บทที่ 3

แนวความคิดและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

การวิเคราะห์อนุกรมเวลาแบบคลาสิคนั้นมีการพยากรณ์ที่ซับซ้อน เมื่อมีอิทธิพลของแนวโน้ม วัฏฏะ และฤดูกาลที่ชัดเจน ส่วนการวิเคราะห์แบบ Box-Jenkins สามารถที่ขจัดปัญหาดังกล่าวได้ (วิชิต หล่อจีระชุณห์กุลและคณะ, 2539) ทำให้การพยากรณ์ข้อมูลที่เป็นอนุกรมเวลาแม่นยำและสามารถนำไปพยากรณ์ค่าของข้อมูลในอนาคตได้

3.1.1 การทดสอบความนิ่งของข้อมูล (Unit Root Test)

การทดสอบความนิ่งของมูลเป็นการแก้ปัญหาการถดถอยที่แท้จริง(Spurious Regression) ที่เกิดจากข้อมูลที่มีแนวโน้ม และการทดสอบความนิ่งยังเป็นการหาสมการการประมาณค่าที่มีความคลาดเคลื่อนต่ำ แล้วสามารถที่จะนิยามคำว่า “นิ่ง (Stationary)” ได้ดังนี้กระบวนการเฟ้นสุ่ม (Stochastic Process) จะถูกเรียกว่า นิ่ง (Stationary) ถ้าค่าเฉลี่ย (Mean) และความแปรปรวนของกระบวนการเฟ้นสุ่ม มีค่าคงที่เมื่อเวลาเปลี่ยนไปและค่าของความแปรปรวนระหว่างสองคาบเวลาจะขึ้นอยู่กับระยะทาง (Distance) หรือความล้าหลัง (Lag) ระหว่างคาบเวลาทั้งสอง และไม่ขึ้นอยู่กับเวลาที่เกิดขึ้นจริงที่ความแปรปรวนร่วมได้ถูกคำนวณ (Gujarati, 2003) และสามารถเขียนเป็นสมการทางคณิตศาสตร์ได้ดังนี้

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

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

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

ถ้าพิจารณาได้ว่าค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) มีค่าคงที่เมื่อเวลาเปลี่ยนไป ในขณะที่ความแปรแรวนร่วม (Covariance) ระหว่างสองคาบเวลาจะขึ้นอยู่กับความห่างระหว่างเวลา (Gap) เท่านั้น ไม่ได้ขึ้นอยู่กับเวลาที่เกิดขึ้นจริง ถ้าพบว่าไม่เป็นไปตามที่กล่าวมาแล้วกระบวนการเฟ้นสุ่มจะถูกเรียกว่ามีลักษณะ “ไม่นิ่ง” (Nonstationary) (Charemza and Deadman, 1992)

การใช้ข้อมูลอนุกรมเวลาโดยไม่ได้ตรวจสอบความนิ่ง (Stationary) จะพบว่าสมการการถดถอยที่มีตัวแปรลักษณะไม่นิ่ง (Nonstationary) ค่าสถิติที่ใช้ในการทดสอบคือ ค่าสถิติ t (t-statistics) จะมีการแจกแจงมาตรฐานไม่ปกติ ซึ่งผลที่ได้จากการประมาณสมการการถดถอยไม่ถูกต้อง (Johnston and Dinardo, 1997) ส่งผลกระทบกับค่าการพยากรณ์ทำให้ความแม่นยำลดน้อยลงหรืออาจไม่ถูกต้อง

ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์ (2542) ได้กล่าวว่า ข้อสมมติฐานเบื้องต้นหลังการประมาณค่าทางเศรษฐมิติโดยใช้ข้อมูลอนุกรมเวลานั้น คือข้อสมมติเกี่ยวกับความนิ่งของข้อมูล โดยมีสมมติฐานแบบจำลองดังนี้

$$Y_t = \alpha + \beta X_t + u_{1t} \tag{1}$$

และ

$$X_t = X_{t-1} + u_{2t}; u_{2t} \sim \text{iid}\,(0, \sigma^2_{u_2}) \tag{2}$$

โดยที่ u_{2t} เป็นอนุกรมของตัวแปรสุ่ม (Random Variables)ที่มีการแจกแจงแบบปกติและอิสระต่อกัน มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนคงที่ ซึ่งตัวแปร X ก็จะเป็นแนวเดินเชิงสุ่ม (Random Walk) และเป็น Integrated Of Order One, I(1) เพราะฉะนั้นตัวแปร Y ก็จะเป็น I(1) ด้วย

