# บทที่ 3
# แนวความคิดและระเบียบวิธีวิจัย

## 3.1 แนวความคิดที่ใช้ในการศึกษา

การพยากรณ์อนุกรมเวลาสำหรับแบบจำลอง ARIMA(p,d,q) นั้นจะอิงใช้ทฤษฎีของ Box-Jenkins (1976) เป็นเครื่องมือในการศึกษาครั้งนี้ โดยในเบื้องต้นต้องพิจารณาว่าอนุกรมเวลานั้นมีคุณสมบัติของอนุกรมเวลาที่มีลักษณะนิ่ง (Stationary) หรือไม่ (Pindyck and Rubinfeld, 1997) โดยการพิจารณาว่าจะเป็นอนุกรมเวลาที่มีลักษณะที่นิ่ง (Stationary) หรือไม่ สามารถพิจารณาได้จากค่าเฉลี่ย ความแปรปรวนและค่าความแปรปรวนร่วมของข้อมูลอนุกรมเวลาดังต่อไปนี้

$$E(y_t) = \mu_y \tag{3.1}$$

กล่าวคือ หากข้อมูลอนุกรมมีลักษณะที่นิ่ง (Stationary) แล้ว ณ ทุกๆ ค่าที่เวลา t ใดๆ จะมีค่าเฉลี่ย $E(y_t)$ คงที่ หรือเท่ากับ $\mu_y$

$$E[(y_t - \mu_y)^2] = \sigma_y^2 \tag{3.2}$$

หากข้อมูลอนุกรมมีลักษณะนิ่ง (Stationary) นอกจากค่าเฉลี่ย $E(y_t)$ คงที่แล้ว ค่าความแปรปรวน (Variance) คงที่ สำหรับทุกๆ ค่าของเวลาที่ t ใดๆ



$$Cov(y_t, y_{t+k}) = \gamma_k \tag{3.3}$$

หากข้อมูลอนุกรมมีลักษณะนิ่ง (Stationary) แล้วอีกเงื่อนไขหนึ่งที่มีความสำคัญคือ ค่าความแปรปรวนร่วม (Covariance) จะต้องมีค่าคงที่ ณ เวลาที่ t ใดๆ

ดังนั้นสามารถสรุปได้ว่า หากข้อมูลอนุกรมเวลานั้นมีลักษณะที่นิ่ง (Stationary) จะมีค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วม (จากสมการที่ 3.1, 3.2 และ 3.3) มีค่าที่

คงที่ ณ ทุกๆ เวลาที่เปลี่ยนแปลงไป ซึ่งสามารถทดสอบข้อมูลว่ามีลักษณะที่นิ่งหรือไม่จากการทดสอบ Unit Root

### 3.1.1. การทดสอบความนิ่งของข้อมูล (Unit Root Test)

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

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

ซึ่งเรียกว่าการทดสอบ Unit root โดยถ้า $|\rho| < 1$ ข้อมูล $x_t$ จะมีลักษณะนิ่ง (stationary) และถ้า $\rho = 1$ แล้วข้อมูล $x_t$ จะมีลักษณะไม่นิ่ง (Nonstationary) อย่างไรก็ตามการทดสอบนี้สามารถทำได้อีกทางหนึ่งซึ่งเหมือนกับสมการ (3.4) กล่าวคือ

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

นั่นก็คือ $X_t = (1+\theta)X_{t-1} + \varepsilon_t$ ซึ่งเทียบสมการที่ (3.4) ดังนี้ คือ $\rho = (1+\theta)$ ถ้าหาก $\theta$ ในสมการ (3.5) มีค่าเป็นลบ จะได้ว่า $\rho$ ในสมการ (3.4) จะมีค่าน้อยกว่า 1 ดังนั้นสามารถจะสรุปได้ว่า การปฏิเสธ $H_0 : \theta = 0$ ซึ่งเป็นการยอมรับ $H_a : \theta < 0$ หมายความว่า $\rho < 1$ และ $x_t$ มี integration of order zero (Charemza and Deadman, 1992) นั่นคือ $x_t$ มีลักษณะนิ่ง (Stationary) และถ้าเราไม่สามารถปฏิเสธ $H_0 : \theta = 0$ ได้ ก็จะหมายความว่า $x_t$ มีลักษณะไม่นิ่ง (Nonstationary)

ถ้า $x_t$ เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) เราสามารถจะเขียน แบบจำลองได้ดังนี้

