บทที่ 3

กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

3.1.1 การพยากรณ์ข้อมูลอนุกรมเวลา

การพยากรณ์ราคาสัญญาซื้อขายสัญญาล่วงหน้าสินค้าข้าวขาว 5% ประเภทข้อตกลงขนาดเล็ก ซึ่งกำหนดรูปแบบวิธีอารีมาให้กับอนุกรมเวลา (time series data) โดยวิธีการของ Box and Jenkins เป็นวิธีวิเคราะห์อนุกรมเวลาที่มีความถูกต้องและเหมาะสมกว่าวิธีการพยากรณ์ระยะสั้นอื่น ซึ่งสามารถพิจารณาการมีลักษณะนิ่งของข้อมูลอนุกรมเวลา ได้ดังนี้

$$\text{Mean} \quad : E(X_t) = \mu = \text{Constant} \tag{3.1}$$

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

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

เมื่อ X_t คือ ค่าตัวแปร x ที่เวลา t ใดๆ

ดังนั้นสามารถสรุปได้ว่า หากข้อมูลอนุกรมเวลานั้นมีลักษณะนิ่ง จะมีค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วม (จากสมการที่ 3.1, 3.2 และ3.3 ตามลำดับ) มีค่าที่คงที่ ณ ทุกๆ เวลาที่เปลี่ยนแปลงไป โดยใช้การทดสอบ unit root เพื่อพิจารณาความนิ่ง (stationary) หรือความไม่นิ่ง (non-stationary) ของข้อมูล

วิธีการของ Box and Jenkins เป็นการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลาโดยการใช้ค่า Autocorrelation Funtion (ADF) และค่า Partial Autocorrelation Funtion (PADF) เป็นหลักในการพิจารณา รูปแบบที่ใช้เลือกจะอยู่ในกลุ่มของรูปแบบ Integrated Aytoregressive-Moving Average order p and q อารีมา (p, d, q) ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากการสังเกตหรือการพยากรณ์ล่วงหน้าและความคลาดเคลื่อนของการพยากรณ์ล่วงหน้า โดยเป็นการรวมส่วนของรูปแบบ AR (p) และ MA (q) เข้าด้วยกัน โดยที่รูปแบบ AR (p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่า $X_{t-1}, X_{t-2}, X_{t-3},\ldots, X_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบMA (q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่าคลาดเคลื่อน $e_{t-1}, e_{t-2}, e_{t-3},\ldots, e_{t-q}$ หรือค่าคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่า

3.1.2 การทดสอบความนิ่งของข้อมูล (Unit Root Test)

การทดสอบ unit root (ทรงศักดิ์ ศรีบุญจิตต์ และ อารี วิบูลย์พงศ์, 2542) ได้อธิบายว่าสามารถทดสอบได้โดยใช้การทดสอบ DF (Dicky-Fuller test) (Dicky and Fuller, 1981) และการทดสอบ ADF (Augmented Dicky-Fuller test) (Said and Dickey 1984) ซึ่งสมมุติฐานว่าง (null hypothesis) ของการทดสอบ DF คือ $H_0 : \rho = 1$ จากสมการ (3.4)

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

ซึ่งเรียกว่า การทดสอบ unit root โดยถ้า $|\rho| < 1$ แสดงว่าข้อมูลX_t จะมีลักษณะนิ่ง หาก $\rho = 1$ แล้วข้อมูลX_t จะมีลักษณะไม่นิ่ง อย่างไรก็ตาม สามารถทำการทดสอบนี้ได้อีกทางหนึ่งซึ่งคล้ายกับสมการ (3.4) คือ

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

ซึ่ง $X_t = (1 + \theta) X_{t-1} + \varepsilon_t$ คือสมการที่ (3.4) นั่นเอง โดยที่ $\rho = (1 + \theta)$ หาก θ ในสมการ (3.5) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3.4) จะมีค่าน้อยกว่า 1 ดังนั้นสามารถจะสรุปได้ว่า การปฏิเสธ $H_0 : \theta = 0$ ซึ่งเป็นการยอมรับ $H_a : \theta < 0$ หมายความว่า $\rho < 1$ และข้อมูลX_t มี integration of order zero (Charemza and Deadman, 1992 : 131) นั่นคือ X_t มีลักษณะนิ่ง แต่ถ้าไม่สามารถปฏิเสธ $H_0 : \theta = 0$ ได้ ก็จะหมายความว่า X_t มีลักษณะไม่นิ่ง

