บทที่ 3
ทฤษฎีและระเบียบวิธีวิจัย

3.1 แนวคิดเกี่ยวกับการพยากรณ์ข้อมูลอนุกรมเวลา

3.1.1 แนวคิดการพยากรณ์อนุกรมเวลา

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

การศึกษาอนุกรมเวลาของราคาสินค้ารายเดือน ณ ตลาดต่างๆ โดยการกำหนดรูปแบบให้กับอนุกรมเวลา เพื่อพิจารณาความมีเสถียรภาพและลักษณะรูปแบบพฤติกรรมราคามีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้องกับสินค้าที่ทำการศึกษาหรือไม่ พร้อมทั้งนำรูปแบบของสมการที่ได้ทำการวิเคราะห์อนุกรมเวลาแล้วนำมาทำนายราคาสินค้าที่ศึกษาไปล่วงหน้าเพื่อนำค่าพยากรณ์ดังกล่าวมาใช้ ในการจัดสรรทรัพยากรที่เหมาะสมต่อไป

3.1.2 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit root test)

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

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

และ $$X_t = x_{t-1} + u_{2t} \quad ; \quad u_{2t} \sim iid(0, \sigma^2_{u2}) \quad (2)$$

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

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

การทดสอบ unit root นั้นสามารถทดสอบได้โดยใช้การทดสอบ DF (Dick-Fuller(DF) test)(Dicky and Fuller, 1981) และการทดสอบ ADF (Augmented Dicky-Fuller test)(Said and Dickey 1984) สมมุติฐานว่าง (null hypothesis) ของการทดสอบ DF (DF test) คือ H_0 : $\rho = 1$ จากสมการ (3)

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{3}$$

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

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \tag{4}$$

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

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{5}$$

และมีแนวโน้มตามเวลาเชิงเส้น เราเขียนแบบจำลองได้ดังนี้

$$\Delta X_t = \alpha + \beta t + \theta X t_{t-1} + \varepsilon_t \tag{6}$$

โดยที่ t = แนวโน้มของเวลา

ถ้าหากสมการที่ (4) (5) และ (6) มีกระบวนการเชิงอัตถดถอยเข้ามาเกี่ยวข้องจะเรียกว่า การทดสอบโดยวิธี Augmented Dickey – Fuller test (ADF test) ได้สมการเป็นดังนี้

$$\Delta x_t = \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{7}$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{8}$$

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{t-1} + \varepsilon_t \tag{9}$$

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

3.1.3 การพยากรณ์โดยวิธี Box-Jenkins

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

วิธีการของ Box-Jenkins เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา รูปแบบที่จะกำหนดให้กับอนุกรมเวลาจะเป็นรูปแบบในกลุ่มของรูปแบบ ARIMA (p,d,q) (Intergrated Autoregreesive-Moving Average order p and q) ซึ่งเป็นการรวมส่วนของรูปแบบ AR (p) และรูปแบบ MA (q) เข้าด้วนกัน ส่วนอันดับของ d คือจำนวนครั้งที่หาผลต่าง (Intergrated) รูปแบบ AR (p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต $\mathbf{Y}_t$ จะขึ้นอยู่กับค่าของ $\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA (q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต $\mathbf{Y}_t$ จะขึ้นอยู่กับค่าของ ความคลาดเคลื่อน $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ หรือความคลาดเคลื่อนที่อยู่ก่อนหน้า q ค่า ซึ่งรูปแบบ AR MA (p,q) มีการกำหนดรูปแบบดังนี้

แบบจำลองที่ใช้ในการพยากรณ์ คือแบบจำลองอารีมา ARIMA (p, d, q) ซึ่งประกอบไปด้วย 3 ส่วนดังนี้ การถดถอยด้วยตัวเอง (Autoregressive; AR: p) การมีอันดับ (Integrated; I: d)

และการเคลื่อนที่ของความคลาดเคลื่อน (Moving Average; MA: q) สำหรับรูปแบบทั่วไปของอารีมา (ARIMA) สามารถเขียนเป็นความสัมพันธ์ได้ดังนี้