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

และถ้า $x_t$ เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) และมีแนวโน้มตามเวลาเชิงเส้น (linear time trend) เราสามารถจะเขียนแบบจำลองได้ดังนี้

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

โดยที่ t = แนวโน้มของเวลา ซึ่งก็จะทำการทดสอบ $H_0 : \theta = 0$ โดยมี $H_a : \theta < 0$ เช่นเดียวกับที่กล่าวมาข้างต้น โดยสรุปแล้ว Dickey and Fuller (1979) ได้พิจารณาสมการถดถอย 3 รูปแบบที่แตกต่างกันในการทดสอบว่ามี unit root หรือไม่ ซึ่ง 3 สมการดังกล่าว ได้แก่

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

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

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

โดยตัวพารามิเตอร์ที่อยู่ในความสนใจในทุกสมการ คือ θ นั่นคือ ถ้า $\theta = 0$; ข้อมูล $x_t$ จะมี unit root โดยการเปรียบเทียบค่าสถิติ DF (DF-test) ที่คำนวณได้กับค่าวิกฤติที่เหมาะสมที่อยู่ในตาราง Dickey-Fuller (Dickey–Fuller tables) (Enders, 1995) หรือกับ ค่าวิกฤติ MacKinnon (MacKinnon critical values) (Gujarati, 2003)

อย่างไรก็ตามค่าวิกฤติ (Critical values) จะไม่เปลี่ยนแปลง ถ้าสมการ (3.5), (3.6), (3.7) ถูกแทนที่โดยกระบวนการถดถอยในตัวเอง (Autoregressive processes)



$$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \qquad (3.8)$$

$$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-1} + \varepsilon_t \qquad (3.9)$$

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \qquad (3.10)$$

จำนวนของ Lagged difference terms ที่จะนำเข้ามารวมในสมการนั้นควรจะมีมากพอที่จะทำให้พจน์ค่าความคลาดเคลื่อน (Error terms) มีลักษณะเป็น Serially independent

และเมื่อนำเอาการทดสอบ DF (DF-test) มาใช้กับสมการ (3.8) – (3.10) เราจะเรียกว่าการทดสอบ ADF (ADF-test) ค่าสถิติทดสอบ ADF (ADF test) มีการแจกแจงเชิงเส้นกำกับ (Asymptotic distribution) เหมือนกับสถิติ DF (DF-test) ดังนั้นก็สามารถใช้ค่าวิกฤติ (critical values) แบบเดียวกัน (Gujarati, 2003)

สำหรับกรณีของการหา lag length ที่เหมาะสมนั้น (Enders, 1995) ได้เสนอแนะว่าวิธีหนึ่งในการหา lag length ก็คือ เริ่มต้นด้วยการให้มี lag length ที่ยาวมากพอ และก็ลดขนาดของ lag length ลงโดยใช้ค่าสถิติทดสอบ t (t-test) หรือค่าสถิติทดสอบ F (F-test) สมมุติว่าเราใช้ lag length เท่ากับ n ถ้าสถิติ t (t – statistic) ของ lag n นั้นๆ ไม่มีนัยสำคัญ ณ ค่าวิกฤติ (critical value) ก็จะต้องทำการประมาณค่าการถดถอยใหม่ โดยใช้ lag length n-1 ทำอย่างนี้ เรื่อยไปจนกระทั่ง lag นั้นมีค่าแตกต่างไปจากศูนย์ อย่างมีนัยสำคัญ

เมื่อพิจารณาการตรวจสอบแล้วว่า อนุกรมเวลาที่ศึกษาไม่เป็น Stationary จะต้องแปลงให้เป็น Stationary เสียก่อน ภายหลังจากนั้นก็ทำตามวิธี Box-Jenkins ต่อไป