การทดสอบ Unit Root สามารถที่จะทดสอบได้จาก DF-test (Dicky-Fuller test) (Dicky-Fuller, 1981) และการทดสอบ ADF-test (Augmented Dicky-Fuller test) (Said and Dicky 1984) โดยมีการตั้งสมมติฐานว่างของการทดสอบคือ

$$H_0: \rho = 1$$

$$H_1: \rho \neq 1$$

จากสมการ

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

ซึ่งเรียกการทดสอบ Unit Root

โดยที่ $|\rho| < 1$;x_t จะมีลักษณะนิ่ง

และ $\rho = 1$;x_t จะมีลักษณะไม่นิ่ง

แต่การทดสอบ Unit Root ยังสามารถทดสอบได้อีกวิธีหนึ่งคือ การตั้งสมมติฐานว่าง

$$H_0: \theta = 0$$

$$H_1: \theta < 0$$

จากสมการ $$\Delta x_t = \theta x_{t-1} + \varepsilon_t \qquad (4)$$

ซึ่งมีคือ $X_t = (1+\theta)X_{t-1} + \varepsilon_t$ นั้นก็เป็นสมการที่ (3) นั้นเอง โดยที่ $\rho = (1+\theta)$ ถ้า θ ในสมการ (4) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ(3) จะมีค่าน้อยกว่า 1 ดังนั้นสามารถที่จะสรุปได้ว่า การปฏิเสธ $H_0: \theta = 0$ แสดงว่า $\rho < 1$ และ X_t มี Integration Of Order Zero (Charemza and Deadman, 1992) นั่นคือ X_t มีลักษณะนิ่ง

ถ้า X_t เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปร่วมอยู่ด้วย (Random Walk With Drift) เราสามารถที่จะเขียนแบบจำลองได้ดังนี้

$$\Delta x_t = \alpha + \theta x_{t-1} + \varepsilon_t \qquad (5)$$

และถ้า X_t เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปร่วมอยู่ด้วย (Random Walk With Drift) และมีแนวโน้มตามเวลาเชิงเส้น (Linear Time Trade) สามารถที่จะเขียนแบบจำลองได้ดังนี้

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \varepsilon_t \qquad (6)$$

โดยที่ t = เวลา ซึ่งก็จะมีทำการทดสอบ $H_0: \theta = 0$ โดยมี $H_1: \theta < 0$ เช่นเดียวกับที่กล่าวมา สรุปแล้ว Dicky-Fuller (1979) ได้พิจารณาสมการถดถอย 3 รูปแบบที่แตกต่างกันในการทดสอบว่ามี Unit Root ได้แก่

$$\Delta x_t = \theta x_{t-1} + \varepsilon_t$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \varepsilon_t$$

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \varepsilon_t$$

โดยตัวพารามิเตอร์ที่อยู่ในความสนใจทุกสมการคือ θ นั่นคือ ถ้า $\theta=0$; X_t จะมี Unit Root โดยการเปรียบเทียบค่าสถิติ t ที่คำนวณได้กับค่าที่เหมาะสมที่อยู่ในตาราง Dicky-Fuller (Dicky-Fuller tables) (Enders, 1995) หรือกับค่าวิกฤติ MacKinnon (Gujarati, 2003) แต่ค่าวิกฤติจะไม่เปลี่ยนแปลง ถ้าสมการ (4) , (5) , (6) ถูกแทนที่โดยกระบวนการเชิงอัตถดถอย และวิธีการทดสอบจะเรียกว่าการทดสอบโดยวิธี Augmented Dickey – Fuller test (ADF test) ได้สมการเป็นดังนี้

$$\Delta x_t = \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \quad (7)$$

$$\Delta x_t = \alpha + \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \quad (8)$$

$$\Delta x_t = \alpha + \beta t + \theta x_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta x_{t-i} + \varepsilon_t \quad (9)$$