ถ้า X_t เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) เราสามารถเขียนแบบจำลองได้ดังนี้

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

และถ้า X_t เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) และ มีแนวโน้มตามเวลาเชิงเส้น (linear time trend) เราสามารถเขียนแบบจำลองได้ดังนี้

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

โดยที่ t = แนวโน้มของเวลา ซึ่งก็จะทำการทดสอบ $H_0 : \theta = 0$ โดยมี $H_a : \theta < 0$ เช่นเดียวกับที่กล่าวมาข้างต้นโดยสรุปแล้ว Dickey and Fuller (1981) ได้พิจารณาสมการถดถอย 3 รูปแบบที่แตกต่างกันในการทดสอบว่ามี unit root หรือไม่ ซึ่ง 3 สมการดังกล่าว ได้แก่

Random walk	$\Delta X_t = X_t - X_{t-1} = \theta X_{t-1} + \varepsilon_t$
Random walk with drift	$\Delta X_t = X_t - X_{t-1} = \alpha + \theta X_{t-1} + \varepsilon_t$
Random walk with drift and trend	$\Delta X_t = X_t - X_{t-1} = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t$

โดยค่าพารามิเตอร์ที่อยู่ในความสนใจในทุกสมการ คือ θ นั่นคือ ถ้า θ = 0 : ข้อมูล X_t จะมี unit root โดยการเปรียบเทียบค่าสถิติ t (t-statistic) ที่คำนวณได้กับค่าที่เหมาะสมที่อยู่ในตาราง Dickey - Fuller (Enders, 1995 : 221) หรือกับค่าวิกฤติ MacKinnon (Gujarati, 2003 : 769)

อย่างไรก็ตามค่าวิกฤติ (critical values) จะไม่เปลี่ยนแปลง ถ้าสมการ (3.5), (3.6) และ (3.7) ถูกแทนที่โดยกระบวนการเชิงอัตถดถอย (autoregressive processes)

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

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

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

(Enders, 1995 : 221 และ Gujarati, 2003 : 720) จำนวนของ lagged difference terms ที่จะนำเข้ามารวมในสมการนั้น ควรจะมีมากพอที่จะทำให้พจน์ของค่าความคลาดเคลื่อน (error terms) มีลักษณะเป็น serially independent และเมื่อนำเอาการทดสอบ DF มาใช้กับสมการ (3.8), (3.9) และ(3.10) เราจะเรียกว่า การทดสอบ ADF ซึ่งมีการแจกแจงเชิงเส้นกำกับ (asymptotic distribution) เหมือนกับค่าสถิติ DF (Dickey - Fuller statistic)

ดังนั้นการหา lag length ที่เหมาะสมนั้น ก็คือ การใช้วิธี Serial Correlation LM test เพื่อหา lag length ที่มีค่า probability มากที่สุด จึงจะถือว่าค่า lag นั้นมีความเหมาะสม

3.1.3 แบบจำลองการพยากรณ์โดยวิธีการของ Box and Jenkins

การพยากรณ์อนุกรมเวลาโดยวิธีการของ Box and Jenkins ในรูปแบบอารีมา (p, d, q) ต้องพิจารณาว่าอนุกรมเวลาเป็น stationary series หรือไม่ (ทรงศิริ แต้สมบัติ, 2539) โดยพิจารณาจาก

1. ค่าเฉลี่ย $E(X_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่ จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมาก จะสรุปว่า $E(X_t)$ คงที่

2. ค่าความแปรปรวน $V(X_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่ จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าแปรปรวนแต่ละส่วนย่อยไม่แตกต่างกันมาก จะสรุปว่า $V(X_t)$ คงที่ความแปรปรวน $V(X_t)$ คงที่