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

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

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

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

1) อนุกรมเวลาที่เป็น Satationary series เป็นอนุกรมเวลา $\{Y_t\}$ ที่มีค่าเฉลี่ยและความแปรปรวนของ Y_t คงที่ นั่นคือค่าเฉลี่ย $E(Y_t)$ และค่าความแปรปรวน $V(Y_t)$,มีค่าคงที่สำหรับแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้มและ/หรืออิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(Y_t)$ ไม่คงที่และอนุกรมเวลาที่มีความแปรปรวนของ Y_t สูงจะเป็นลักษณะของอนุกรมเวลาที่ $V(Y_t)$ มีค่าไม่คงที่ซึ่งจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่เป็น Stationary series นอกจากนั้นอนุกรมเวลาที่เป็น Sationary series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้วยังจะต้องมีค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตที่ lag K ขึ้นอยู่กับค่า K อย่างเดียว อนุกรมเวลาที่กำหนดรูปแบบ AR MA (p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น Stationary series แล้ว

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

2.1 การหาผลต่างปกติ (Regular differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้มนั่นคือถ้าอนุกรมเวลา $\{\mathbf{Y}_t\}$ มีแนวโน้มอยู่ในอนุกรมเวลาจะแปลงให้เป็นอนุกรมเวลาใหม่

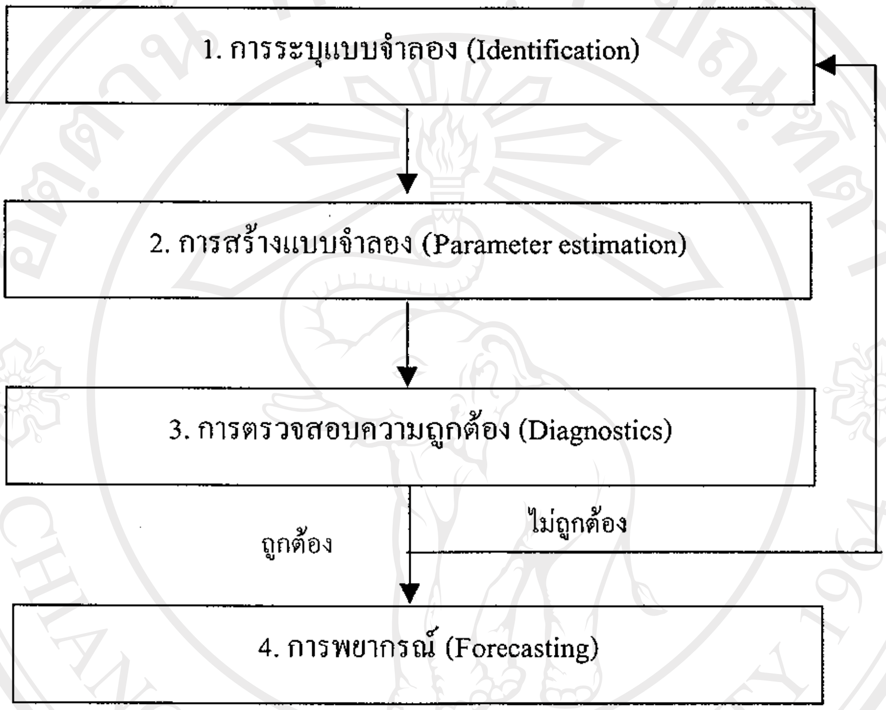
ที่ไม่มีแนวโน้ม $\{\mathbf{Z}_t\}$ โดย $\mathbf{Z}_t = \nabla^d \mathbf{Y}_t$ โดย d เป็นลำดับของการหาผลต่างและ ∇ คือผลต่างของตัวแปร เช่นเมื่อ d=1 จะได้ $\mathbf{Z}_t = \nabla \mathbf{Y}_t = \mathbf{Y}_t$ เมื่อ d = 2 จะได้ $\mathbf{Z}_t = \nabla^2 \mathbf{Y}_t = \nabla(\mathbf{Y}_t - \mathbf{Y}_{t-1}) = \nabla\mathbf{Y}_t - \nabla\mathbf{Y}_{t-1} = \mathbf{Y}_t - 2\mathbf{Y}_{t-1} + \mathbf{Y}_{t-2}$ เป็นต้น จำนวนครั้งที่หาผลต่างจะขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม่เป็น stationaries series ต้องหาผลต่างต่อไปต่อไป โดยทั่วไป ถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ d = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติก (Quadratic) จะใช้ d = 2

