บทที่ 3

กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการศึกษานี้จะหารูปแบบจำลองที่เหมาะสมในการประมาณราคาตั๋วสัญญาล่วงหน้าในตลาดล่วงหน้าสินค้าเกษตรโดยวิธีของ Box and Jenkins โดยใช้ข้อมูลอนุกรมเวลา และต้องใช้ข้อมูลอนุกรมเวลาที่มีลักษณะนิ่ง (stationary) โดยการศึกษาต้องทำการทดสอบ unit root ก่อนที่จะหารูปแบบจำลองที่เหมาะสมโดยวิธีอารีมาเพื่อให้ค่าความคลาดเคลื่อนจากการพยากรณ์ลดลง

3.1.1 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลา

การพยากรณ์ตั๋วสัญญาล่วงหน้าในตลาดล่วงหน้าสินค้าเกษตร โดยกำหนดรูปแบบให้กับอนุกรมเวลาเพื่อพิจารณาความมีเสถียรภาพของราคาว่ามีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้องกับสินค้าที่ต้องการศึกษาหรือไม่ รวมทั้งนำรูปแบบของสมการที่ได้ทำการวิเคราะห์อนุกรมเวลาแล้วจึงนำมาทำนายราคาสินค้าที่ศึกษาไปล่วงหน้าเพื่อทำการพยากรณ์ โดยกำหนดแบบจำลองอนุกรมเวลาในรูปแบบ ARIMA Model โดยวิธีการของ Box – Jenkins นั้นเป็นวิธีการพยากรณ์ที่มีความถูกต้องและเหมาะสมกว่าวิธีอื่นในการพยากรณ์ระยะสั้น สามารถพิจารณาได้จากความนิ่ง (stationary) ของค่าเฉลี่ย ความแปรปรวนและความแปรปรวนร่วมของข้อมูลอนุกรมเวลาดังต่อไปนี้

ค่าเฉลี่ย : $E(X_t) = \text{Constant} = \mu$ (3.1)

ค่าความแปรปรวน : $V(X_t) = \text{Constant} = \sigma^2$ (3.2)

ค่าความแปรปรวนร่วม : $Cov(X_t, X_{t+k}) = E(X_t-\mu)(X_{t+k}-\mu) = \sigma-\mu$ (3.3)

ดังนั้นสามารถสรุปได้ว่า หากข้อมูลอนุกรมเวลานั้นมีลักษณะนิ่ง (stationary) จะมีค่าเฉลี่ย ความแปรปรวนและความแปรปรวนร่วม (จากสมการที่ 3.1, 3.2 และ 3.3) มีค่าที่คงที่ ณ ทุกๆ เวลาที่เปลี่ยนแปลงไป ซึ่งสามารถทดสอบข้อมูลว่ามีลักษณะที่ความนิ่งหรือไม่ จากการทดสอบ unit root

วิธีการของ Box – Jenkins เป็นการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลาโดยการใช้ค่า Autocorrelation Function (ADF) และค่า Partial Autocorrelation Function (PADF) เป็นหลักในการพิจารณา รูปแบบที่ใช้เลือกจะอยู่ในกลุ่มของรูปแบบ Integrated Autoregressive-Moving Average order p and q อารีมา (p, d, q) ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากการสังเกตหรือการพยากรณ์ล่วงหน้าและความคลาดเคลื่อนของการพยากรณ์ล่วงหน้า โดยเป็นการรวม

ส่วนของรูปแบบ 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 ส่วนรูปแบบMA (q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่าคลาดเคลื่อน e_{t-1}, e_{t-2}, e_{t-3},….., e_{t-q} หรือค่าคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่า

3.1.2 การทดสอบความนิ่งของข้อมูล (unit root test)

แนวคิดการทดสอบ Unit Root (Gujarati, 2003) สามารถทำได้โดยใช้การทดสอบ DF (Dicky-Fuller test) (Dicky and Fuller, 1981) และการทดสอบ ADF (augmented Dicky-Fuller test) (Said and Dickey, 1984) ซึ่งมีสมมุติฐานว่าง (null hypothesis) ของการทดสอบ DF คือ H_0 : $\rho = 1$ จากสมการ (3.4) ดังนี้

$$X_t = \rho X_{t-1} + \varepsilon t \tag{3.4}$$

ซึ่งเรียกว่าการทดสอบ unit root โดยถ้า $|\rho| < 1$ แสดงว่าข้อมูล X_t จะมีลักษณะนิ่ง (stationary) แต่ถ้า $\rho = 1$ แล้วข้อมูล X_t จะมีลักษณะไม่นิ่ง (nonstationary) นอกจากนี้การทดสอบนี้สามารถทำได้อีกทางหนึ่งซึ่งคล้ายกับสมการ (4) กล่าวคือ