### 3.1.2. วิธีการพยากรณ์ด้วยวิธี Box-Jenkins

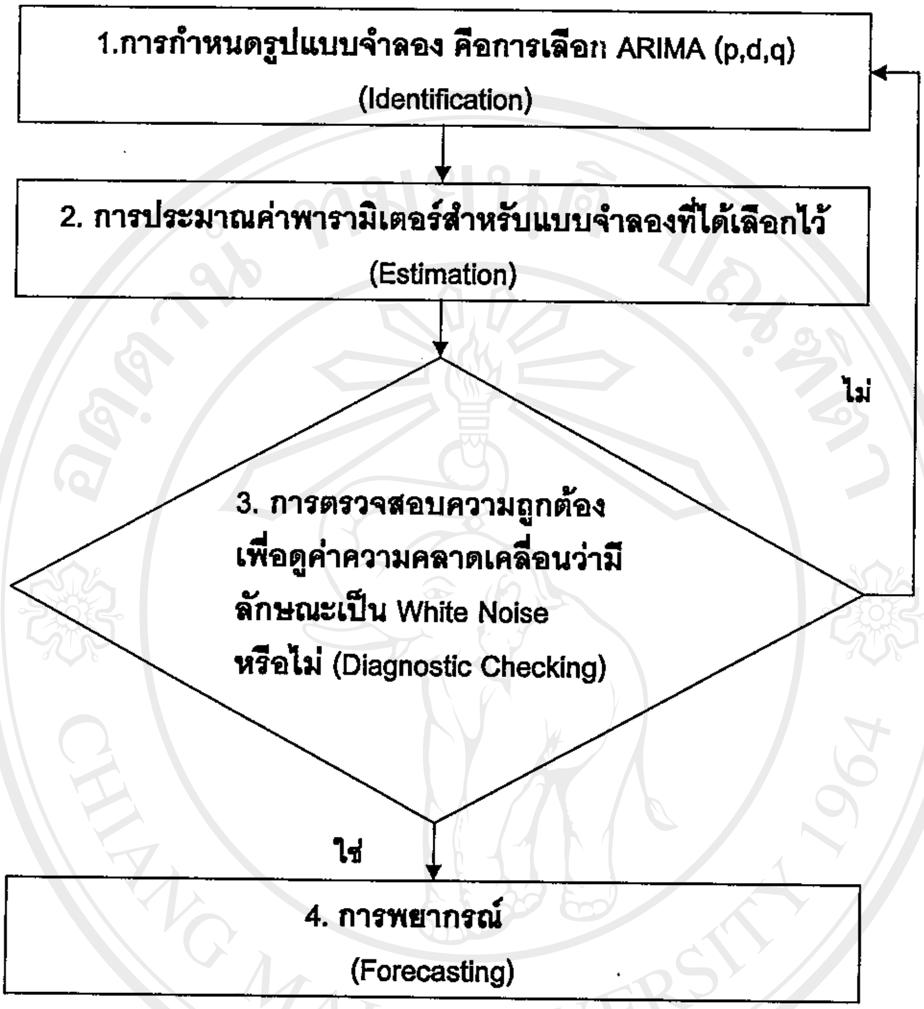
แบบจำลองที่ใช้ในการพยากรณ์คือ แบบจำลองอารีมา ARIMA (p,d,q) ซึ่งจะประกอบไปด้วย 3 ส่วนดังนี้ การถดถอยด้วยตัวเอง (Autoregressive; AR :p) การมีอันดับ (Integrated; I: d) และการเคลื่อนที่ของความคลาดเคลื่อน (Moving Average; MA: q) สำหรับรูปแบบทั่วไปของอารีมา (ARIMA) สามารถเขียนเป็นความสัมพันธ์ได้ดังนี้

$$\Delta^d y_t = \delta + \phi_1 \Delta^d y_{t-1} + \phi_2 \Delta^d y_{t-2} + \ldots + \phi_p \Delta^d y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_t - \ldots - \theta_q \varepsilon_{t-q} \qquad (3.11)$$

จากสมการที่ (3.11) ประกอบด้วยรูปแบบ AR(p) กล่าวคือ ค่าสังเกต $Y_t$ ขึ้นอยู่กับค่า $Y_{t-1}$ …… $Y_{t-p}$ หรือกล่าวว่าขึ้นอยู่กับค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า อีกทั้งรูปแบบ MA(q) หรือกล่าวว่า ค่าสังเกต $Y_t$ ขึ้นอยู่กับค่าความคลาดเคลื่อน $\varepsilon_{t-1}$ …. $\varepsilon_{t-q}$ กล่าวอีกนัยหนึ่งคือ ขึ้นอยู่กับค่าความคลาดเคลื่อนก่อนหน้า q ค่า และรูปแบบ I(d) เกิดจากการหาผลต่าง (Difference) ของอนุกรมเวลา

ขั้นตอนของ Box-Jenkins ประกอบด้วย 4 ขั้นตอน ประกอบด้วย ขั้นตอนที่หนึ่งคือการกำหนดรูปแบบจำลอง (Identification) ขั้นตอนที่สองคือการประมาณค่า (Estimation) ขั้นตอนที่สามคือการวิเคราะห์ความถูกต้อง (Diagnostic Checking) และขั้นตอนสุดท้ายคือการพยากรณ์ (Forecasting) ตามลำดับ ดังจะพิจารณาจากรูปที่ 3.1