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

2.3 การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น Stationary series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาล โดยที่ค่า d และ D ควบคู่กันไปซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และ D จะมีค่าเท่าใดนั้นขึ้นอยู่กับว่าเมื่อหาผลต่างปกติและผลต่างฤดูกาล แล้วอนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม่เป็น Stationary series ต้องหาผลต่างต่อไป เช่น อนุกรมเวลารายเดือน ที่มีทั้งแนวโน้มและฤดูกาล เมื่อ d = 1 และ D = 1 จะแปลงอนุกรมเวลาเดิม $\{\mathbf{y}_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{\mathbf{z}_t\}$ ซึ่ง $\mathbf{Z}_t = \nabla\nabla_{12} \mathbf{Y}_t = \nabla(\mathbf{Y}_t - \mathbf{Y}_{t-12}) = \nabla \mathbf{Y}_t - \nabla\mathbf{Y}_{t-12} = \mathbf{Y}_t - \mathbf{Y}_{t-1} - \mathbf{Y}_{t-12} + \mathbf{Y}_{t-13}$ เป็นต้น

2.4 การหาลอกการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือ แปลงอนุกรมเวลาเดิม $\{\mathbf{y}_t\}$ ให้เป็นอนุกรมเวลาใหม่ $\{\mathbf{z}_t\}$ โดย $\mathbf{Z}_t = \ln(\mathbf{Y}_t)$ การแปลงนี้จะทำเมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ นั่นคือ $V(\mathbf{Y}_t)$ สำหรับค่าเวลา t ต่างๆ

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



ภาพ 3.1 แสดงขั้นตอนการพยากรณ์โดยวิธี Box and Jenkins (Gujarati , 2003)

1. การกำหนดแบบจำลอง (Identification)

การกำหนดแบบรูปแบบ (Indentification) ให้กับอนุกรมเวลาที่เป็น Stationary Series เป็นการหารูปแบบ ARMA (p.q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยที่

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

$$r_k = \frac{\sum_{t=a}^{n-k}\left(Y_{t-q}\right)\left(Y_{t+k-q}\right)}{\sum_{t=a}^{n}\left(Y_{t-q}\right)^2} \tag{10}$$

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

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

Partial Autocorrelation : ρ_{kk} คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยพิจารณาเปรียบเทียบค่า Partial Autocorrelation (r_{kk}) ของอนุกรมเวลาตัวอย่าง กับค่า Partial Autocorrealtion (ρ_{kk}) ของอนุกรมเวลาของประชากรที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_{kk} = \frac{\sum_{j=1}^{k-1}\left(r_{k-1,j}\right)\left(r_{k-j}\right)}{\sum_{j=1}^{k-1}\left(r_{k-1,j}\right)\left(r_j\right)} \tag{11}$$

