บทที่ 3

แนวความคิด ทฤษฎี และระเบียบวิธีการศึกษา

3.1 แนวความคิด และทฤษฎี

การศึกษาข้อมูลอนุกรมเวลา (Time series data) ที่ได้มาจากกระบวนการเชิงสุ่ม (Random process) ตัวแปรส่วนใหญ่มักจะมีลักษณะไม่นิ่ง (Non-stationary) กล่าวคือ ค่าเฉลี่ย (Mean) และค่าความแปรปรวน (Variances) จะมีค่าไม่คงที่เปลี่ยนแปลงไปตามกาลเวลา ทำให้ความสัมพันธ์ระหว่างตัวแปรของสมการมีความสัมพันธ์ไม่แท้จริง (Spurious regression) โดยสังเกตได้จากค่าสถิติบางค่า อาทิ ค่า t-statistic จะมีการแจกแจงที่ไม่เป็นมาตรฐาน และค่า R^2 ที่สูงในขณะที่ ค่า Durbin-Watson statistic (DW) อยู่ในระดับต่ำ แสดงให้เห็นถึงลักษณะความสัมพันธ์กันเองของค่าคลาดเคลื่อนในระดับสูง (High level of autocorrelated residuals) จึงเป็นการยากที่จะยอมรับได้ในทางเศรษฐศาสตร์ (Enders, 1995) หากนำข้อมูลที่ไม่นิ่งไปใช้งาน จะทำให้การเปรียบเทียบค่าในตารางมาตรฐานไม่ถูกต้อง นำไปสู่การลงความเห็นที่ผิดพลาดและการถดถอยที่ไม่ถูกต้อง (Spurious regression)ได้ (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542) ซึ่งข้อมูลอนุกรมเวลาที่มีลักษณะนิ่ง (Stationary) นิยามได้ ดังนี้

ค่าเฉลี่ย (Mean) : $E(X_t) = \mu$ = constant (3.1)

ความแปรปรวน (Variance) : $V(X_t) = \sigma^2$ (3.2)

ความแปรปรวนร่วม (Covariance) : $\mathrm{cov}(X_t, X_{t+k}) = E(X_t - \mu)(X_{t+k} - \mu) = \sigma_k - \mu$ (3.3)

โดย X_t คือค่าตัวแปร X ที่เวลา t ใดๆ

3.1.1 การทดสอบความนิ่งของข้อมูล (Unit Root Test)

การทดสอบ Unit root สามารถทดสอบโดยใช้การทดสอบ DF (Dickey Fuller test) (Dickey and Fuller, 1981) และการทดสอบ ADF (Augmented Dickey Fuller test) (Said and Dickey, 1984) โดยกำหนดสมมุติฐานว่าง (Null hypothesis) ของการทดสอบ DF (DF Test) คือ $H_0 : \rho = 1$ และ $H_a : |\rho| < 1$ จากสมการ (3.4)

$$X_t = \rho X_{t-1} + \hat{e}_t \tag{3.4}$$

ถ้า $|\rho| < 1$ แล้ว ข้อมูลของตัวแปร X_t จะมีลักษณะนิ่ง (Stationary) และถ้า $\rho = 1$ แล้ว X_t จะมีลักษณะไม่นิ่ง (Non-stationary) อย่างไรก็ตาม การทดสอบนี้สามารถทำได้อีกทางหนึ่ง ซึ่งเหมือนกับสมการ (3.4) กล่าวคือ

$$\Delta X_t = \theta X_{t-1} + \hat{e}_t \tag{3.5}$$

จะได้ว่า $X_t = (1+\theta) X_{t-1} + \hat{e}_t$ คือ สมการที่ (3.3) นั่นเอง โดยที่ $\rho = (1+\theta)$ ถ้า θ ในสมการ (3.5) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3.4) มีค่าน้อยกว่า 1 สรุปได้ว่า การปฏิเสธ $H_0 : \theta = 0$ เป็นการยอมรับ $H_a : \theta < 1$ หมายความว่า $|\rho| < 1$ และ X_t มีระดับผลต่างเท่ากับศูนย์ (Integration of order zero) (Charemza and Deadman, 1992: 131) นั่นคือ X_t มีลักษณะนิ่ง (Stationary) และถ้าไม่สามารถปฏิเสธ $H_0 : \theta = 0$ ได้ ก็หมายความว่า X_t มีลักษณะไม่นิ่ง (Non-stationary)

ทั้งนี้ Dickey and Fuller (1979) ได้พิจารณาสมการถดถอย 3 รูปแบบ ที่แตกต่างกันในการทดสอบ Unit root ดังนี้