รูปที่ 3.1 แสดงขั้นตอนของ Box and Jenkins




ที่มา: Gujarati (2003)

**1. การกำหนดแบบจำลอง (Identification)** ให้กับอนุกรมเวลาที่มีลักษณะนิ่ง (Stationary Series) เป็นการหารูปแบบ ARMA (p,q) ที่คาดว่าเหมาะสมให้กับอนุกรมเวลาโดยพิจารณาที่สหสัมพันธ์ (Autocorrelation: $\rho_k$) โดยที่ $\rho_k$ มีค่าอยู่ในช่วง [-1,1] คือการวัดความสัมพันธ์ของข้อมูลของแต่ละช่วงเวลา โดยมีช่วงที่ย้อนหลังไป k หน่วยเวลา สำหรับวิธีการหาค่า $\rho_k$ สามารถหาได้จากความแปรปรวนร่วม(Covariance) ของค่ากลุ่มตัวอย่างที่ Lag k (Sample Covariance) เทียบกับค่าแปรปรวน (Variance) ของกลุ่มตัวอย่าง (Sample Variance) ซึ่งจะพิจารณาได้จากสมการดังต่อไปนี้

$$\rho_k = \frac{\text{Covariance at lag k}}{\text{Variance}} \tag{3.12}$$

$$\hat{\rho}_k = \frac{\sum(y_t - \hat{y})(y_{t+k} - \hat{y})}{\sum(y_t - \bar{y})^2} \quad (3.13)$$

อย่างไรก็ตามเนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวแปรอิสระที่เป็นค่าความล่าช้าของ (Lag) ของตัวแปรตาม (Autoregressive) และที่เป็นสหสัมพันธ์ของค่าความคลาดเคลื่อน (Moving average) โดยทั้งนี้เนื่องจาก Autocorrelation Function (ACF) จะใช้ในการอธิบายสหสัมพันธ์ของของค่าความคลาดเคลื่อน แต่ไม่สามารถใช้อธิบายสหสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง Patial Autocorrelation Function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule-walker (Pindyck and Rubinfeld, 1996) ดังนี้

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + .... + \phi_p \quad (3.14)$$

ถ้า k มากกว่า p จะได้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + .... + \phi_p\rho_{k-p} \quad (3.15)$$

จากข้างต้นเมื่อทราบถึงคอเรลโลแกรม (Correlogram) ของ ACF และ PACF ก็นำมาหารูปแบบที่มีความเป็นไปได้โดยสามารถพิจารณาความเป็นไปได้ของแบบจำลองจากตารางดัง 3.1 ต่อไปนี้

ตารางที่ 3.1 ตารางแสดงการพิจารณา 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)

จากตารางที่ 3.1 จะสามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็น ค่าที่ p ของ AR(p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรม เกิดขึ้น 1 แท่ง แปลได้ว่าแบบจำลองอาจมีลักษณะเป็น AR(1) ในทางตรงกันข้าม สำหรับ MA(q) นั้นก็จะมี ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น ยกตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ สามารถสรุปได้ว่าแบบจำลองควรมีลักษณะเป็น MA(2) และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองควรจะเป็น ARMA(p,q) แต่อย่างไรก็ตามหลักการดังกล่าวก็เป็นเพียงเครื่องช่วยการพิจารณาในระดับหนึ่งเท่านั้น เนื่องจากวิธีการดังกล่าวมีลักษณะที่เป็นศิลป์ (Art) มากกว่าศาสตร์ (Science) ดังนั้นเพื่อประเมินแบบจำลองว่าแบบจำลองใดมีความเหมาะสมที่จะใช้เป็นตัวแทนกลุ่มข้อมูลจริง สามารถพิจารณาได้จากค่าสถิติดังต่อไปนี้เพื่อประกอบในการตัดสินใจ

- **ค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root–mean–square Error: RMS)** โดยจะเป็นการวัดว่าค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองมีความแตกต่างกันมากน้อยเพียงใด ซึ่งหากค่า RMSE มีค่าเท่ากับศูนย์ จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงไร ก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากเพียงนั้น (Pindyck and Rubinfeld, 1996) ซึ่งค่าสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) แสดงได้ดังนี้

