าพารามิเตอร์แล้ว จึงทำการทดสอบแบบจำลองโดยการพิจารณา Q–statistic จากคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่างว่าแตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ เพื่อเข้าสู่ขั้นตอนการพยากรณ์ต่อไป

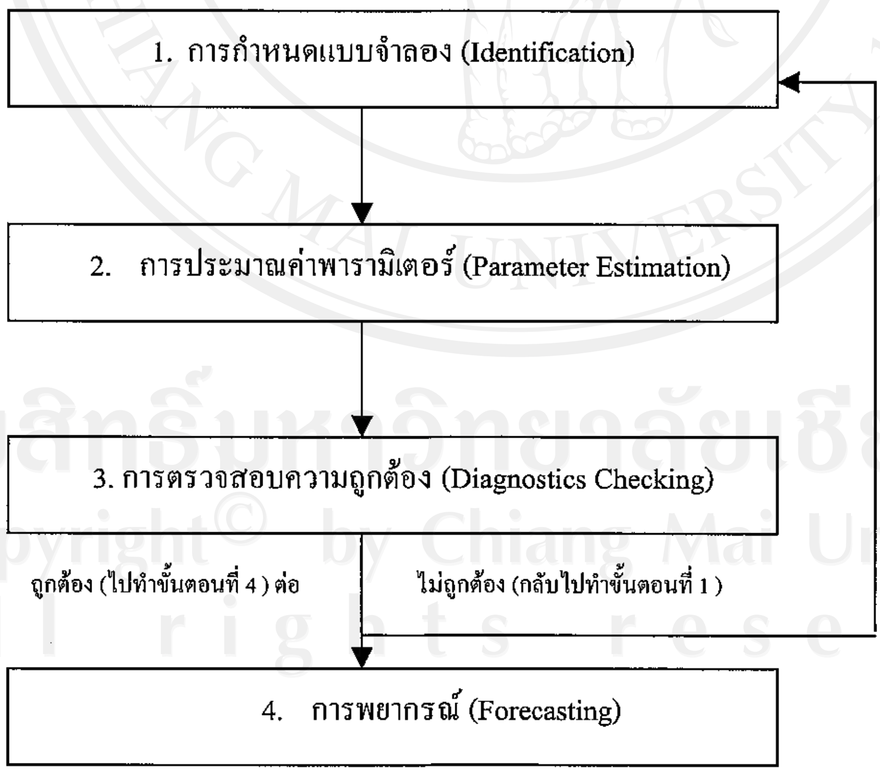
3.2.5) การพยากรณ์ ทำการพยากรณ์ยอดขายสินค้าประกันอุบัติเหตุหมู่ โดยทำการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง Historical Forecast, Ex-post Forecast และ Ex-ante Forecast แสดงได้รูป 3.5

รูปที่ 3.5 แสดงช่วงเวลาที่ใช้ในการพยากรณ์จริง



ขั้นตอนต่าง ๆ ทั้งหมดสามารถสรุปเป็นแผนภูมิแสดงการวิจัยตามวิธี Box-Jenkins ได้ดังนี้

รูปที่ 3.6 การแสดงขั้นตอนของ Box–Jenkins



ที่มา : Gujarati (2003)