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \tag{3.5}$$

สมมติฐานว่าง(null hypothesis) ของการทดสอบคือ $H_0 : \theta = 0$ และ $H_a : \theta < 0$ โดย $\rho = (1+\theta)$ ซึ่งถ้า θในสมการ(3.5) มีค่าเป็นลบ แสดงว่า $\rho < 1$ (ρ จากสมการ(3.4)) ดังนั้นสามารถจะสรุปได้ว่าถ้าปฏิเสธ H_0:$\theta = 0$ หรือยอมรับ H_a:$\theta < 0$ หมายความว่า X_t มีลักษณะนิ่ง และมี integration of order zero (Charemza and Deadman, 1992) ในทางตรงข้าม ถ้ายอมรับ $H_0 : \theta = 0$ ได้ จะหมายความว่า X_t มีลักษณะไม่นิ่ง โดยการพิจารณาค่าสถิติDF เปรียบเทียบค่าวิกฤต MacKinnon (MacKinnon critical values) (Gujarati, 2003) หรือค่าวิกฤตจากตาราง Dickey-Fuller (Enders, 1995) โดยถ้า X_t มีแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) สามารถเขียนแบบจำลองได้ดังสมการที่ (3.6) และถ้ามีแนวโน้มตามเวลาเชิงเส้นรวมอยู่ด้วย (linear time trend) จะได้แบบจำลองดังสมการที่ (3.7) ตามลำดับ

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{3.6}$$

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t \tag{3.7}$$

อย่างไรก็ตามค่าวิกฤติ (critical value)จะไม่เปลี่ยนแปลง หากสมการที่ (3.5) , (3.6) , (3.7) มีกระบวนการเชิงอัตถดถอยเข้ามาเกี่ยวข้องแล้ว ซึ่งเรียกว่า การทดสอบโดยวิธี ADF (augmented Dickey – Fuller (ADF test) โดยได้สมการเป็นดังนี้

$$\Delta X_{t-1} = \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.8}$$

$$\Delta X_{t-1} = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.9}$$

$$\Delta X_{t-1} = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.10}$$

ซึ่งค่าสถิติที่ได้จากการทดสอบ ADF มีการแจกแจงเชิงเส้นกำกับ (asymptotic distribution) เหมือนกับสถิติ DF ดังนั้นก็สามารถใช้ค่าวิกฤติ ในการพิจารณาแบบเดียวกัน(Gujarati, 2003) และใช้ F (F-test) ในการทดสอบว่าแบบจำลองที่เหมาะสมนั้นมีจุดแกนและแนวโน้มเวลาหรือไม่ ส่วนในการเลือก lag length ที่เหมาะสมนั้นจะใช้วิธี Serial Correlation LM test เพื่อหา lag length ที่มีค่า probability มากที่สุด จึงจะถือว่าค่า lag นั้นมีความเหมาะสม

เมื่อพิจารณาการตรวจสอบแล้วว่าอนุกรมเวลาที่ศึกษาไม่มีลักษณะนิ่ง จะต้องแปลงให้มีลักษณะนิ่งเสียก่อน หลังจากนั้นจึงทำตามวิธี Box-Jenkins ต่อไป

3.1.3 แบบจำลองการพยากรณ์โดยวิธี Box – Jenkins

การพยากรณ์อนุกรมเวลาโดยวิธี Box – Jenkins ในรูปแบบ อารีมา (p, d, q) ต้องพิจารณาว่าอนุกรมเวลาเป็น stationary series หรือไม่ โดยพิจารณาจาก

1) ค่าเฉลี่ย $E(X_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมากจะสรุปว่า $E(X_t)$ คงที่