Random walk $$\Delta X_t = \theta X_{t-1} + \hat{e}_t \tag{3.6}$$

Random walk with drift $$\Delta X_t = \alpha + \theta X_{t-1} + \hat{e}_t \tag{3.7}$$

Random walk with drift and trend $$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \hat{e}_t \tag{3.8}$$

โดยตัวพารามิเตอร์ที่อยู่ในความสนใจของทุกสมการ คือ θ ถ้า θ = 0 แล้ว X_t จะมี Unit root หรือมีลักษณะไม่นิ่ง (Non-stationary) โดยการเปรียบเทียบค่าสถิติ t (t-statistic) ที่คำนวณได้กับค่าวิกฤต MacKinnon (MacKinnon critical values) (Gujarati, 2003: 815)

เนื่องจากวิธีทดสอบ DF ไม่สามารถทดสอบตัวแปรกรณีที่เป็น Serial correlation ในค่า Error term ($\hat{e}_t$) ที่มีลักษณะความสัมพันธ์กันเองระดับสูง จึงมีการเพิ่ม Lag length; $\sum_{i=1}^{p}\phi_i \Delta x_{t-i}$ เข้าไปในสมการทางด้านขวามือของสมการ (3.6), (3.7) และ (3.8) ที่เสนอโดย David Dickey และ Wayne Fuller (Pindyck and Rubinfeld, 1998) ซึ่งรู้จักกันดีในชื่อของ Augmented Dickey-Fuller test ดังนี้

Random walk $$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{p}\phi_i \Delta X_{t-i} + \hat{e}_t \qquad (3.9)$$

Random walk with drift $$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{p}\phi_i \Delta X_{t-i} + \hat{e}_t \qquad (3.10)$$

Random walk with drift and trend $$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{p}\phi_i \Delta X_{t-i} + \hat{e}_t \qquad (3.11)$$

สำหรับจำนวน Lag length (p) ที่ใส่เข้าไปนั้น จะขึ้นอยู่กับความเหมาะสมของแต่ละงานวิจัย หรือสามารถใส่จำนวน Lag ไปจนกระทั่งไม่เกิดปัญหา Autocorrelation ในส่วนของ Error term (Pindyck and Rubinfeld, 1998)

Enders (1995) ได้เสนอวิธีการเลือก Lag length ที่เหมาะสมในการทดสอบ Unit root ของตัวแปรว่าควรเริ่มต้นจาก Lag length ที่สูงพอ เช่นที่ P* แล้วดูว่าสัมประสิทธิ์ของ Lag length ที่ P* นั้นแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติหรือไม่ โดยดูจากค่า t-statistic ถ้าพบว่าสัมประสิทธิ์ของ Lag length ที่ P* นั้นไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติ ก็ทำการทดสอบ Unit root ของตัวแปรนั้นโดยใช้ Lag length เท่ากับ P*-1 จนกระทั่ง Lag length ที่ใช้นั้นแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติ

ในการทดสอบ Augmented Dickey Fuller (ADF Test) เพื่อแก้ปัญหา Serial correlation ที่เกิดขึ้น โดยกำหนด Null hypothesis ของ ADF Test ในสมการ (3.9), (3.10) และ (3.11) คือ $H_0 : \theta = 0$ และ $H_a : \theta < 0$ โดยเปรียบเทียบค่า t-statistic ที่คำนวณได้ กับค่าวิกฤต MacKinnon (MacKinnon critical values) ถ้าสามารถปฏิเสธสมมติฐานได้ แสดงว่าตัวแปรที่นำมา

ทดสอบเป็น Intergrated of order zero แทนได้ด้วย $X_t \sim I(0)$ ในขณะเดียวกันสามารถทดสอบโดยใช้ค่า F-statistic ซึ่งเป็น Joint hypothesis (Φ_1 , Φ_2 และ Φ_3) เป็นสถิติทดสอบเปรียบเทียบกับค่า Dickey Fuller tables (Enders, 1995) ภายใต้ข้อสมมติฐานสมการ (3.9); $\theta = \alpha = 0$ ใช้ Φ_1 statistic สมการ (3.10); $\beta_t = \theta = \alpha = 0$ ใช้ Φ_2 statistic และ $\beta_t = \theta = 0$ ใช้ Φ_3 statistic ในการทดสอบ ซึ่งค่าสถิติดังกล่าวสามารถคำนวณได้ดังนี้