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)} \qquad (3.16)$$

กำหนดให้ $Y_t^s$ คือค่าที่ประมาณจากแบบจำลอง

$Y_t^a$ คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

- **ค่า Theil's inequality coefficient (U)** โดยในหลักการเบื้องต้น พบว่าสมการที่ใช้นี้ยังคงมีหลักการที่คล้ายกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือ ค่าสถิตินี้จะมีค่าอยู่ระหว่าง 0 และ 1 ทั้งนี้หากค่า U มีค่าเท่ากับศูนย์ นั่นก็หมายความว่าค่าที่ได้จากการประมาณ

มีค่าเท่ากันพอดีกับค่าที่เป็นข้อมูลจริง แสดงถึงแบบจำลองที่ประมาณได้เป็นแบบจำลองที่เป็นตัวแทนข้อมูลจริงได้อย่างดีที่สุด ในขณะที่ถ้า U มีค่าเท่ากันกับหนึ่ง แปลว่าแบบจำลองที่ประมาณได้เป็นแบบจำลองที่แย่ที่สุด (Pindyck and Rubinfeld, 1996) ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่า U ที่น้อยๆ ดังจะพิจารณาได้จากสมการที่ (3.17)

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

กำหนดให้ $Y_t^s$ คือค่าที่ประมาณจากแบบจำลอง

$Y_t^a$ คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

- **$R^2$** คือการวัดว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใด หากค่านี้กับ 1 นั่นก็หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% ในทางกลับกันหากค่านี้มีค่าเท่ากับ 0 แปลความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่อย่างไรก็ตามพบว่าหากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมากๆ ก็จะทำให้ค่า $R^2$ มากขึ้นด้วย ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการที่ (3.18)

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

- **Adjusted $R^2$ ($\bar{R}^2$)** ซึ่งจะมีชั่งน้ำหนัก (Weighted) กันระหว่างตัวแปรที่เพิ่มเข้าไป กับค่า $R^2$ ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการที่ (3.19)

$$\bar{R}^2=1-\frac{\sum \hat{u}_i^2/(n-k)}{\sum y_i^2/(n-1)} \tag{3.19}$$

- **Akaike Information Criterion (AIC)** คือค่าสถิติที่ประยุกต์คล้ายกับ Adjusted $R^2$ แต่ใช้รูปแบบการใส่ค่าลอกกาลิทึม ฐานธรรมชาติ (Natural logarithm) โดยหาก

ค่าสถิตินี้มีค่าน้อยเพียงใด นั่นก็แปลว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (Lag Length) ที่เหมาะสมอีกด้วย ซึ่งแสดงในสมการที่ (3.20)

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

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

n คือค่าสังเกตทั้งหมด

จากค่าสถิติข้างต้นทั้งหมดจะนำมาใช้ประกอบในการพิจารณาเลือกแบบจำลอง ARIMA (p,d,q) ที่เหมาะสมที่สุด โดยจะคัดเลือกแบบจำลองในขั้นตอนนี้ไว้ 3-4 แบบจำลองเพื่อทำการเลือกอีกครั้งในขั้นตอนการพยากรณ์ เพื่อที่จะเปรียบเทียบว่าแบบจำลองใดจะมีความสามารถในการพยากรณ์มากที่สุด

2. **การประมาณค่าพารามิเตอร์ (Estimation)** คือการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (AR) และ รูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (MA) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (Simple least square) แต่สามารถที่จะใช้วิธีการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์ได้ หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3. **การวิเคราะห์ความถูกต้อง (Diagnostics checking)** เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลอง จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่การตรวจสอบจะทำได้หลายวิธียกตัวอย่างเช่น การพิจารณาคอเรลโลแกรมของสหสัมพันธ์ในตัวเอง (Autocorrelaion) ของกลุ่มตัวอย่าง ($\rho_k$) แต่อย่างไรก็ตาม (Gujarati, 2003) ได้เสนอ การทดสอบวิเคราะห์ความเหมาะสมของแบบจำลอง โดยการใช้ทดสอบของ Box และ Pierce ซึ่งจะแสดงได้โดยใช้ Q Statistic ดังในสมการที่ 3.21

$$Q = n\sum_{k=1}^{m}\hat{\rho}_k^2 \tag{3.21}$$

กำหนดให้: n คือจำนวนของข้อมูล

m คือค่า Lag Length

จากสมการที่ 3.21 โดยมีการกำหนดค่า Q-Statistic เพื่อเป็นการทดสอบว่า สหสัมพันธ์ในตัวเองของค่าความคลาดเคลื่อนจากการประมาณ (Estimated Residuals) ทุกช่วงเวลาที่ห่างกัน k มีความเป็นอิสระหรือไม่ จากสมมุติฐานดังต่อไปนี้