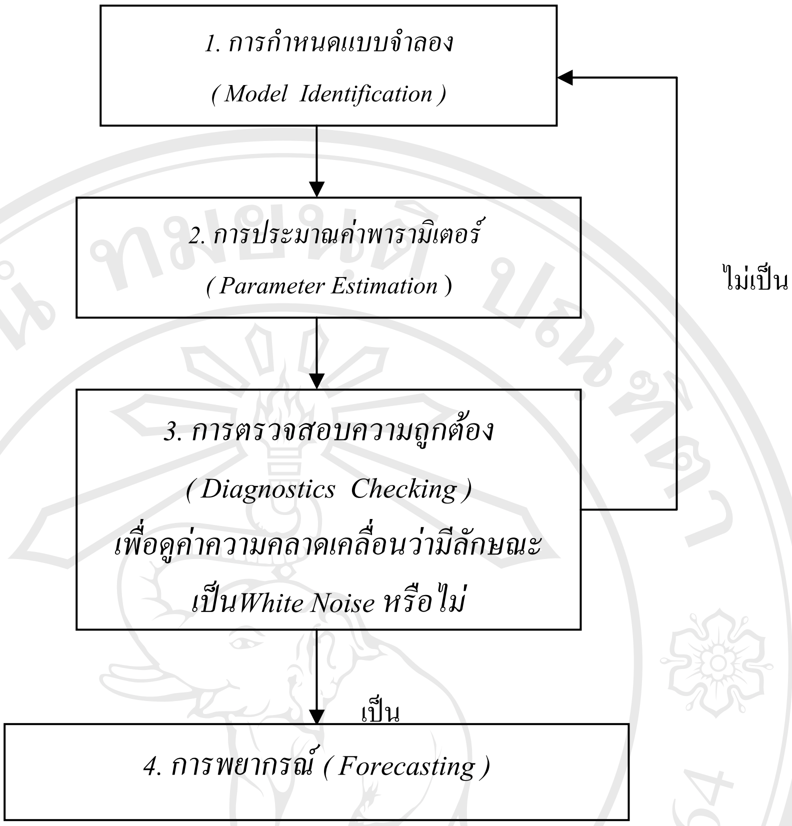
2) ค่าความแปรปรวน $V(X_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าแปรปรวนแต่ละส่วนย่อยไม่แตกต่างกันมากจะสรุปว่า $V(X_t)$ คงที่

3) พิจารณาแนวโน้มและปัจจัยฤดูกาล ด้วยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและปัจจัยฤดูกาลมักจะเห็นได้ชัดเจนได้จากรูปที่เรียกว่าคอเรลโรแกรม (Correlogram)

4) พิจารณาคอเรลโลแกรม ของสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ stationary จะมีค่าลดลงค่อนข้างเร็ว เมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้น ถ้าค่าสัมประสิทธ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้าจะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่าสัมประสิทธ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างสูงที่ k = L, 2L, 3L จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลของฤดูกาล และถ้าการเคลื่อนไหวของค่าคอเรลโลแกรมของค่าสัมประสิทธ์สหสัมพันธ์ (r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบภายใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วพบว่าอนุกรมเวลาที่ศึกษาไม่เป็น stationary ดังนั้น ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่เป็นstationary จะต้องแปลงอนุกรมเวลาให้เป็น stationary เสียก่อน โดยหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลได้อนุกรมเวลาที่เป็น stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลาเดิม โดยการหาลอการิซึม $Z = \ln(X_t)$ จนกว่าจะได้อนุกรมเวลาใหม่ ที่มีความแปรปรวนคงที่จากอนุกรมเวลาใหม่เป็น stationary series แล้วจึงทำตามขั้นตอนของ Box – Jenkins ต่อไป

ขั้นตอนของ Box-Jenkins ประกอบด้วย 4 ขั้นตอน ดังนี้ ขั้นตอนที่หนึ่ง คือการกำหนดรูปแบบจำลอง (model identification) ขั้นตอนที่สอง คือ การประมาณค่าพารามิเตอร์ (parameter estimation) ขั้นตอนที่สามคือ วิเคราะห์ความถูกต้อง (diagnostic checking) และขั้นตอนสุดท้ายคือการพยากรณ์ (forecasting) ตามลำดับดังรูปที่ 3.1



รูปที่ 3.1 แสดงขั้นตอนการพยากรณ์โดยวิธี Box and Jenkins

1) การกำหนดรูปแบบจำลอง (identification) ให้กับอนุกรมเวลาที่มีลักษณะนิ่ง (stationary series) เป็นการหารูปแบบ ARMA (p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยที่พิจารณาที่สหสัมพันธ์ (autocorrelation : ρ_k) โดยที่ ρ_k มีค่าเท่ากับ $-1< \rho_k < 1$ คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยพิจารณาเปรียบเทียบค่า autocorrelation (r_k) ของอนุกรมเวลาตัวอย่างกับค่า autocorrelation (ρ_k) ของอนุกรมเวลาของประชากรที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$\rho_k = \frac{\gamma_k}{\gamma_0} \qquad (3.11)$$