$$\Phi_i = \frac{(N-k)(SSR_R - SSR_{UR})}{r(SSR_{UR})} \quad (3.12)$$

โดยที่ SSR_R = the sum of square of residuals from the restricted model

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

N = number of observations

k = number of parameters estimated in the unrestricted model

r = number of restriction

กรณีที่ผลการทดสอบสมมติฐานพบว่า X_t มี Unit root หรือไม่มีความนิ่ง ให้ทำการแปลงข้อมูลโดยหาผลต่าง (Difference) ไปจนสามารถปฏิเสธสมมติฐานที่ว่า X_t เป็น Non-stationary เพื่อหา Order of integration (d) [$X_t \sim I(d)$; $d > 0$] สำหรับวิธีทดสอบว่าข้อมูลเป็น Non-stationary และมีอันดับความสัมพันธ์ของข้อมูล (Order of integration) ที่มากกว่า 0 [$X_t \sim I(d)$] จะทดสอบตามรูปแบบสมการดังต่อไปนี้ (วิโชติ ตั้งศักดาพร, 2540)

$$\Delta^{d+1} x_t = \alpha + \beta t + (\rho - 1)\Delta^d x_{t-1} + \sum_{i=1}^{p} \phi_i \Delta^{d+1} x_{t-i} + \hat{e}_t \quad (3.13)$$

ภายหลังจากทราบค่า d (Order of integration) แล้วต้องทำ Difference ตัวแปรเท่ากับ d+1 ครั้ง ตามกระบวนการของ Box-Jenkin's method (1970) ก่อนที่จะนำตัวแปรดังกล่าวมาทำ Regression เพื่อหลีกเลี่ยงปัญหา Spurious regression

3.1.2 ทฤษฎี ARIMA โดยวิธี Box-Jenkins

วิธีการวิเคราะห์อนุกรมเวลาของ Box-Jenkins เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา โดยเริ่มต้นจากฟังก์ชั่นสหสัมพันธ์ในตัวเอง (Autocorrelation Function, ACF) และฟังก์ชั่นสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function, PACF) รูปแบบที่เลือกใช้จะอยู่ในกลุ่มของรูปแบบ ARIMA (p,d,q) หรือเรียกว่า Autoregressive Integrated Moving-Average order p and q ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากการสังเกตหรือค่าพยากรณ์ก่อนหน้า [Autoregressive process of order p, AR(p)] และความคลาดเคลื่อนของการพยากรณ์ก่อนหน้า [Moving Average process of order q, MA (q)] โดยรูปแบบ AR(p) คือรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่า X_{t-1}, X_{t-2} ,..., X_{t-p} หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่าและรูปแบบ MA (q) คือรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่าของความคลาดเคลื่อน $\hat{e}_{t-1}$, $\hat{e}_{t-2}$,..., $\hat{e}_{t-p}$ หรือค่าความคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่า ซึ่งรูปแบบ ARMA (p,q) กำหนดได้ ดังนี้

AR(p) คือ $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \delta + \hat{e}_t$ (3.14)

MA(q) คือ $X_t = \mu + \hat{e}_t - \theta_1 \hat{e}_{t-1} - \theta_2 \hat{e}_{t-2} - \ldots - \theta_q \hat{e}_{t-q}$ (3.15)

ARMA (p, q) คือ $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \delta + \hat{e}_t - \theta_1 \hat{e}_{t-1} - \ldots - \theta_q \hat{e}_{t-q}$ (3.16)

การหาผลต่างปกติ (Regular differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้ม นั่นคือ ถ้าอนุกรมเวลา X_t มีแนวโน้มอยู่ในอนุกรมเวลา จะแปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม (Z_t) โดย

$$Z_t = \nabla^d X_t \qquad (3.17)$$

d คือ ลำดับของการหาผลต่าง

∇ คือ ผลต่างของตัวแปร

เช่น เมื่อ d = 1 จะได้ $Z_t = \nabla X_t = X_t - X_{t-1}$ (3.18)

เมื่อ d = 2 จะได้ $Z_t = \nabla^2 X_t = \nabla (X_t - X_{t-1}) = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} - X_{t-2}$ (3.19)

จำนวนครั้งที่หาผลต่างจะขึ้นอยู่กับว่า เมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม่เป็น Stationary series ต้องหาผลต่างต่อไป โดยทั่วไปถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ d = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติกจะใช้ d = 2

การหาผลต่างฤดูกาลของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรฤดูกาลเข้ามาเกี่ยวข้องจะต้องแปลงอนุกรมเวลาเดิม X_t ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล (Z_t) โดย

$$Z_t = \nabla_L^D X_t \qquad (3.20)$$

D คือ ลำดับของการหาผลต่างฤดูกาล

L คือ จำนวนฤดูกาลต่อปี

เช่น อนุกรมเวลารายเดือน (L=12)