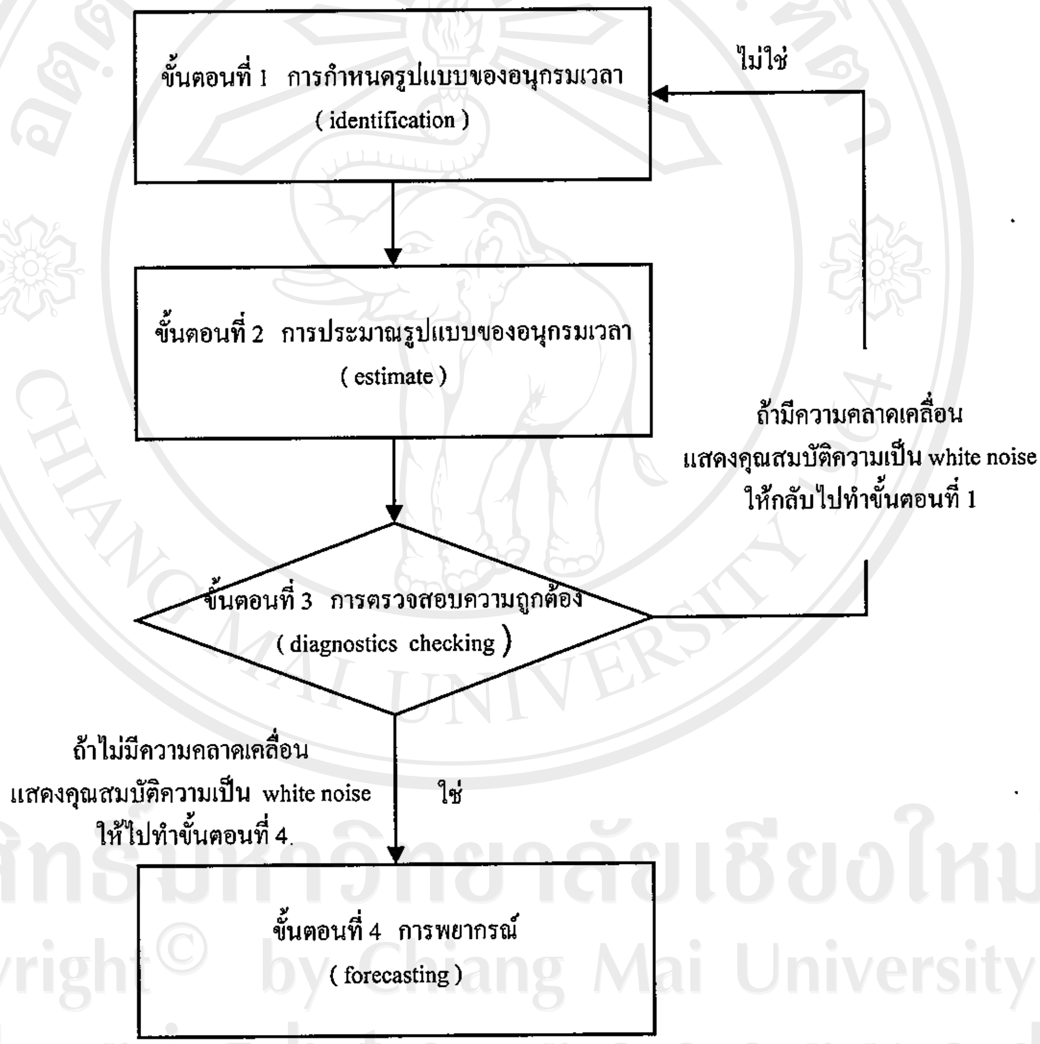
3. พิจารณาแนวโน้มและปัจจัยฤดูกาล ด้วยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและปัจจัยฤดูกาล มักจะเห็นได้ชัดเจนได้จากรูปที่เรียกว่า คอเรลโรแกรม (correlogram)

4. พิจารณาคอเรลโลแกรม ของสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง (r_k) กรณีที่อนุกรมเวลามีลักษณะนิ่ง ค่าคอเรลโลแกรมของสัมประสิทธิ์สหสัมพันธ์ (r_k) จะมีค่าลดลงค่อนข้างเร็ว เมื่อ k มีค่าเพิ่มขึ้นมาก หากค่าสัมประสิทธ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้า จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่าสัมประสิทธ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้าและมีค่าค่อนข้างสูงที่ k = L, 2L, 3L จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลของฤดูกาล และถ้าการเคลื่อนไหวของค่าคอเรลโลแกรมของสัมประสิทธ์สหสัมพันธ์ (r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบภายใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วพบว่าอนุกรมเวลาที่ศึกษามีลักษณะไม่นิ่ง ดังนั้นก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่มีลักษณะไม่นิ่ง จะต้องแปลงอนุกรมเวลาให้มีลักษณะนิ่งก่อน โดยหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่มีลักษณะนิ่ง ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลได้อนุกรมเวลาที่มีลักษณะนิ่ง แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลาเดิม โดยการหาลอการิซึม $Z = \ln(X_t)$ จนกว่าจะได้อนุกรมเวลาใหม่ ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่เป็น stationary series แล้วจึงทำตามขั้นตอนของ Box and Jenkins ต่อไป