การพิจารณาเปรียบเทียบแต่ละรูปแบบ ต้องพิจารณา r_k , r_{kk} กับ ρ_k และ ρ_{kk} พร้อมกันหลาย ๆ ค่า จึงมักจะพิจารณาจากรูปที่เรียกว่าคอเรลโลแกรม (Correlogram) ที่ได้จากการพล็อต r_k , r_{kk} , ρ_k และ ρ_{kk} ในช่วงเวลา k ดังนั้นการพิจารณาเปรียบเทียบ จะเป็นการเปรียบเทียบ Correlogram ของค่า Autocorrelation ของอนุกรมเวลาตัวอย่าง (r_k) กับค่า Autocorrelation ของอนุกรมเวลาของประชากร (ρ_k) และ Correlogram ของค่า Partial Autocorrelation ของอนุกรมเวลาตัวอย่าง (r_{kk}) กับค่า Partial Autocorrelation ของอนุกรมเวลาประชากร (ρ_{kk}) สำหรับแต่ละรูปแบบจะมี Correlogram ของ ρ_k และ ρ_{kk} ต่างกัน อนุกรมเวลาที่จะนำมากำหนดรูปแบบจะต้องเป็นอนุกรมเวลาที่ Stationary เท่านั้น หากไม่จำเป็น Stationary จะต้องแปลงให้เป็น Stationary เสียก่อน

การกำหนดลำดับขั้น p,q ในแบบจำลอง (Identiftying the dependence order of model) ขั้นตอนนี้คือการระบุว่าแบบจำลองนี้ควรจะมี Autoregressive,p เท่าใด Differencing, d ที่ลำดับเท่าใด และ Moving average ,q เท่าใด โดยพิจารณาจาก ACF และ PACF ซึ่งอาจจะใช้ตาราง 4 ดังต่อไปนี้พิจารณาร่วม

ตาราง 3.1 ตารางแสดงการพิจารณา ACF และ PACF

ชนิดของแบบจำลอง	รูปแบบของ ACF	รูปแบบของ PACF
AR(p)	ลู่โค้งเข้าหาแกน (Tails off)	เกิดค่าที่ชัดเจนเพียง p ค่าแล้วหายไป (Cut off after Lag p)
MA(q)	เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป (Cut off after Lag p)	ลู่โค้งเข้าหาแกน (Tails off)
ARMA(p,q)	ลู่โค้งเข้าหาแกน (Tails off)	ลู่โค้งเข้าหาแกน (Tails off)

ที่มา : Gujarati, N. (2003)

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

1.1 ค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE)

ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root mean square Error: RMSE) เป็นการวัดค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองว่ามีความแตกต่างกันมากน้อยเพียงใด ซึ่งหากค่า RMSE มีค่าเท่ากับศูนย์ จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงไรก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากเพียงนั้น สามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)} \tag{12}$$

กำหนดให้ Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

1.2 ค่าTheil's inequality coefficient

โดยในหลักการเบื้องต้น พบว่าสมการที่ใช้นี้ยังคงมีหลักการที่คล้ายกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือค่าสถิตินี้จะมีค่าอยู่ระหว่าง 0 และ 1 ทั้งนี้หากค่า U มีค่าเท่ากับศูนย์ นั่นก็หมายความว่าค่าที่ได้จากการประมาณมีค่าเท่ากันพอดีกับค่าที่เป็นข้อมูลจริง แสดงถึงแบบจำลองที่ประมาณได้เป็นแบบจำลองที่เป็นตัวแทนข้อมูลจริงได้อย่างดีที่สุด ในขณะที่ถ้า U มีค่าเท่ากับหนึ่ง แปลว่าแบบจำลองที่ประมาณได้เป็นแบบจำลองที่ไม่ดีที่สุด ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่า U ที่น้อย ๆ ดังจะพิจารณาได้จากสมการที่ (13)

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^w - Y_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^s\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^a\right)^2}} \tag{13}$$

กำหนดให้ Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

1.3 ค่า Adjusted R^2

ค่า Adjusted R^2 คือ การวัดค่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใด หากค่านี้เท่ากับ 1 ก็หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100 % ในทางกลับกัน หากค่านี้มีค่าเท่ากับ 0 แปลความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่อย่างไรก็ตาม พบว่าหากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมาก ๆ ก็จะทำให้ค่า R^2 มากขึ้นด้วย ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการที่ (14) ดังนั้นเพื่อปรับปรุงข้อจำกัดข้างต้น จึงเกิดค่าสถิติใหม่ คือค่า Adjusted R^2 ($\overline{R}^2$) ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการที่ (15)

$$R^2 = 1 - \frac{\Sigma \hat{u}_i}{\Sigma y_i^2} \tag{14}$$

$$\overline{R}^2 = 1 - \frac{\Sigma \hat{u}_i^2 / (n-k)}{\Sigma y_i^2 / (n-1)} \tag{15}$$