เมื่อ D = 1 จะได้ $Z_t = \nabla_{12} X_t$ หรือ $Z_t = X_t - X_{t-12}$ (3.21)

และเมื่อ D = 2 จะได้ $Z_t = \nabla^2_{12} X_t$ หรือ $Z_t = \nabla^2(X_t - X_{t-12}) = X_t - 2X_{t-12} + X_{t-24}$ (3.22)

ผลต่างนี้จะทำกี่ครั้งขึ้นอยู่กับว่า เมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม่เป็น Stationary series ต้องหาผลต่างต่อไป

การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น Stationary series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาล d และ D ควบคู่กันไปซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และ D จะมีค่าเท่าไรนั้น ขึ้นอยู่กับว่าเมื่อหาผลต่างและผลต่างฤดูกาลแล้ว อนุกรมเวลาใหม่เป็น Stationary series หรือไม่ ถ้ายังไม่เป็น Stationary series ต้องหาผลต่างต่อไป เช่น อนุกรมเวลารายเดือนที่มีทั้งแนวโน้มและฤดูกาล เมื่อ d = 1 และ D = 1 จะแปลงอนุกรมเวลาเดิม X_t ให้เป็นอนุกรมเวลาใหม่ Z_t โดย

$$Z_t = \nabla\nabla_{12} X_t = \nabla(X_t - X_{t-12}) = X_t - X_{t-1} - X_{t-12} - X_{t-13} \qquad (3.23)$$

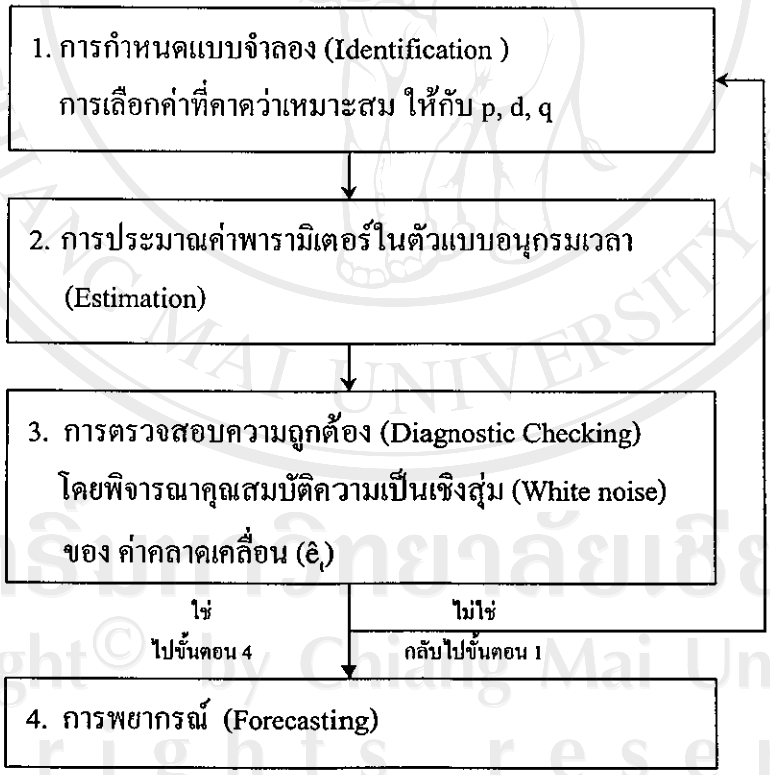
การหาค่า Correlogram ของค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ Stationary ค่า Correlogram ของ Autocorrelation (r_k) จะมีค่าลดลงค่อนข้างเร็วเมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นถ้าค่า Autocorrelation (r_k) มีค่าลดลงค่อนข้างช้า จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่า Autocorrelation (r_k) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างสูงที่ k = L, 2L, 3L จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลฤดูกาล และถ้าการเคลื่อนไหวของค่า Correlogram ของ Autocorrelation (r_k) มีลักษณะคล้ายลูกคลื่นโดยคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วว่า อนุกรมเวลาที่ศึกษาไม่เป็น Stationary ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลา จะต้องแปลงอนุกรมเวลาให้เป็น Stationary เสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ในส่วนอนุกรมเวลาที่มีอิทธิพลฤดูกาล ให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary และหากอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาล ให้หาผลต่างและผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary

แต่ถ้าอนุกรมเวลามีความแปรปรวน $[V(X_t)]$ ไม่คงที่สำหรับค่าเวลา t ต่างๆ ให้แปลงอนุกรมเวลาเดิม X_t โดยการหาลอการิทึมของค่าสังเกตให้อนุกรมเวลา นั่นคือการแปลงอนุกรมเวลาเดิม X_t ให้เป็นอนุกรมเวลาใหม่ Z_t โดย $Z_t = In(X_t)$

จากอนุกรมเวลาใหม่ที่เป็น Stationary แล้วจะทำตามขั้นตอนการพยากรณ์โดยวิธีของ Box และ Jenkins 4 ขั้นตอน ดังนี้

ภาพ 3.1 ขั้นตอนการพยากรณ์โดยวิธีของ Box และ Jenkins



ที่มา : Gujarati (2003)

1) การกำหนดแบบจำลอง (Identification)

เป็นการหารูปแบบ ARMA(p,q) ที่คาดว่าเหมาะสมให้กับอนุกรมเวลาที่เป็น Stationary series ซึ่งค่าสหสัมพันธ์ในตัวเอง (Autocorrelation Function (ACF), ρ_k) มีค่า $|\rho_k| < 1$ โดยพิจารณาเปรียบเทียบค่า ACF ของอนุกรมเวลาตัวอย่าง (r_k) กับค่า ACF ของอนุกรมเวลาของประชากร (ρ_k) ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_k = \frac{\sum_{t=a}^{n-k}(x_{t-q})(x_{t-k-q})}{\sum_{t=a}^{n}(x_{t-q})^2} \tag{3.24}$$

โดยที่ $x_t = \sum_{t=a}^{n} x_t$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

ค่าสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function (PACF), ρ_{kk}) เป็นการวัดความสัมพันธ์ของแต่ละช่วงเวลา ซึ่งมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยพิจารณาเปรียบเทียบค่า PACF ของอนุกรมเวลาตัวอย่าง (r_{kk}) กับค่า PACF ของอนุกรมเวลาของประชากร (ρ_{kk}) ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_{kk} = \frac{\sum_{j=1}^{k-1}(r_{k-1,j})(r_{k-j})}{\sum_{j=1}^{k-1}(r_{k-1,j})(r_j)} \tag{3.25}$$

การพิจารณาเปรียบเทียบแต่ละรูปแบบต้องพิจารณา r_k, r_{kk} กับ ρ_k และ ρ_{kk} พร้อมกันหลายค่า ซึ่งพิจารณาจากรูปที่เรียกว่า คอเรลโลแกรม (Correlogram) ที่ได้จากการพล๊อต r_k, r_{kk}, ρ_k และ ρ_{kk} ในช่วงเวลา k โดยเปรียบเทียบ Correlogram ของค่า ACF ของอนุกรมเวลาตัวอย่าง (r_k) กับค่า ACF ของอนุกรมเวลาประชากร (ρ_k) และ Correlogram ของค่า PACF ของอนุกรมเวลาตัวอย่าง (r_{kk}) กับค่า PACF ของอนุกรมเวลาประชากร (ρ_{kk}) โดยแต่ละรูปแบบจะมี Correlogram ของ ρ_k และ ρ_{kk} ต่างกัน

ในการกำหนดแบบจำลองว่าควรจะมี Autoregressive (p), Differencing (d) และMoving average (q) ที่ลำดับเท่าใด จะพิจารณาจาก ACF และ PACF ดังนี้

ตาราง 3.1 แสดงการพิจารณา ACF และ PACF

ชนิดของแบบจำลอง	รูปแบบของ ACF	รูปแบบของ PACF
AR(p)	ลู่โค้งเข้าหาแกน	เกิดค่าที่ชัดเจนเพียง p ค่าแล้วหายไป
MA(q)	เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป	ลู่โค้งเข้าหาแกน
ARMA(p,q)	ลู่โค้งเข้าหาแกน	ลู่โค้งเข้าหาแกน

ที่มา : Gujarati (2003)

จากตาราง 3.1 ในกำหนดรูปแบบของแบบจำลอง หาก Correlogram ของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ขณะที่ Correlogram ของ PACF เกิดค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนแท่งของค่าที่เกิดขึ้นมา ให้นับเป็นค่าที่ p ของ AR (p) เช่น เมื่อ Correlogram ของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่ง Correlogram เกิดขึ้น 1 แท่ง นั่นคือ แบบจำลองจะมีลักษณะเป็น AR (1)

สำหรับ MA (q) จะมี ACF เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น เช่น ถ้าค่า ACF เกิดแท่ง Correlogram ขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สรุปได้ว่าแบบจำลองจะมีลักษณะเป็น MA (2)