โดยที่ $\gamma_k = Cov(Z_t,Z_{t-k}) = E[(X_t-\mu)(X_{t-k}-\mu)]$

$\gamma_0 = Cov(Z_t,Z_{t-0}) = E[(X_t-\mu)^2]$

โดยที่ ρ_k เป็นการประมาณค่าของประชากร เป็นการประมาณค่าที่ต้องสุ่มมาจากตัวอย่างประชากร ซึ่งจะทำได้ง่ายและประหยัด ดังนั้นจึงได้กำหนด r_k เป็น autocorrelation ที่มาจากตัวอย่าง โดยมีสูตรดังนี้

$$r_k = \frac{C_k}{C_0} \tag{3.12}$$

โดยที่ $C_k = Cov(X_t, X_{t-k}) = E[(X_t - \overline{X})(X_{t-k} - \overline{X})]$

$C_0 = Cov(X_t, X_{t-0}) = E[(X_t - \overline{X})^2]$

ซึ่งสามารถพิสูจน์ เป็นสมการได้ดังนี้

$$\rho_k = \frac{\sum_{t=a}^{n-k}(X_{t-q})(X_{t+k-q})}{\sum_{t=a}^{n}(X_{t-q})^2} \tag{3.13}$$

โดยที่ $X_t = \sum_{t=a}^{n}(X_t)$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตามเนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวตามแปรอิสระ ที่เป็นค่าความล่าช้า (lag) ของตัวแปรตาม (autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (moving average) เนื่องจาก autocorrelation function (ACF) จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อนแต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง patial autocorrelation function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule-Walker (Pindyck and Rublinfeld, 1996) ดังนี้

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p\rho \tag{3.14}$$

ถ้า k มากกว่า p จะได้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \tag{3.15}$$

การกำหนดลำดับขั้น p , q ในแบบจำลอง (identifying the dependence order of model) ขั้นตอนคือการระบุว่าแบบจำลองนี้ควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใด และ moving average, q เท่าใด โดยพิจารณาจาก ACF และ PACF ซึ่งใช้ตาราง 1 ดังต่อไปนี้พิจารณา

ตารางที่ 3.1 แสดงการพิจารณา 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 สามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วหายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็น ค่าที่ p ของ AR(p) ยกตัวอย่างเช่นเมื่อพิจารณา คอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรมเกิดขึ้น 1 แท่ง แปลได้ว่าแบบจำลองควรมีลักษณะเป็น AR(1) สำหรับ MA(q) นั้นก็จะมี ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่งและหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลดเข้าหาแกนระนาบ สามารถสรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA (2) และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรจะเป็น ARMA (p,q) และเมื่อรวมกันกับการทดสอบความนิ่ง (stationary) ในขั้นตอนที่ 1 แล้ว จะสามารถหาค่าของ difference ได้ ซึ่งผลจากการ difference จำนวน d ครั้งนั้นก็จะได้แบบจำลอง ARIMA (p,d,q) แต่อย่างไรก็ตามหลักการดังกล่าวก็เป็นเพียงเครื่องช่วยการพิจารณาในระดับหนึ่งเท่านั้น

ดังนั้นเพื่อประเมินแบบจำลองว่าแบบจำลองใดมีความเหมาะสมที่จะใช้เป็นตัวแทนกลุ่มข้อมูลจริง สามารถพิจารณาได้จากค่าสถิติดังต่อไปนี้เพื่อประกอบในการตัดสินใจ

1. ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (root mean square error: RMSE) โดยจะเป็นการวัดค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองมีความแตกต่างกันมากน้อยเพียงใด หากค่า RMSE มีค่าเท่ากับศูนย์ จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงใดก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้มากเพียงนั้น สามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s - X_t^a)} \tag{3.16}$$