ซึ่งการทดสอบนี้จะมีการทดสอบเหมือนกับการทดสอบ DF-test จึงสามารถใช้ค่าวิกฤตในการพิจารณาแบบเดียวกัน ส่วนการหา Lag Length ตามวิธีของ Enders (1995) นั้นให้เริ่มต้นที่ Lag Length ที่มากพอสมควรค่าหนึ่งแล้วค่อยๆลดค่า Lag Length ลงเรื่อยๆ เมื่อพบว่าค่าสถิติ t-test หรือ F-test ที่ใช้ในการทดสอบนั้นไม่มีนัยสำคัญ ณ ระดับความเชื่อมั่นที่กำหนดจนกระทั่งค่าสถิติมีนัยสำคัญจึงจะถือว่าค่า Lag นั้นมีความเหมาะสม

3.1.2 แนวคิดการวิเคราะห์ข้อมูลอนุกรมเวลา Box-Jenkins

วิธีการของ Box – Jenkins นั้นเป็นวิธีการพยากรณ์ที่มีความแม่นยำกว่าวิธีอื่นเป็นวิธีการพยากรณ์ที่มีประสิทธิภาพและเหมาะสมกับลักษณะของข้อมูลที่มีลักษณะเป็นอุนกรมเวลาและหากต้องการจะพยากรณ์ช่วงเวลาที่ยาวนานควรนำข้อมูลที่ทันสมัยมาปรับค่าพยากรณ์ที่ได้ทำไว้แล้วเพื่อให้ได้ค่าความคลาดเคลื่อนจากการพยากรณ์ลดลง

วิธีของ Box-Jenkins เป็นการวิเคราะห์ตามหลักคณิตศาสตร์ที่ใช้ข้อมูลที่มีจำนวนมาก โดยการใช้สมการทางคณิตศาสตร์ให้เหมาะสมกับข้อมูลที่เป็นอนุกรมเวลา สมการที่จะกำหนดให้กับข้อมูลอนุกรมเวลาจะเป็นสมการทางคณิตศาสตร์ในกลุ่มของรูปแบบ ARIMA (p,d,q) (Intergrated Autoregreesive-Moving Average Order p And q) ซึ่งเป็นการรวมส่วนของรูปแบบ AR (p) และรูปแบบ MA (q) เข้าด้วยกัน ส่วนอันดับของ d คือจำนวนครั้งที่หาผลต่าง (Intergrated) รูปแบบ AR (p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่าของ $Y_{t-1},..., Y_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA (q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่าของ ความคลาดเคลื่อน $\varepsilon_{t-1},..., \varepsilon_{t-q}$ หรือความคลาดเคลื่อนที่อยู่ก่อนหน้า q ค่า ซึ่งรูปแบบ AR MA (p,q) มีการกำหนดรูปแบบดังนี้

AR(p) คือ $Y_t = \theta_0 + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t$

MA(q) คือ $Y_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

ARMA(p,q) คือ $Y_t = \theta_0 + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

ARIMA(p,d,q) คือ $\Delta^d Y_t = \theta_0 + \phi_1 \Delta^d Y_{t-1} + ... + \phi_p \Delta^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

การกำหนดรูปแบบที่เหมาะสมให้กับอนุกรมเวลาจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) อนุกรมเวลาที่จะนำมาศึกษาเพื่อประโยชน์ในการพยากรณ์นั้นการเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่างๆ ได้แก่ ตัวแปรวัฏจักร (Cyclical Factor) เหตุการณ์ที่ผิดปกติ (Irregular Movement) แนวโน้ม (Trend) และตัวแปรฤดูกาล (Seasonal Factor) โดยวิธีการพยากรณ์ Box-Jenkins จะสามารถแบ่งวิเคราะห์อนุกรมเวลาออกเป็น 2 ประเภทดังนี้