และถ้า ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองจะเป็น ARMA (p,q) และเมื่อรวมกับการทดสอบความนิ่ง (Stationary) ในขั้นตอนที่ 1 ซึ่งผลจากการทำ Difference จำนวน d ครั้งนั้น ก็จะได้แบบจำลอง ARIMA (p,d,q)

2) การประมาณค่าพารามิเตอร์ในตัวแบบอนุกรมเวลา (Estimation)

เป็นการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (Autoregressive process, AR) และรูปแบบการเคลื่อนที่ของค่าความคลาดเคลื่อน (Moving Average process, MA) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (Simple least square) แต่สามารถใช้วิธีการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะนำไปใช้ในการพยากรณ์ได้ หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด สถิติที่ใช้ในการทดสอบความมีนัยสำคัญของค่าตัวประมาณ ใช้สถิติ t (t-statistic)

ในการพิจารณาความแปรปรวนที่เกิดขึ้นกับอนุกรมเวลา X_t จะพิจารณาค่าสถิติที่สำคัญๆ ได้แก่ ค่า Adjusted R^2 ซึ่งจะอธิบายว่า ตัวแบบอนุกรมเวลาที่ได้คัดเลือกไว้สามารถอธิบายอนุกรมเวลา X_t ได้มากน้อยเพียงใด โดย

$$\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{3.26}$$

Durbin-Watson statistic (DW) ใช้ทดสอบปัญหาสหสัมพันธ์ในตัวเอง (Autocorrelation) หาก DW มีค่าเข้าใกล้ 2 แสดงว่าไม่มีปัญหา Autocorrelation หมายความว่า ตัวแบบอนุกรมเวลาที่ใช้ในการศึกษา อยู่ในระดับที่น่าเชื่อถือได้

Akaike information criterion (AIC) เป็นค่าที่แสดงระดับความคลาดเคลื่อนในการพยากรณ์ ค่ายิ่งน้อยยิ่งดี หมายความว่า ตัวแบบอนุกรมเวลานั้นมีความคลาดเคลื่อนในการพยากรณ์น้อย

$$AIC = e^{2k-n}\frac{\sum \hat{u}_i^2}{n} = e^{2k-n}\frac{RSS}{n} \tag{3.27}$$

โดย $\sum \hat{u}_i^2$ คือ ผลรวมของค่าความคลาดเคลื่อน

n คือ ค่าสังเกตทั้งหมด

3) การตรวจสอบความถูกต้อง (Diagnostic Checking)

ในการตรวจสอบความถูกต้อง จะพิจารณาคุณสมบัติความเป็นเชิงสุ่ม (White noise) ของค่าประมาณการของความคลาดเคลื่อน (Estimated residual, $\hat{e}_t$) โดยใช้ค่า Q-statistic ของ Box-Pierce ซึ่งกำหนดสมมุติฐาน $H_0 : P_1(\hat{e}_t) = P_2(\hat{e}_t) = \ldots = P_k(\hat{e}_t) = 0$ ถ้าค่า Q-statistic ของตัวแบบอนุกรมเวลาไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ 10 % แสดงว่า $\hat{e}_t$ มีคุณสมบัติความเป็นเชิงสุ่ม (White noise) หรือ $\hat{e}_t$ มีการกระจายแบบปกติ (Normal distribution) มีค่าเฉลี่ย (Mean) เท่ากับ 0 และค่าความแปรปรวน เท่ากับ $\sigma^2 I$ [$\hat{e}_t \sim NID(0, \sigma^2 I)$] แสดงว่า $\hat{e}_t$ ไม่มีสหสัมพันธ์ในตัวเอง (Autocorrelation) และมีความแปรปรวนไม่แตกต่างกัน (Heteroscedasticity) ซึ่งหมายความว่าตัวแบบอนุกรมเวลาดังกล่าว ได้ผ่านการตรวจสอบความถูกต้อง (Diagnostic Checking) และมีความเหมาะสมที่จะใช้ในการพยากรณ์ต่อไป แต่หากพบว่าแบบจำลองที่กำหนดนั้นไม่เหมาะสม จะต้องทำตามขั้นตอน 1 เพื่อกำหนดรูปแบบของแบบจำลองใหม่ สำหรับค่า Q-statistic ของ Box-Pierce มีค่า เท่ากับ

$$\text{Q - statistic} = N\sum_{k=1}^{m}\rho_k^2 \tag{3.28}$$

โดย n คือ จำนวนของข้อมูล

m คือ ค่า lag length

4) การพยากรณ์ (Forecasting)