1.4 ค่า Akaike Information Criterion (AIC)

ค่า Akaike Information Criterion (AIC) คือ ค่าสถิติที่ประยุกต์คล้ายกับ $\overline{R}^2$ แต่ใช้รูปแบบการใส่ค่าลอกกาลิทึมฐานธรรมชาติ (Natural logarithm) โดยหากค่าสถิตินี้มีค่าน้อยเพียงใด นั่นก็แปลว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (Lag Length) ที่เหมาะสมอีกด้วย ซึ่งแสดงในสมการที่ (16)

$$\ln AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{\Sigma \hat{u}_i^2}{n}\right) \tag{16}$$

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

n คือ ค่าสังเกตทั้งหมด

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

2. การประมาณค่าพารามิเตอร์ (Parameter estimation)

การประมาณค่าพารามิเตอร์ คือการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (AR) และรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (MA) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (Simple Least Square) แต่สามารถที่จะใช้วิการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์ได้ หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3. การตรวจสอบความถูกต้อง (Diagnostics)

การตรวจสอบแบบจำลอง เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลอง จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธี ยกตัวอย่าง เช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลอง โดยใช้การทดสอบของ Box และ Pierce ซึ่งจะแสดงได้โดยใช้ Q statistic ดังในสมการที่ (15)

$$Q = n\sum_{k=1}^{m}\rho_k^2 \tag{17}$$

กำหนดให้ n คือจำนวนของข้อมูล

m คือค่า Lag Length

จากสมการที่ (15) ค่า Q นั้นจะพบว่ามีการแจกแจงเป็นแบบ Chi-square ที่มีดีกรีเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมติฐานว่า สมมติฐานว่าง คือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น White Noise นั่นก็แปลว่าแบบจำลองมีลักษณะปราศจากอัตสหสัมพันธ์ (Autocorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นปราศจากอัตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่1 เพื่อกำหนดรูปแบบจำลองใหม่

4. การการพยากรณ์ (Forecasting)

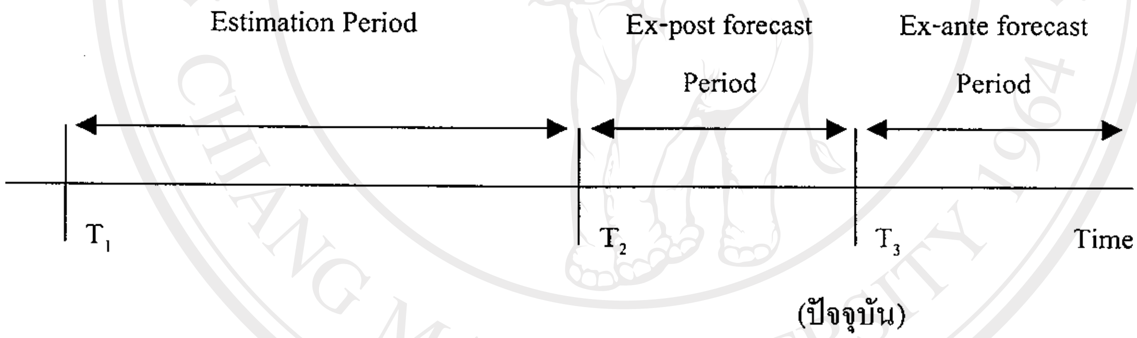
การพยากรณ์จะทำได้ทั้งแบบจุด (Point Forecast) และแบบช่วง (Interval Forecast) ซึ่งการพยากรณ์ควรจะเป็นการพยากรณ์ในช่วงระยะเวลาสั้นๆ เพื่อความเหมาะสมกับเครื่องมือทางเศรษฐมิติ ในการศึกษาครั้งนี้คือ อารีมา