1. อนุกรมเวลาที่เป็น Stationary Series เป็นอนุกรมเวลา (Y_t) ที่มีค่าเฉลี่ย $E(Y_t)$ และความแปรปรวน $V(Y_t)$ ของ Y_t คงที่ สำหรับแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้มและอิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(Y_t)$ ไม่คงที่ อีกทั้งที่อนุกรมเวลาที่มีความแปรปรวนของ Y_t สูงจะเป็นลักษณะของอนุกรมเวลาที่ $V(Y_t)$ มีค่าไม่คงที่ ซึ่งจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่มี Stationary Series นอกจากนั้นอนุกรมเวลาที่มี Sationary Series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้วยังจะต้องมีค่าสัมประสิทธิ์สหสัมพันธ์แบบในตัวเองที่ lag K ขึ้นอยู่กับค่า K อย่างเดียว อนุกรมเวลาที่กำหนดรูปแบบ ARMA (p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น Stationary Series แล้ว

2. อนุกรมเวลาที่ Nonstationary Series เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น Stationary Series การจะหารูปแบบ ARMA (p,q) ให้กับอนุกรมเวลาดังกล่าวได้จะต้องปรับของข้อมูลของอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ Stationary Series เสียก่อนการปรับอนุกรมเวลาที่ไม่เป็น Stationary Series ให้เป็นอนุกรมเวลาที่เป็น Stationary Series อาจทำได้ด้วยวิธีการต่างๆตามลักษณะของข้อมูลได้ดังนี้

2.1 การหาผลต่างปกติ (Regular Differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้มนั่นคือ ถ้าอนุกรมเวลา ($\mathbf{Y}_t$) มีแนวโน้มอยู่ในอนุกรมเวลาจะปรับให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม ($\mathbf{Z}_t$)

โดย $\mathbf{Z}_t = \nabla^m \mathbf{Y}_t$ โดย m เป็นลำดับของการหาผลต่าง

และ ∇ คือผลต่างของตัวแปร

เช่น $\nabla Z = Z_t - Z_{t-1}$

$$\nabla^m Z = \nabla^{m-1} Z_t - \nabla^{m-1} Z_{t-1}$$

จากการกำหนดรูปแบบอนุกรมเวลา Box-Jenkins มักเขียนอยู่ในอนุกรมเวลาย้อนหลัง และกำหนดสัญลักษณ์ย้อนหลัง B เรียกว่า Backward Shift Operater

ดังนั้น $B Z_t = Z_{t-1}$

และ $B^m Z_t = Z_{t-m}$

จาก $\nabla Z_t = Z_t - Z_{t-1}$

$\nabla Z_t = Z_t - BZ_t$

$\nabla Z_t = (1-B) Z_t$

$\nabla Z_t = (1-B)$

เช่นเมื่อ m=1 จะได้ $\mathbf{Z}_t = \nabla \mathbf{Y}_t = \mathbf{Y}_t$ เมื่อ m = 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 ถ้ายังไม่เป็นอนุกรม Stationaries Series ต้องหาผลต่างต่อไป โดยทั่วไป ถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ m = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติก (Quadratic) จะใช้ m = 2

2.2 การหาผลต่างฤดูกาล (Seasonal Differencing) ของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรเข้ามาเกี่ยวข้อง จะต้องปรับอนุกรมเวลาเดิม (Y_t) ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีอิทธิพลฤดูกาล (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 $Z_t = \nabla^2(Y_t - Y_{t-12})$ เป็นต้น ผลต่างนี้จะทำกี่ครั้งขึ้นกับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป

2.3 การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาล การปรับให้อนุกรมเวลาเป็น Stationary Series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาล โดยที่ค่า d และ D ควบคู่กันไปซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และ ค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และD จะมีค่าเท่าใดนั้นขึ้นอยู่กับว่าเมื่อหาผลต่างปกติและผลต่างฤดูกาล แล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป เช่น อนุกรมเวลารายเดือน ที่มีทั้งแนวโน้มและฤดูกาล เมื่อ d = 1 และ D = 1 จะแปลงอนุกรมเวลาเดิม (Y_t) ให้เป็นอนุกรมเวลาใหม่ (Z_t) ซึ่ง $Z_t = \nabla\nabla_{12} Y_t = \nabla(Y_t - Y_{t-12}) = \nabla Y_t - \nabla Y_{t-12} = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$ เป็นต้น