ขั้นตอนการพยากรณ์ของ Box and Jenkins ประกอบด้วย 4 ขั้นตอน ดังนี้ ขั้นตอนที่หนึ่ง คือ การกำหนดรูปแบบจำลอง (identification) ขั้นตอนที่สอง คือ การประมาณค่าแบบจำลอง (estimation) ขั้นตอนที่สาม คือ วิเคราะห์ความถูกต้อง (diagnostic checking) และขั้นตอนสุดท้าย คือ การพยากรณ์ (forecasting) ตามลำดับ ดังรูป 3.1

รูป 3.1 ขั้นตอนการพยากรณ์ โดยวิธีการของ Box and Jenkins



3.1.3.1. การกำหนดรูปแบบจำลอง (identification) ให้กับอนุกรมเวลาที่เป็น stationary series เป็นการหารูปแบบ ARMA(p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลา โดยที่พิจารณาที่สหสมพันธ์ (autocorrelation : ρ_k คือ การวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา) มีค่าเท่ากับ $-1< \rho_k < 1$ โดยเปรียบเทียบค่าautocorrelation (r_k) ของอนุกรมเวลาตัวอย่าง กับค่า autocorrelation (ρ_k) ของอนุกรมเวลาของประชากรที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ดังสมการ (3.11)

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{3.11}$$

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

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

โดยที่ ρ_k เป็นการประมาณค่าของประชากร เป็นการประมาณค่าที่ต้องสุ่มมาจากตัวอย่างประชากร ซึ่งจะทำได้ง่ายและประหยัด ดังนั้นจึงได้กำหนด r_k เป็น autocorrelation ที่มาจากตัวอย่าง ดังสมการ (3.12)

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

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

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

ซึ่งสามารถพิสูจน์ เป็นสมการ (3.13)

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

เมื่อ $X_t = \sum_{t=a}^{n}(X_t)$

q คือ จำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตามเนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวแปรอิสระ ที่เป็นค่าความล่าช้า (lag) ของตัวแปรตาม (autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (moving average) เนื่องจาก ACF จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อน แต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง PACF จะใช้วัดความสัมพันธ์ดังกล่าว ซึ่งสามารถพิจารณาได้จากสมการ Yule-Walker (Pindyck and Rublinfeld, 1997) 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) ขั้นตอนคือ การระบุว่าแบบจำลองนี้ควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใด และ moving average , q เท่าใด โดยใช้ตาราง 3.1 พิจารณาจากค่า ACF และ PACF

ตาราง 3.1 การพิจารณาค่า Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF)

ชนิดของแบบจำลอง	รูปแบบของ ACF	รูปแบบของ PACF
AR(p)	ลู่โค้งเข้าหาแกน (tails off)	เกิดค่าที่ชัดเจนเพียง p ค่า แล้วหายไป (cut off after lag p)
MA(q)	เกิดค่าที่ชัดเจนเพียง q ค่า แล้วหายไป (cut off after lag q)	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) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรมเกิดขึ้น 1 แท่ง แปลได้ว่าแบบจำลองควรมีลักษณะเป็น AR(1) สำหรับ MA(q) นั้นก็จะมี ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลดเข้าหาแกนระนาบ สามารถสรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(2) และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรจะเป็น ARMA(p,q) และเมื่อรวมกันกับการทดสอบความนิ่ง ในขั้นตอนที่ 1 แล้ว จะสามารถหาค่าของ difference ได้ ซึ่งผลจากการ difference จำนวน d ครั้ง ก็จะได้แบบจำลอง ARIMA(p,d,q) แต่อย่างไรก็ตามหลักการดังกล่าวก็เป็นเพียงเครื่องช่วยการพิจารณาในระดับหนึ่งเท่านั้น

ดังนั้นการประเมินแบบจำลองที่มีความเหมาะสม เพื่อนำมาใช้เป็นตัวแทนของกลุ่มข้อมูลจริง ประกอบในการตัดสินใจสามารถพิจารณาได้จากค่าสถิติดังต่อไปนี้

1. ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (root mean square error : RMSE) โดยจะเป็นการวัดค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองมีความแตกต่างกันมากน้อยเพียงใด ถ้าค่า RMSE มีค่าเท่ากับ 0 จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นค่า RMSE มีค่าน้อยเพียงใด ก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากเพียงนั้น โดยสมการค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ดังสมการ (3.12)

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

เมื่อ X_t^s คือ ค่าที่ประมาณจากแบบจำลอง

X_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

2. Theil's Inequality Coefficient โดยในหลักการเบื้องต้น พบว่าสมการที่ใช้นี้ยังคงมีหลักการที่คล้ายคลึงกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือ ค่าสถิตินี้จะมีค่าอยู่ระหว่าง 0 และ 1 หากค่า U มีค่าเท่ากับ 0 ก็หมายความว่าค่าที่ได้จากการประมาณมีค่าเท่ากันพอดีกับค่าที่เป็นข้อมูลจริง แสดงถึงแบบจำลองที่ประมาณได้เป็นแบบจำลองที่เป็นตัวแทนข้อมูลได้อย่างดีที่สุด หากค่า U มีค่าเท่ากับ 1 หมายความว่าแบบจำลองที่ประมาณได้เป็นตัวแทนข้อมูลที่แย่ที่สุด ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่า U ที่น้อยๆ พิจารณาจากสมการ (3.17)

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

เมื่อ X_t^s คือ ค่าที่ประมาณจากแบบจำลอง

X_t^a คือ ค่าข้อมูลจริง

T คือ จำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

3. R^2 คือ การวัดค่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใด ถ้าค่า R^2 เท่ากับ 1 หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% ในทางกลับกัน หากค่า R^2 มีค่าเท่ากับ 0 หมายความว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่ถ้ามีการเพิ่มตัวแปรอิสระเข้าไปในสมการมาก ก็จะทำให้ค่า R^2 มากขึ้นด้วย จึงเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการ จากสมการที่ (3.17) ดังนั้นเพื่อปรับปรุงข้อจำกัดดังกล่าว จึงเกิดค่าสถิติใหม่ คือค่า Adjusted $R^2(\overline{R}^2)$ ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังสมการที่ (3.18) และ(3.19)

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

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

4. Akaike Information Criterion (AIC) คือ ค่าสถิติที่ประยุกต์คล้ายกับ $\overline{R}^2$ แต่ใช้รูปแบบการใส่ค่าลอกาลิทึมฐานธรรมชาติ (natural logarithm) ถ้าค่า $\overline{R}^2$ มีค่าน้อยเพียงใด ก็หมายความว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ค่า $\overline{R}^2$ เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (lag length) ที่เหมาะสมอีกด้วย ดังสมการที่ (3.20)

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

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

n คือ ค่าสังเกตทั้งหมด

5. Schwarz Criterion (SC) เป็นวิธีการวัดปรับค่าได้อย่างดี (goodness of fit) ของแบบจำลองได้ดีกว่า AIC เพราะ SC เป็นวิธีประยุกต์ที่คล้ายกับ AIC ดังสมการ (3.21)

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

จากค่าสถิติข้างต้นทั้งหมดจะนำมาใช้ประกอบในการพิจารณาเลือกแบบจำลอง ARIMA(p, d, q) ที่เหมาะสมที่สุด โดยจะคัดเลือกแบบจำลองในขั้นตอนนี้ประมาณ 3 ถึง 4 แบบจำลอง เพื่อนำมาเลือกอีกครั้งในขั้นตอนของการพยากรณ์ที่จะทำการเปรียบเทียบว่าแบบจำลองใดจะมีความสามารถในการพยากรณ์มากที่สุด

3.1.3.2. การประมาณค่าแบบจำลอง (estimation) คือ การประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (autoregressive : AR : p) และรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (moving average : MA : q) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (simple least square) แต่สามารถที่จะใช้วิธีการถดถอยแบบไม่เป็นเชิงเส้น (nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์ได้ หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3.1.3.3. การตรวจสอบแบบจำลอง (diagnostic checking) เมื่อกำหนดรูปแบบและประมาณค่าแบบจำลองแล้ว จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบสามารถจะทำได้หลายวิธี ยกตัวอย่างเช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลองโดยใช้การทดสอบของ Box and Pierce ซึ่งใช้ Q-statistic ดังสมการ (3.22)