ในการเลือกตัวแบบอนุกรมเวลาที่มีความเหมาะสมที่สุด เพื่อใช้ในการพยากรณ์ต่อไปนั้น จะพิจารณาจากค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root mean squared error: RMSE) และค่าสัมประสิทธิ์ Theil (Theil's inequality coefficient) ที่มีค่าต่ำสุด ซึ่งค่าดังกล่าวจะแสดงค่าคลาดเคลื่อนในการพยากรณ์ ค่ายิ่งน้อยยิ่งดี หมายความว่า ตัวแบบอนุกรมเวลาที่คัดเลือกนั้นให้ค่าพยากรณ์ที่มีความคลาดเคลื่อนน้อย

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)} \tag{3.29}$$

โดย Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

สำหรับ ค่าสัมประสิทธิ์ Theil (Theil's inequality coefficient, U) มีค่าเท่ากับ

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^a)^2}} \tag{3.30}$$

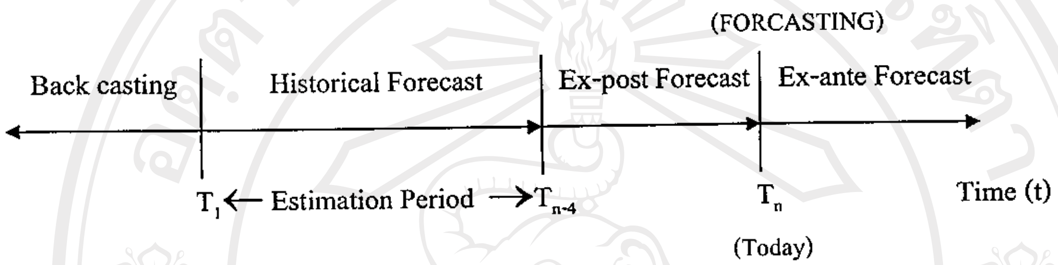
โดย Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

ในการพยากรณ์จะแบ่งออกเป็น 3 ช่วงเวลา โดยพิจารณาค่าสถิติที่ได้จากการพยากรณ์ของแต่ละแบบจำลอง ที่ผ่านการตรวจสอบความถูกต้องในขั้นตอนที่ผ่านมา แล้วเปรียบเทียบกับค่าข้อมูลจริงที่มีอยู่ เพื่อพิจารณาความเหมาะสมของสมการที่จะใช้ในการพยากรณ์ต่อไป

ภาพ 3.2 ช่วงการพยากรณ์



ที่มา : Pindyck and Rubinfeld (1998)

ภาพ 3.2 กำหนดให้ T_1 เป็นเวลาใดๆ ในช่วงที่ผ่านมา และ T_n เป็นเวลาปัจจุบัน จะได้ว่า ช่วง Historical Forecast คือช่วงพยากรณ์ที่กำหนดค่าพยากรณ์เริ่มต้นจาก T_1 ถึง T_{n-4} ค่าพยากรณ์ที่ได้จะพิจารณาเปรียบเทียบกับค่าข้อมูลจริงที่มีอยู่ ซึ่งแบบจำลองใดให้ค่าพยากรณ์ใกล้เคียงกับข้อมูลจริงมากที่สุดจะนำไปใช้ในการพยากรณ์ต่อไป

ช่วง Ex-post Forecast เป็นการนำแบบจำลองในช่วง Historical Forecast ไปทำการพยากรณ์ เพื่อเปรียบเทียบค่าพยากรณ์ที่ได้กับค่าข้อมูลจริงที่มีอยู่ ซึ่งในการพยากรณ์จะกำหนดช่วงเวลาเริ่มต้นต่อจากช่วงเวลาสุดท้ายของ ช่วง Historical Forecast นั่นคือ T_{n-4} ถึงเวลาปัจจุบัน T_n หรืออาจกล่าวได้ว่า เป็นการกำหนดช่วงพยากรณ์ย้อนกลับไป เช่น จาก T_n ย้อนกลับไป 4 ช่วงเวลา คือ ที่ T_{n-4} ผลที่ได้จากการพยากรณ์จะเป็นการพิสูจน์ว่าแบบจำลองในช่วง Historical Forecast เมื่อนำไปใช้ในการพยากรณ์ จะมีความแม่นยำเพียงใด

จากนั้นจะใช้แบบจำลองที่มีความเหมาะสมในช่วง Ex-post Forecast ทำการพยากรณ์ล่วงหน้า (Ex-ante Forecast) โดยกำหนดช่วงพยากรณ์เริ่มต้นต่อจากเวลาปัจจุบัน (T_n) เป็นต้นไป