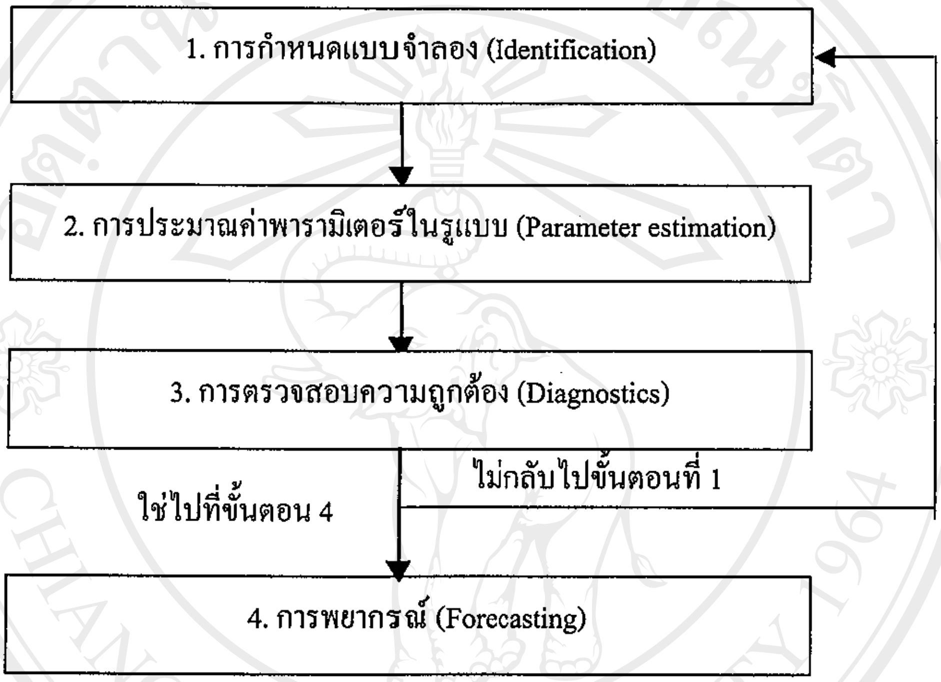
2.4 การหาลอกการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือ ปรับอนุกรมเวลาเดิม (Y_t) ให้เป็นอนุกรมเวลาใหม่ (Z_t) โดย $Z_t = \ln(Y_t)$ การใช้วิธีทางคณิตศาสตร์นี้จะทำเมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ นั่นคือ $V(Y_t)$ สำหรับค่าเวลา t ต่างๆ

3.1.3 แบบจำลองการพยากรณ์ โดยวิธี Box-Jenkins

การพิจารณาแบบจำลองโดยวิธี Box-Jenkins จะต้องพิจารณาว่าอนุกรมเวลาที่ศึกษามีลักษณะ "นิ่ง" ก่อนที่จะทำการกำหนดแบบจำลอง สำหรับอนุกรมเวลาที่มีลักษณะ "ไม่นิ่ง" จะต้องทำการปรับข้อมูลให้มีลักษณะ "นิ่ง" โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างและผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ ให้แปลงอนุกรมเวลาเดิมโดยการหาลอการิทึม ($Z_t = \ln Z_t$) จนกว่าจะได้อนุกรมเวลาใหม่ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่ที่เป็น Stationary Series แล้วจะทำตามขั้นตอนของ Box – Jenkins ดังนี้

ขั้นตอนการพยากรณ์โดยวิธี Box-Jenkins ประกอบด้วย 4 ขั้นตอน ดังนี้

1. การกำหนดรูปแบบจำลอง (Identification)
2. การประมาณค่า (Estimation)
3. วิเคราะห์ความถูกต้อง (Diagnostic Checking)
4. การพยากรณ์ (Forecasting)



รูป 3.1 แสดงขั้นตอนของ Box and Jenkins
ที่มา: Gujarati (2003)

1) การกำหนดแบบจำลอง (Identification)

สำหรับอนุกรมเวลาที่เป็น Stationary Series แล้ว เป็นการหารูปแบบจำลอง ARMA (p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาที่จะศึกษาโดยที่

