บทที่ 3

ระเบียบวิธีวิจัย

3.1 วิธีการทดสอบ Unit Root

การทอสอบ Unit Root โดยวิธี Augmented Dickey-Fuller Test หรือ ADF Test เพื่อดูว่าข้อมูลที่นำมาศึกษามีลักษณะ "นิ่ง" (Stationary) หรือ "ไม่นิ่ง" (Non- Stationary) ซึ่งรูปแบบสมการที่นำมาใช้ทดสอบคือ

$$\Delta X_t = \theta x_t + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t \tag{62}$$

$$\Delta X_t = \alpha + \theta x_t + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t \tag{63}$$

$$\Delta X_t = \alpha + \beta t + \theta x_t + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t \tag{64}$$

โดย X_t คือ ราคาข้าวหอมมะลิ ราคายางพารารมควัน ชั้น 3 และราคาน้ำตาล ณ เวลาที่ t

X_{t-1} คือ ราคาข้าวหอมมะลิ ราคายางพารารมควัน ชั้น 3 และราคาน้ำตาล ณ เวลาที่ t-1

$\alpha, \beta, \theta, \phi$ คือ ค่าพารามิเตอร์

ε_t คือ ค่าความคลาดเคลื่อนเชิงสุ่มของราคาสินค้าทั้ง 3 ชนิด

โดยรูปแบบของสมการที่ (62) คือ แนวเดินเชิงสุ่ม (None) สมการที่ (63) คือ แนวเดินเชิงสุ่มที่มีจุดตัดแกน (Intercept) สมการที่ (64) คือ แนวเดินเชิงสุ่มที่มีจุดตัดแกน และแนวโน้มเวลา (Trend & Intercept)

3.2 การหาสมการ ARIMA (p,d,q)

หาสมการพยากรณ์ราคาสินค้าเกษตรทั้ง 3 ชนิด คือ ข้าวหอมมะลิ ยางพารารมควันชั้น 3 และน้ำตาล โดยใช้แบบจำลองARIMA ซึ่งเขียนสมการได้ดังนี้

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-p} \quad (65)$$

โดยที่ y_t คือ ค่าสังเกตของอนุกรมเวลา ณ เวลาที่ t

δ คือ ค่าคงที่

$\phi_1, \phi_2, \ldots, \phi_p$ คือ พารามิเตอร์ของ Autoregressive

$\theta_1, \theta, \ldots, \theta_q$ คือ พารามิเตอร์ของ Moving Average

p คือ อันดับของ Autoregressive

q คือ อันดับของ Moving Average

ε_t คือ ค่าความคาดเคลื่อน ณ เวลาที่ t

3.3 วิธีการทดสอบ Long Memory

3.3.1 Test for Long Memory: การทดสอบ R/S

การทดสอบ R/S test ใช้ในการคำนวณค่า พารามิเตอร์ H ที่ใช้วัด ความหนาแน่นของ long range dependenceในอนุกรมเวลา

อนุกรมเวลาของช่วง T แบ่งออกเป็น n sub-series ของช่วง m และสำหรับทุกๆ sub-series โดยแต่ละ sub-series จะมีค่า m = 1,…,n, to เพื่อหาค่ามัธยฐาน (E_m) และส่วนเบี่ยงเบนมาตรฐาน (S_m) และขจัด ค่ามัธยฐานตัวอย่าง ที่ $Z_{i,m} = X_{i,m} - E_m$ สำหรับ i = 1,…..,m

หลังจากนั้นจึง สร้างอนุกรมเวลาโดยใช้รูปแบบของ $W_{i,m} = \sum_{j=1}^{i} Z_{j,m}$ โดยที่ i = 1,..,m และ เพื่อหา ระยะของ $R_m = \max\{W_{1,m,\ldots}, W_{n,m}\} - \min\{W_{1,m,\ldots}, W_{n,m}\}$

การกำหนด rescaled range ของ R_m โดยใช้ $\frac{Rm}{Sm}$ แบบเดียวกับในกรณีของอนุกรมเวลาสามารถหาค่า R, S และ H ตามข้อกำหนดต่อไปนี้

$$R = k \times T^{0.5} \quad (66)$$

กำหนดให้ R คือ ระยะ ที่อยู่ในตัวแปร , k คือ ค่าคงที่ และ T คือ ช่วงความยาวของเวลา