3.2 ระเบียบวิธีการศึกษา

การศึกษาความเคลื่อนไหวและพยากรณ์ราคาส่งออกข้าวในครั้งนี้ ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลทุติยภูมิ โดยใช้ราคาส่งออกข้าว (FOB) ของไทย ชนิดข้าวขาว 100 % ชั้น 2 เป็นรายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2531 ถึงเดือนธันวาคม พ.ศ. 2546 จำนวน 192 ตัวอย่าง แทนด้วยสัญลักษณ์ P_t ข้อมูลดังกล่าวเป็นข้อมูลอนุกรมเวลา (Time series data) ที่ได้มาจากกระบวนการเชิงสุ่ม (Random process) ซึ่งตัวแปร P_t มีลักษณะไม่นิ่ง (Non-stationary) จึงได้แปลงอนุกรมเวลาเดิม (P_t) โดยการหาลอการิทึมของค่าสังเกตให้อนุกรมเวลา นั่นคือการแปลงอนุกรมเวลาเดิม P_t ให้เป็นอนุกรมเวลาใหม่ $\ln P_t$ โดย $\ln P_t = \ln(P_t)$ เพื่อให้ค่าความแปรปรวนของตัวแปรคงที่

การเลือก Lag length ที่เหมาะสมในการทดสอบ Unit root ของตัวแปร P_t ตามวิธีของ Enders (1995) เนื่องจากข้อมูลที่ใช้เป็นข้อมูลรายเดือน ซึ่งในหนึ่งเดือนมี 4 สัปดาห์ ดังนั้น จึงกำหนดค่า P-lag เริ่มต้น เท่ากับ 4

จากนั้นทำการทดสอบ Unit root โดยใช้วิธีทดสอบของ David Dickey และ Wayne Fuller (Pindyck and Rubinfeld, 1998) หรือวิธี Augmented Dickey-Fuller test ซึ่งกำหนดรูปแบบสมการได้ ดังนี้

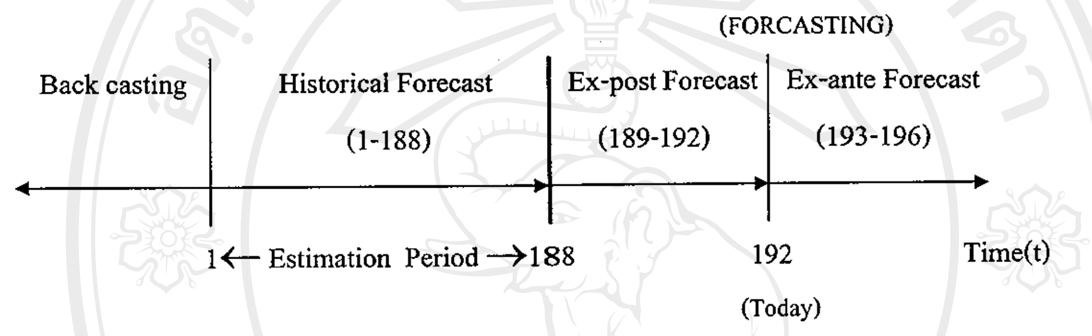
Random walk
$$\Delta \ln P_t = \theta P_{t-1} + \sum_{i=1}^{p} \phi_i \Delta P_{t-i} + \hat{e}_t \qquad (3.31)$$

Random walk with drift
$$\Delta \ln P_t = \alpha + \theta P_{t-1} + \sum_{i=1}^{p} \phi_i \Delta P_{t-i} + \hat{e}_t \qquad (3.32)$$

Random walk with drift and trend
$$\Delta \ln P_t = \alpha + \beta t + \theta P_{t-1} + \sum_{i=1}^{p} \phi_i \Delta P_{t-i} + \hat{e}_t \qquad (3.33)$$

ในการพยากรณ์ได้กำหนดช่วงการพยากรณ์ 3 ช่วงเวลาดังนี้ ช่วง Historical Forecast กำหนดค่าพยากรณ์เริ่มต้นจากค่าที่ 1 ถึง 188 ช่วงที่สอง Ex-post Forecast กำหนดช่วงเวลาเริ่มต้นค่าที่ 189 ถึง 192 และช่วง Ex-ante Forecast กำหนดช่วงพยากรณ์ต่อไปอีก 4 ช่วงเวลา คือค่าที่ 193 ถึง 196 แสดงได้ดังภาพ 3.3

ภาพ 3.3 การกำหนดช่วงการพยากรณ์



ที่มา : จากการคำนวณ