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

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{10}$$

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 เป็นการประมาณค่าของประชากรซึ่งทำการประมาณค่าได้ยาก จึงต้องทำการประมาณค่าที่สุ่มมาจากตัวอย่างของประชากรซึ่งจะทำได้ง่ายและประหยัด ดังนั้นจึงได้กำหนด r_k เป็น Autocorrelation ที่มาจากตัวอย่างโดยมีสูตรดังนี้

$$r_k = \frac{c_k}{c_0} \tag{11}$$

โดยที่

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

แต่อย่างไรก็ตามอนุกรมเวลามักจะมีปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวแปรตามที่เป็นค่าระยะห่าง (Lag) ของตัวแปรตาม (Autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (Moving Average) เนื่องจาก Autocorrelation Function (ACF) จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อน แต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตามซึ่ง Patial Autocorrelation Function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule-walker (Pindyck and Rubinfeld, 1996) ดังนี้

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p \tag{10}$$

ถ้า k >p จะได้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p} \tag{11}$$

การกำหนดลำดับขั้น p,q ในแบบจำลอง (Identiftying The Dependence Order of Model) ขั้นตอนนี้เป็นการระบุว่าแบบจำลองนี้ควรจะมี Autoregressive(p) เท่าใด Differencing(d) ที่ลำดับเท่าใด และ Moving Average (q) เท่าใด โดยพิจารณาจาก ACF และ PACF ซึ่งอาจจะใช้ ตาราง 3.1 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 (2003)

จากตาราง 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)

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 ดังนั้นเพื่อประเมินแบบจำลอง ว่าแบบจำลองใดมีความเหมาะสมที่จะใช้เป็นตัวประมาณค่าของกลุ่มข้อมูลจริง สามารถพิจารณาได้จากค่าสถิติดังต่อไปนี้เพื่อประกอบในการตัดสินใจ

1. ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE) โดยจะเป็นการวัดค่าความคลาดเคลื่อนระหว่างค่าจริงเทียบกับค่าที่ประมาณได้จากแบบจำลองมีความแตกต่างกันมากน้อยเพียงใด ซึ่งหากค่า RMSE มีค่าน้อยแสดงว่าสมการที่ได้จากแบบจำลองมีความคลาดเคลื่อนน้อย ซึ่งจะเป็นการประมาณสมการที่ดีและสามารถที่จะอธิบายได้ว่าค่าที่พยากรณ์ได้มีความโน้มเอียงต่ำ ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงไรก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดี ดังนั้นที่จะพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

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

กำหนดให้ Z_t^s คือค่าที่ประมาณจากแบบจำลอง

Z_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

2. Theil's inequality coefficient เป็นค่าสถิติที่สามารถบอกถึงมาตราวัดสัมพัทธ์ของค่าความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) ว่าเกิดมาจากส่วนใด การพิจารณาค่าสัมประสิทธิ์ Theil's inequality coefficient จะทำให้รู้ถึงสาเหตุที่มีผลต่อค่า MSE ของตัวแบบการพยากรณ์ (วิชิต หล่อจีระชุณห์กุลและคณะ, 2539)

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

กำหนดให้ Z_t^s คือค่าที่ประมาณจากแบบจำลอง

Z_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

3. ค่า R^2 คือเป็นการวัดค่าตัวแปรอิสระที่จะสามารถอธิบายตัวแปรตามว่ามีความสัมพันธ์ได้มากน้อยเพียงใด หากค่า R^2 เท่ากับ 1 ก็หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100 % ในทางตรงข้าม หากค่านี้มีค่าเท่ากับ 0 ความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้ 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)

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

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

4. Akaike Information Criterion (AIC) คือค่าสถิติที่ประยุกต์คล้ายกับ $\overline{R}^2$ แต่ใช้รูปแบบการใส่ค่าลอกกาลิทึมฐานธรรมชาติ (Natural Logarithm) โดยหากค่าสถิตินี้มีค่าน้อยเพียงใดนั่นก็แปลว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าข้อนหลัง (Lag Length) ที่เหมาะสมอีกด้วย ซึ่งแสดงในสมการที่ (16)