กำหนดให้ R/S คือ rescaled range, m จำนวนครั้งของกาสำรวจ, k คือ ค่าคงที่ และ *H* คือ Hurst exponent จะสามารถนำมาใช้ในอนุกรมเวลาขนาดใหญ่ได้

$$\frac{R}{S} = k \times m^{H} \tag{67}$$

ค่า Hurst exponent หาได้จาก :

$$\log(R/S)\text{m} = \log \text{k} + \text{H} \log \text{m} \tag{68}$$

และมีข้อกำหนดว่า :

ถ้า *H* value = 0.5 อนุกรมเวลาจะเป็นเคลื่อนไปอย่างสุ่ม และ เป็นอิสระ

ถ้า *H* value =(0, 0.5) อนุกรมเวลาจะเป็นแบบไม่คงตัว กระบวนการจะคลอบคลุมเพียงแค่วงแคบ เมื่อเทียบกับกรณีการเคลื่อนไปอย่างสุ่ม

ถ้า *H* value =(0.5, 1) อนุกรมเวลา จะเป็นชุดข้อมูลที่คงตัว กระบวนการจะคลอบคลุมเป็นวงกว้าง เมื่อเทียบกับกรณีการเคลื่อนไปอย่างสุ่ม

3.3.2 Test for Long Memory: การทดสอบ Modified R/S

โดยสมมติให้ข้อมูลอนุกรมเวลา $\{x_t\}, t = 1,2,\ldots,N$ โดยที่ $y_j = \sum_{i=1}^{j} x_i, j = 1,2,\ldots,N$ และ ความแปรปรวนของตัวอย่าง $S_j^2 = j^{-1}\sum_{i=1}^{j}(x_i - j^{-1}y_j)^2, j = 1,2,\ldots,N$ ซึ่งจะได้ R/S-statistic ดังนี้

$$R/S(j) = 1/S_j\left[\max_{0\le t\le j}(y_t - \frac{t}{j}y_j) - \min_{0\le t\le j}(y_t - \frac{t}{j}y_j)\right] \tag{69}$$

Lo (1991) ได้เสนอ Modified R/S ซึ่งพัฒนามาจาก การทดสอบ เนื่องจาก R/S test มีความอ่อนไหวอย่างมากต่อ short-range dependence โดยแทนที่ S_j ในสมการที่ (23) เป็น S_q (modified standard deviation) ที่คำนึงถึง autocovariances ใน q lags แรก ซึ่งจะช่วยละอิทธิพลของ short-range dependence ที่อยู่ในข้อมูล แทนที่จะพิจารณาหลาย ๆ lag อย่างในสมการที่ (23) ก็พิจารณาเพียง lag j = N โดย S_q มีค่าดังนี้

$$S_q = \left(\frac{1}{N} \sum_{j=1}^{N} (x_j - \bar{x}_N)^2 + \frac{2}{N} \sum_{j=1}^{q} \omega_j(q) \left[\sum_{i=j=1}^{N} (x_i - \bar{x}_N)(x_{i-j} - \bar{x}_N) \right] \right)^{1/2} \tag{70}$$

โดยที่ $\bar{x}_N$ เป็นค่าเฉลี่ยของอนุกรมเวลา

$$\omega_j(q) = 1 - \frac{j}{q+1}, q < N \tag{71}$$

ดังนั้นจะได้ Lo's Modified R/S statistic ดังสมการ

$$Q_{N,q} = \frac{1}{S_q} \left\{ \max_{0 \le i \le N} \sum_{j=1}^{i} (x_j - \bar{x}_N) - \min_{0 \le i \le N} \sum_{j=1}^{i} (x_j - \bar{x}_N) \right\} \tag{72}$$