ขั้นตอนต่าง ๆ ในการพยากรณ์ซึ่งสามารถแบ่งได้ 3 ขั้นตอนดังนี้

4.3 Estimation Period หมายถึง การใช้แบบจำลองเพื่อคาดคะเนในช่วงข้อมูลตั้งแต่ T_1 ถึง T_2 จากข้อมูลทั้งหมดที่มีถึง T_2

4.3 Ex-Post Forecast หมายถึง การพยากรณ์โดยใช้ข้อมูล T_2 ถึง ข้อมูลปัจจุบันT_3 เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ของแต่ละแบบจำลอง โดยนำค่าพยากรณ์ที่ได้มาเปรียบเทียบกับข้อมูลที่เรามีอยู่จริง

4.3 Ex-ante Forecast หมายถึง การพยากรณ์ในช่วงเวลาต่อไปถัดจาก T_3 เป็นต้นไป



ภาพ 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์ (Pindyck and Rubinfeld, 1997)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ปริมาณการซื้ออาหารจากฝ่ายโภชนาการระหว่างประเทศ บริษัท การบินไทย จำกัด (มหาชน) โดยวิธี ARIMA เป็นการนำเอาข้อมูลอนุกรมเวลามาหารูปแบบแบบจำลองที่เหมาะสมที่สุด โดยมีขั้นตอนดังนี้

1. การหาลอการิทึมธรรมชาติ (Natural Logarithm : ln) ข้อมูลอนุกรมเวลาปริมาณการซื้ออาหาร และ X_t ได้ถูกเปลี่ยนเป็น Q_t ตัวแปรที่จะใช้ในสมการต่างๆจึงถูกเปลี่ยนเป็น $\ln Q_t$

2. การทดสอบความนิ่งของข้อมูล เป็นการพิจารณาว่าข้อมูลอนุกรมเวลามีลักษณะนิ่งหรือไม่ โดยการทดสอบ Unit Root ดังสมการต่อไปนี้

$$\Delta Q_t = \theta Q_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta Q_{t-1} + \varepsilon_t \qquad (18)$$

$$\Delta Q_t = \alpha + \theta Q_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta Q_{t-1} + \varepsilon_t \qquad (19)$$

$$\Delta Q_t = \alpha + \beta t + \theta Q_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta Q_{t-1} + \varepsilon_t \qquad (20)$$

3. การกำหนดรูปแบบแบบจำลอง (Identification) ARIMA (p,d,q) โดยการพิจารณาคอเรลโลแกรม Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF) เพื่อจะสามารถระบุว่าแบบจำลองควรจะมี Autoregressive (p) เท่าใด และ Moving Average (q) เท่าใด โดยจะเลือกมาเพื่อหาแบบจำลองที่เหมาะสมที่สุด

4. การประมาณค่าพารามิเตอร์ เพื่อที่จะนำค่าที่ได้นั้นไปทำการพยากรณ์ปริมาณซื้อในอนาคตได้

5. การตรวจสอบความถูกต้อง ทำการทดสอบโดยพิจารณาจากค่า Q-Satistic จากคอเรล-โลแกรมของอัตตสหสัมพันธ์ของกลุ่มข้อมูลตัวอย่าง

6. การพยากรณ์ ทำได้โดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ Historical Forecast คือ ตั้งแต่ เดือนมกราคม ค.ศ. 1999 ถึงเดือนกันยายน ค.ศ. 2003, Ex-post Forecast ตั้งแต่เดือนตุลาคม ค.ศ. 2003 ถึงเดือนธันวาคม ค.ศ. 2003 และ Ex-ante Forecast ตั้งแต่เดือนมกราคม ค.ศ. 2004 ถึงเดือนมีนาคม ค.ศ. 2004