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

กำหนดให้ $\Sigma \hat{u}_i^2$ คือผลรวมของค่าความคลาดเคลื่อน

n คือค่าสังเกตทั้งหมด

5. Schwarz Criterion (SC) คือ วิธีการวัดปรับได้อย่างดี (Goodness of Fit) ของแบบจำลองได้ดีกว่า Akaike Information Criterion (AIC) เป็นวิธีที่ประยุกต์คล้ายกับ AIC และเป็นวิธีหนึ่งใน Information Criteria โดย Information Criteria จะประกอบด้วย Schwarz Criterion (SC), the Hannan – Quinn Criterion (HQ), the Final Prediction Error (FPF) และ Akaike Information Criterion (AIC) สามารถเขียนสมการ SC ได้ดังสมการที่ (17)

$$SC = \log\left(\frac{\underline{e}'\underline{e}}{n}\right) + \frac{k \log n}{n} \tag{17}$$

2) การประมาณค่าพารามิเตอร์ในรูปแบบ (Estimation)

เมื่อได้ตัวแบบที่เหมาะสมแล้ว การประมาณค่าพารามิเตอร์ในตัวแบบที่ได้กำหนดไว้เป็นการกำหนดค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (AR) หรือรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (MA) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย แต่สามารถที่จะใช้วีธิการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์ได้ หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3) การตรวจสอบแบบจำลอง (Diagnostics)

เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลอง จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธี ยกตัวอย่างเช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลอง โดยใช้การทดสอบของ Box - Pierce ซึ่งจะแสดงได้โดยใช้ Q- statistic ดังในสมการที่ (18)

$$Q(k) = n\sum_{k=1}^{m} \rho_k^2 \tag{18}$$

กำหนดให้ n คือจำนวนของข้อมูล

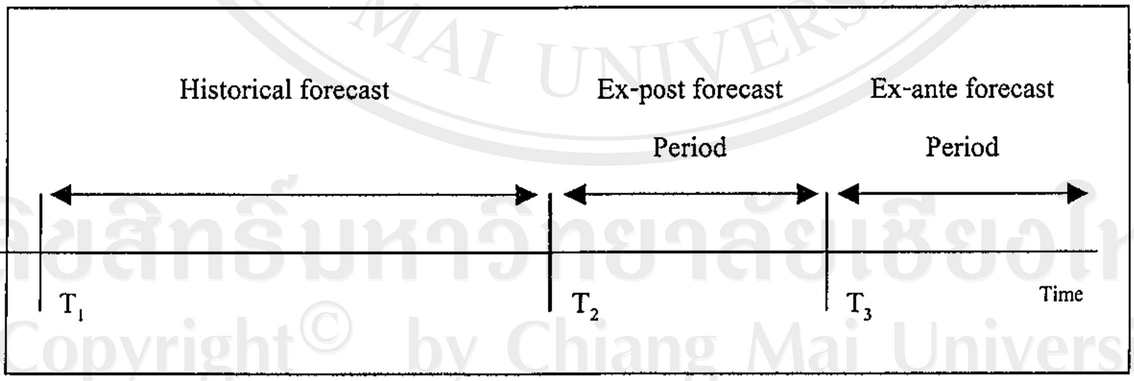
m คือค่า Lag Length

จากสมการที่ (18) ค่า Q(k) นั้นจะพบว่ามีการแจกแจงแบบ χ^2 (Chi-square) ที่มีองศาอิสระเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมติฐานว่า สมมติฐานว่างคือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น White Noise นั่นก็หมายความว่าแบบจำลองมีลักษณะไม่มีอัตสหสัมพันธ์ (Autocorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นไม่มีอัตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

4) การการพยากรณ์ (Forecasting)

เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากพิจารณาความถูกต้องแล้ว ก็สามารถนำแบบจำลองที่ได้จากการประมาณค่าไปใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลองโดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ

- Historical forecast เป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา (T_2)
- Ex-post forecast คือการพยากรณ์ โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์แล้วเปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์
- Ex-ante forecast ซึ่งเป็นการพยากรณ์ข้อมูลไปข้างหน้า



รูป 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา: Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