Lo การมีอยู่ของระบบความจำระยะยาว (long memory) $N^{-1/2} Q_{N,q}$ จะเบนเข้าหาช่วง Brownian Bridge

$$W = \max_{0 \le r \le 1} V(r) - \min_{0 \le r \le 1} V(r) \tag{73}$$

โดย V คือ Standard Brownian bridge

$$V(r) = B(r) - rB(1) \tag{74}$$

โดยที่ B คือ Standard Brownian motion

การกระจายของตัวแปรสุ่ม W รู้จักกันใน

$$P(W \le x) = 1 + 2\sum_{j=1}^{\infty} (1 - 4x^2 j^2) e^{-2x^2 j^2} \tag{75}$$

Lo ใช้ช่วง [0.809,1.862] ที่ 95% (asymptotic) ยอมรับ null hypothesis

H_0 = { ไม่มี long memory, เช่น H = 0.5}

H_1 = { มี long memory, เช่น 0.5< H <1}

ค่า critical อยู่ในช่วงตามตาราง 2.1

3.3.3 Test for Long Memory: GPH Test

กระบวนการ GPH Test แสดงถึง การประมาณค่า OLS estimator ของ d จากสมการถดถอย

$$\ln\left[I(\xi)\right] = a - \hat{d}\ln\left[\sin^2\left(\frac{\xi_\lambda}{2}\right)\right] + e_\lambda \qquad , \lambda = 1,2,\ldots,V \tag{76}$$

โดยที่

$$I(\xi) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} e^{it\xi}(x_t - \bar{x})\right|^2 \tag{77}$$

และสมการ คือ Periodogram (การประมาณค่าความหนาแน่น ของ spectral) ของค่า x ที่ frequency เหมือนกันกับ ค่า bandwidth v ถูกเลือกไว้สำหรับ (ξ) , $T \to \infty$ $v \to \infty$ แต่ $\frac{v}{T} \to 0$พิจารณาว่า อิทธิพลของ T จะอยู่ระหว่าง (0.5,0.6) และสมมติฐานหลักของกระบวนการความทรงจำระยะยาว ความชันของสมการถดถอย d เท่ากับศูนย์ และ ค่า t-statistics สามารถใช้ในการแสดงผลการทดสอบได้

3.4 การหาสมการ ARFIMA (p,d,q)

หาสมการพยากรณ์ราคาสินค้าเกษตรทั้ง 3 ชนิด คือ ข้าวหอมมะลิ ยางพารารมควันชั้น 3 และน้ำตาล โดยใช้แบบจำลองARFIMA ซึ่งได้ใช้แบบจำลองนี้กับข้อมูลชุดความจำระยะยาว (Long memory) ซึ่งแบบจำลอง ARFIMA(p,d,q) สามารถเขียนสมการความสัมพันธ์ ได้ดังนี้

$$\phi\left(\beta\Delta^d y_t\right) = \delta + \theta(\beta)\varepsilon_t \tag{78}$$

กับ

$$\phi(\beta) = 1 - \phi_1(\beta) - \phi_2(\beta)^2 - \ldots. - \phi_p(\beta)^p \tag{79}$$

และ

$$\theta(\beta) = 1 - \theta_1(\beta) - \theta_2(\beta)^2 - \ldots. - \theta_q(\beta)^q \tag{80}$$

โดย

$\delta =$ ค่าคงที่

$\theta(\beta) =$ การทำงานของ moving-average ที่ order q

$\varepsilon_t =$ error term ของสมการ

$\phi(\beta) =$ การทำงานของ autoregressive ที่ order p

$\Delta^d y_t =$ ผลต่างที่ order d ของข้อมูลอนุกรมเวลา y_t

ถ้า d อยู่ที่ (0 ,0.5) ในแบบจำลอง ARFIMA อธิบายว่า Long memory หรือ Long range มี ความสัมพันธ์กันในทิศทางบวก

ถ้า d อยู่ที่ (-0.5 ,0) ในแบบจำลอง ARFIMA อธิบายว่า Intermediate memory หรือ Long range มี ความสัมพันธ์กันในทิศทางลบ

ถ้า d อยู่ที่ (0.5 ,1) ในแบบจำลอง ARFIMA อธิบายว่า เป็นค่าเฉลี่ยย้อยหลัง และไม่มีผลกระทบระยะยาวกับมูลค่าในอนาคต