กำหนดให้ X_t^s คือค่าที่ประมาณจากแบบจำลอง

X_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

2. Theil's inequality coefficient โดยในหลักการเบื้องต้น พบว่าสมการที่ใช้นี้ยังคงมีหลักการที่คล้ายกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือค่าสถิตินี้จะมีค่าอยู่ระหว่าง 0 และ 1 ทั้งนี้หากค่า U มีค่าเท่ากับศูนย์ นั่นหมายความว่าค่าที่ได้จากการประมาณมีค่าเท่ากันพอดีกับค่าที่เป็นข้อมูลจริงแสดงถึงแบบจำลองที่ประมาณได้เป็นแบบจำลองที่เป็นตัวแทนข้อมูลได้อย่างดีที่สุด ในขณะที่ถ้า U มีค่าเท่ากันกับหนึ่ง แปลว่าแบบจำลองที่ประมาณได้เป็นแบบจำลองที่แย่ที่สุด ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่าU ที่น้อยๆ ดังจะพิจารณาได้จากสมการที่ (3.17)

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s - X_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^a)^2}} \tag{3.17}$$

กำหนดให้ X_t^s คือค่าที่ประมาณจากแบบจำลอง

X_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

3. R^2 คือการวัดค่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใด หากค่านี้เท่ากับ 1 ก็หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% ในทางกลับกัน หากค่านี้มีค่าเท่ากับ 0 ความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่อย่างไรก็ตาม พบว่าหากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมากๆ ก็จะทำให้ค่า R^2 มากขึ้นด้วย ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการที่ (3.17) ดังนั้นเพื่อปรับปรุงข้อจำกัดดังกล่าวข้างต้น จึงเกิดค่าสถิติใหม่ คือค่า adjusted R^2 ($\overline{R}^2$)ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการที่ (3.18)

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum x_i^2} \tag{3.18}$$

$$\overline{R}^2 = 1 - \frac{\sum \hat{u}_i^2/(n-k)}{\sum x_i^2/(n-1)} \quad (3.19)$$

4. Akaike information criterion (AIC) คือค่าสถิติที่ประยุกต์คล้ายกับ $\overline{R}^2$ โดยหากค่าสถิตินี้มีค่าน้อยเพียงใด ก็แปลว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น นอกจากนี้ ค่าสถิตินี้ยังเหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (lag length) ที่เหมาะสมอีกด้วย ซึ่งแสดงในสมการที่ (3.20)

$$AIC = \left(\frac{2k}{n}\right) + \log\left(\frac{\sum \hat{u}_i^2}{n}\right) \quad (3.20)$$

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

n คือ ค่าสังเกตทั้งหมด

5. Schwarz criterion (SC) เป็นวิธีการวัดปรับได้อย่างดี (goodness of fit) ของแบบจำลองได้ดีกว่า Akaike information criterion (AIC) ซึ่งเป็นวิธีที่ประยุกต์คล้ายกับ AIC สามารถเขียนในรูปสมการ SC ได้ดังสมการ (3.21)

$$SC = \log\left(\frac{\sum \hat{u}_i^2}{n}\right) + \left(\frac{2k \log n}{n}\right) \quad (3.21)$$

จากค่าสถิติข้างต้นทั้งหมดจะนำมาใช้ประกอบในการพิจารณาเลือกแบบจำลองARIMA (p , d , q) ที่เหมาะสมที่สุด โดยจะคัดเลือกแบบจำลองในขั้นตอนนี้ไว้ 3-4 แบบจำลองเพื่อการเลือกอีกครั้งในขั้นตอนการพยากรณ์ เพื่อที่จะเปรียบเทียบว่าแบบจำลองใดจะมีความสามารถในการพยากรณ์มากที่สุด

2) การประมาณค่าพารามิเตอร์ในรูปแบบ (estimation) คือการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (AR) และรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (MA) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (simple least square) แต่สามารถที่จะใช้วิธีการถดถอยแบบไม่เป็นเชิงเส้น (nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์ได้ หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3) การตรวจสอบแบบจำลอง (diagnostics checking) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลอง จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธียกตัวอย่าง เช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ในตัวเองของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลองโดยใช้การทดสอบของ Box-Pierce ซึ่งจะแสดงได้โดยใช้ Q-statistic ดังในสมการที่ (3.22)

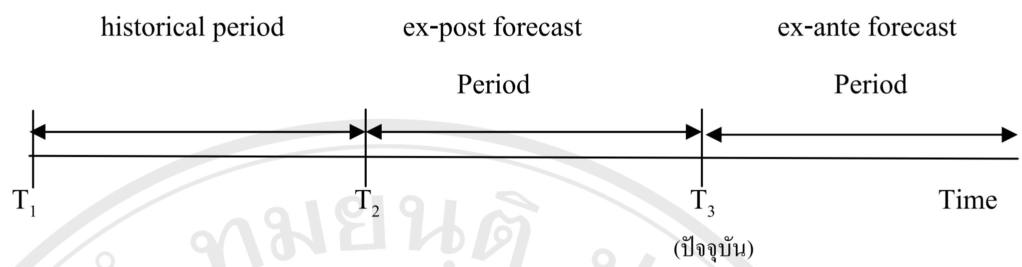
$$\text{Q-statistic} = n\sum_{k=1}^{m}\rho_k^2 \tag{3.22}$$