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

เมื่อ n คือ จำนวนของข้อมูล

m คือ ค่า lag length

จากสมการ (3.22) ค่า Q-statistic ของแบบจำลองมีการแจกแจงแบบ chi-square ที่มีดีกรีเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมุติฐานว่าง : H_0 คือ ค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น 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 (non - autocorrelation) หากตรวจสอบพบว่าแบบจำลองนั้นปราศจากอัตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่ถ้าแบบจำลองนั้นไม่เหมาะสมต้องกลับไปทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

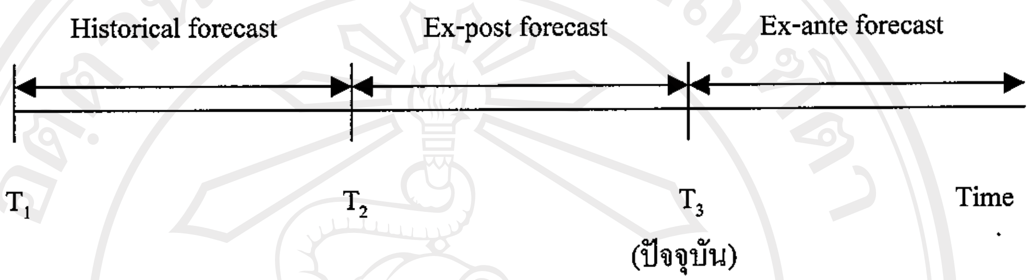
3.1.3.4. การพยากรณ์ (forecasting) เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากการวิเคราะห์ความถูกต้องแล้ว ก็สามารถนำแบบจำลองใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลองโดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ

1) ช่วง Historical forecast เป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา (T_2)
2) ช่วง Ex-post forecast เป็นการพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์ เพื่อเปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า Root Mean Square Error ค่า Theil's Inequality Coefficient และค่า Akaike information criterion (AIC) จะพิจารณาค่าสถิติทั้ง 3 ค่า ที่มี

ค่าน้อยที่สุด ซึ่งได้จากการทำการพยากรณ์เมื่อเลือกรูปแบบจำลองที่ดีที่สุด แล้วจึงนำแบบจำลองนั้นมาทำการพยากรณ์

3) ช่วง Ex-ante forecast เป็นการพยากรณ์ข้อมูลไปข้างหน้า

รูป 3.2 ช่วงเวลาที่ใช้ในการพยากรณ์



ที่มา : Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีการวิจัย

การพยากรณ์ราคาสัญญาล่วงหน้าของข้าวขาว 5% ประเภทข้อตกลงขนาดเล็ก โดยวิธีอารีมา เป็นการนำเอาข้อมูลอนุกรมเวลามาหาแบบจำลองที่เหมาะสมที่สุด มีขั้นตอนดังต่อไปนี้

3.2.1 การทดสอบความนิ่งของข้อมูล (unit root test) เพื่อทำการทดสอบว่าราคาสัญญาล่วงหน้า ว่ามีลักษณะนิ่งหรือไม่

3.2.2 การกำหนดแบบจำลอง ARIMA(p,d,q) โดยพิจารณาคอเรลโลแกรม ค่า autocorrelation function (ACF) และ partial autocorrelation function (PACF) เพื่อที่จะสามารถระบุว่าแบบจำลองควรจะมี ACF, autoregressive (p) และ moving average (q) เท่าใด โดยการสร้างหลายๆ แบบจำลองไว้หลายรูปแบบจำลอง เพื่อหาแบบจำลองที่เหมาะสมที่สุด

3.2.3 การประมาณค่าแบบจำลอง (estimation) เป็นการนำเอารูปแบบ ARIMA(p,d,q) จากแบบจำลองที่เหมาะสม มาประมาณค่าพารามิเตอร์ ประกอบการพิจารณาความมีนัยสำคัญทางสถิติด้วยค่า t-statistic

3.2.4 การตรวจสอบความถูกต้อง (diagnostic checking) เพื่อทำการหาแบบจำลองที่เหมาะสม และประมาณค่าแบบจำลองแล้ว จึงทำการทดสอบแบบจำลอง ประกอบการพิจารณากับค่า Q – statistic , ค่า Akaike information criterion และค่า Schwarz criterion

3.2.5 การพยากรณ์ (forecasting) ราคาสัญญาล่วงหน้าของข้าวขาว 5% ประเภทข้อตกลงขนาดเล็ก แบ่งออกเป็น 3 ช่วง ได้แก่ Historical forecast, Ex-post forecast และ Ex-ante forecast ดังรูป 3.2