ในการศึกษาครั้งนี้ได้ทำการศึกษามูลค่าการส่งออกรถยนต์นั่งและชิ้นส่วนของประเทศไทย โดยศึกษาจากข้อมูลของมูลค่าการส่งออกรถยนต์นั่งและชิ้นส่วนในอดีตตั้งแต่ พ.ศ. 2536-2546 เพื่อที่พยากรณ์มูลค่าการส่งออกรถยนต์นั่งและชิ้นส่วนในอนาคต โดยวิธี Box-Jenkins ซึ่งขั้นตอนการพยากรณ์ดังนี้

1) การเปลี่ยนอนุกรม C_t ซึ่งในการศึกษาครั้งนี้คือมูลค่าการส่งรถยนต์นั่งและชิ้นส่วนให้อยู่ในรูปของลอกกาลิทึมฐานธรรมชาติ (Natural Logarithm : ln) คือ LnC_t พร้อมทั้งสมการการประมาณค่า

2) การทดสอบความนิ่งของอนุกรมเวลา LnC_t ว่ามีลักษณะนิ่งหรือไม่ กล่าวคือเป็นการทดสอบ Unit Root นั่งเอง

3) การกำหนดตัวแบบ ARIMA(p,d,q) ว่าควรจะมีลักษณะอย่างไร โดยสามารถที่จะพิจารณาได้จากการค่า Autocorrelation Function(ACF) และค่า Partial Autocorrelation Function(PACF) พร้อมทั้งการเลือกรูปแบบ ARIMA(p,d,q) โดยเป็นการรวมรูปแบบAR(p) และรูปแบบMA(q) อยู่ในรูปของสมการ ARIMA ดังนี้

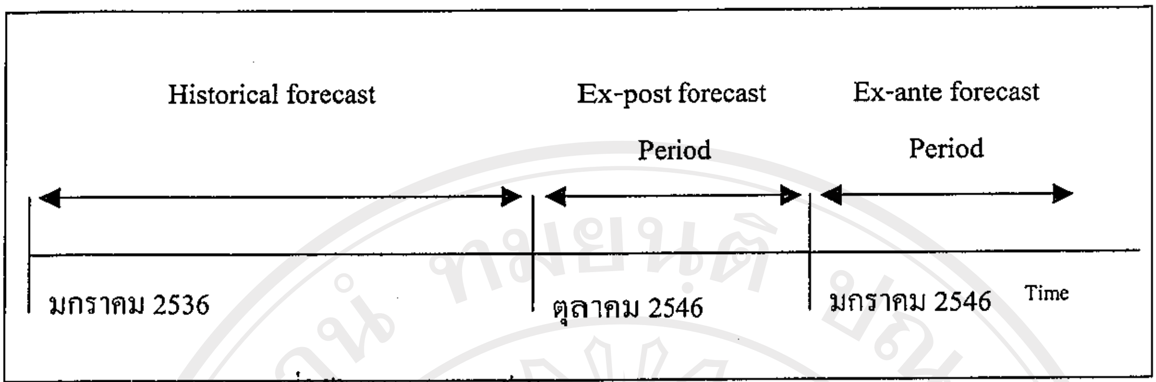
ARIMA(p,d,q) คือ

$$\Delta^d \mathrm{Ln}(C_t) = \theta_0 + \phi_1 \Delta^d \mathrm{Ln}(C_{t-1}) + \dots + \phi_1 \Delta^d \mathrm{Ln}(C_{t-q}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

4) การประมาณค่าพารามิเตอร์ เพื่อที่จะนำค่าสัมประสิทธิ์ที่ประมาณได้ไปทำการพยากรณ์มูลค่าการส่งออกรถยนต์นั่งและชิ้นส่วนในอนาคต

5) การตรวจความถูกต้องของค่าสัมประสิทธิ์ของสมการการประมาณค่าที่ได้ และทำการทดสอบตัวแบบที่ได้จากค่า Q-statistic (จากสมการที่ 18)

6) การพยากรณ์เป็นขั้นตอนสุดท้าย โดยจะทำการพยากรณ์มูลค่าการส่งออกรถยนต์นั่งและชิ้นส่วน สามารถที่จะแบ่งการพยากรณ์ได้ออกเป็น 3 ช่วงคือ ช่วง Forecast Historical ช่วงEx-post Forecast และช่วง Ex-ante Forecast



รูป 3.3 แสดงช่วงเวลาที่ใช้ในการพยากรณ์จริง