กำหนดให้ n คือ จำนวนของข้อมูล

m คือ ค่า lag length

จากสมการ(3.22) ค่า Q ของแบบจำลองมีการแจกแจงเป็นแบบ chi-square ที่มีดีกรีเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมุติฐานว่าง สมมติฐานว่าง คือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น white noise หรือ e_t มีการกระจายแบบปกติ (normal distribution) มีค่าเฉลี่ยเท่ากับ 0 และมีค่าความแปรปรวนเท่ากับ $\sigma^2 I[e_t \sim NID(0,\sigma^2 I)]$ แสดงว่า e_t มีลักษณะปราศจากอัตสหสัมพันธ์ (autocorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นปราศจากอัตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องกลับไปทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

4) การพยากรณ์ (forecasting) เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากการวิเคราะห์ความถูกต้องแล้ว ก็สามารถนำแบบจำลองมาใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลองโดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง historical forecast อันเป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา (T_2) การพยากรณ์ช่วง ex-post forecast คือการพยากรณ์โดยการตัดข้อมูลบางส่วนแล้วทำการพยากรณ์ โดยเปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ ซึ่งพิจารณาค่า root mean square error (RMSE) , Theil's inequality coefficient (TIC) และค่า Schwarz criterion (SC) เมื่อได้ค่าสถิติทั้งสามค่าแล้วจะพิจารณาค่าที่มีค่าน้อยที่สุด เพื่อเลือกรูปแบบจำลองที่ดีที่สุดแล้วจึงนำแบบจำลองนั้นมาทำการพยากรณ์แบบ ex-ante forecast กล่าวคือเป็นการพยากรณ์ไปข้างหน้าซึ่งยังไม่มีข้อมูลจริงเกิดขึ้น ดังรูปที่ 3.2



รูปที่ 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา : Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีการวิจัย

การพยากรณ์ราคาสัญญาล่วงหน้ายางพาราแผ่นรมควันชั้น 3 โดยวิธีอารีมา เป็นการนำเอาข้อมูลอนุกรมเวลามาหาแบบจำลองที่เหมาะสมที่สุด มีขั้นตอนดังต่อไปนี้

1. การทดสอบความนิ่งของข้อมูล (unit root test) เพื่อทำการทดสอบว่าราคาสัญญาล่วงหน้ามี ลักษณะนิ่งหรือไม่ โดยใช้สมการดังต่อไปนี้

$$\Delta X_{t-1} = \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.8}$$

$$\Delta X_{t-1} = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.9}$$

$$\Delta X_{t-1} = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.10}$$

2. การกำหนดแบบจำลอง ARIMA (p,d,q) โดยพิจารณาคอเรลโลแกรม autocorrelation function (ACF) และ partial autocorrelation function (PACF) เพื่อที่จะสามารถระบุว่าแบบจำลองควรจะมี autocorrelation function (ACF) , autoregressive (p) และ moving average (q) เท่าใด งานวิจัยนี้จึงสร้างแบบจำลองไว้หลายรูปแบบจำลองเพื่อหาแบบจำลองที่เหมาะสมที่สุด

3. การประมาณค่าพารามิเตอร์ คือการนำเอารูปแบบ ARIMA (p,d,q) ที่เลือกจากขั้นตอนการหารูปแบบที่เหมาะสมมาประมาณค่าพารามิเตอร์ โดยพิจารณาความมีนัยสำคัญทางสถิติด้วย T-statistic

4. การตรวจสอบความถูกต้องเพื่อทำการหาแบบจำลองที่เหมาะสม และประมาณค่าพารามิเตอร์แล้ว จึงทำการทดสอบแบบจำลอง โดยพิจารณาจากค่า Q – statistic , AIC และ SC

5. การพยากรณ์จะทำการพยากรณ์ราคาสัญญาล่วงหน้ายางพาราแผ่นรมควันชั้น 3 โดยแบ่งการพยากรณ์ออกเป็น 3 ช่วง ได้แก่ historical forecast , ex–post forecast และ ex–ante forecast