ถ้าเป็นใน Short memory d=0 จะสอดคล้องกับค่ามาตรฐานตามวิธี ARMA

3.5 วิธีการเลือก Best Model

เลือกรูปแบบจำลองที่ดีของแบบจำลอง ARIMA และแบบจำลอง ARFIMA โดยพิจารณาจากค่า Akaike Information Criterion (AIC) และ Schwartz Criterion (SC) รูปแบบจำลองที่ให้ค่า AIC และ SC น้อยที่สุดจะเป็นรูปแบบที่ดีที่สุด โดย Akaike Information Criterion (AIC) และ Schwartz Criterion (SC) สามารถคำนวณได้ดังนี้

Akaike Information Criterion (AIC) $-2l/\eta + 2k/\eta$

Schwartz Criterion (SC) $-2l/\eta + k\log\eta/\eta$

โดยที่ k เป็นจำนวนพารามิเตอร์ที่ทำการประมาณค่า

η เป็นจำนวนของค่าสังเกต

l เป็นค่าของ Log Likelihood function ที่ใช้พารามิเตอร์ที่ถูกประมาณค่า k ตัว

3.6 การวัดประสิทธิภาพการพยากรณ์

การวัดประสิทธิภาพการพยากรณ์ของแบบจำลอง ARIMA และแบบจำลอง ARFIMA โดยจะใช้ค่าความคลาดเคลื่อนร้อยละเฉลี่ย (The Mean Absolute Percentage : MAPE), รากที่สองของค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสอง (Root Mean Error: RMSE) และค่าสัมประสิทธิ์ Theil (Theil's Ineqvality Coefficient: U) ซึ่งค่าทั้ง 3 แบบ มีค่าน้อย แสดงว่ารูปแบบจำลองเป็นรูปแบบที่ดีในการพยากรณ์ มีสูตรดังนี้

ก) ค่าความคลาดเคลื่อนร้อยละเฉลี่ย (The Mean Absolute Percentage Error : MAPE) ซึ่งสูตรของ MAPE คือ

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|$$

เมื่อให้ Y_t คือค่าที่แท้จริง และ $\hat{Y}_t$ คือค่าที่คาดการไว้

ข) รากที่สองของค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE) ซึ่งของ RMSE คือ

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t - Y_t)^2}$$

เมื่อให้ $\hat{Y}_t$ คือ ค่าประมาณจากแบบจำลอง

Y_t คือ ค่าที่แท้จริง / ข้อมูลจริง

ค) ค่าสัมประสิทธิ์ Theil (Theil's Ineqvality Coefficient: U) ซึ่งสูตรของ U คือ

$$U = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t - Y_t)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t)} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t)^2}}$$

เมื่อให้ $\hat{Y}_t$ คือ ค่าประมาณจากแบบจำลอง

Y_t คือ ค่าที่แท้จริง / ข้อมูลจริง

3.7 การวิเคราะห์อนุกรมเวลาแบบคลาสสิก

การวิเคราะห์อนุกรมเวลาแบบคลาสสิกของราคาสินค้าเกษตรทั้ง 3 ชนิด คือ คือ ข้าวหอมมะลิ ยางพารารมควันชั้น 3 และน้ำตาล ที่มีองค์ประกอบของ การเปลี่ยนแปลงแนวโน้ม การเปลี่ยนแปลงฤดูกาล การเปลี่ยนแปลงวัฏจักร และการเปลี่ยนแปลงที่ไม่สม่ำเสมอ

โดยการวิเคราะห์แนวโน้ม (Trend) จะใช้วิธีกำลังสองน้อยที่สุด การวิเคราะห์ฤดูกาล (Seasonal) จะใช้วิธีอัตราส่วนต่อแนวโน้ม การวิเคราะห์วัฏจักร (Cyclical) จะกำจัดแนวโน้มและฤดูกาล แล้วทำการขจัดการเปลี่ยนแปลงที่ไม่สม่ำเสมอโดยวิธีน้ำหนักถ่วงทวิ ซึ่งจะใช้การถ่วงน้ำหนักเป็น 1:2:1 และการหาการเปลี่ยนแปลงไม่สม่ำเสมอ โดยใช้วิธีการขจัด แนวโน้ม ฤดูกาล และวัฏจักร