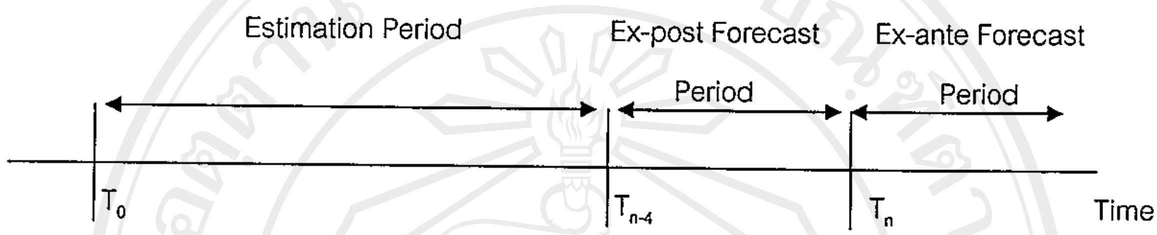
$$H_0: \rho_1(\hat{\varepsilon}_t) = \rho_2(\hat{\varepsilon}_t) = \ldots = \rho_k(\hat{\varepsilon}_t) = 0$$
$$H_a: \rho_1(\hat{\varepsilon}_t) \neq \rho_2(\hat{\varepsilon}_t) \neq \ldots \neq \rho_k(\hat{\varepsilon}_t) \neq 0$$

ทั้งนี้ค่า Q นั้นจะพบว่ามีการแจกแจงเป็นแบบ Chi-square ที่มีดีกรีเท่ากับ m ซึ่งอยู่ภายใต้ข้อสมมุติฐานว่า สมมุติฐานว่างคือค่าความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น White Noise นั่นก็แปลว่าแบบจำลองมีลักษณะปราศจากสหสัมพันธ์ (Autocorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นมีปราศจากสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

**4. การพยากรณ์ (Forecasting)** คือการพยากรณ์ล่วงหน้าโดยอาศัยแบบจำลองที่เหมาะสมมากที่สุด เนื่องจากการพยากรณ์ข้อมูลล่วงหน้านั้น จะทำให้เกิดข้อจำกัดที่ว่าความแม่นยำของข้อมูลที่ได้จากการพยากรณ์นั้น มีความน่าเชื่อถือมากน้อยเพียงใด และแบบจำลอง ARIMA เหมาะสำหรับการพยากรณ์ในระยะสั้น ดังนั้นเพื่อที่จะทราบว่าแบบจำลองที่ประมาณขึ้นมานั้น สามารถที่จะพยากรณ์ราคาได้ถูกต้องมากน้อยเพียงใด จึงได้ใช้การพยากรณ์แบบ Ex-post Forecast กล่าวคือ เป็นการพยากรณ์ข้อมูล ณ ช่วงเวลาที่มีข้อมูลจริงเกิดขึ้นแล้ว ยกตัวอย่างเช่น จะลดจำนวนค่าสังเกตการของอนุกรมเวลาลงจากข้อมูลที่มีทั้งหมด n ข้อมูล เหลือ n-4 ข้อมูล แล้วทำการถดถอยข้อมูลใหม่เพื่อดูค่า RMSE (Root Mean Squared Error) และ Theil Inequality Coefficient และทำการพยากรณ์ล่วงหน้า (Ex-post Forecast) จำนวน 4 ข้อมูล เพื่อเปรียบเทียบกับข้อมูลจริงที่มีอยู่ และก็จะได้ค่า RMSE (Root Mean Squared Error) และ

Theil Inequality Coefficient แล้วใช้ค่าสถิติดังกล่าวประกอบการพิจารณาเลือกแบบจำลองที่มีความเหมาะสม สามารถพิจารณาช่วงการพยากรณ์ได้ดังรูปต่อไปนี้

รูปที่ 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์




ที่มา: (Pindyck and Rubinfeld, 1997)

ภายหลังจากที่เลือกแบบจำลองที่ใช้เป็นตัวแทนอนุกรมเวลาข้อมูลได้แล้ว จะทำการพยากรณ์ล่วงหน้า (Ex-ante forecast) กล่าวคือเป็นการพยากรณ์ในช่วงเวลาที่ยังไม่มีข้อมูลจริงเกิดขึ้นต่อไป

### 3.2 ระเบียบวิธีวิจัย

เนื้อไก่.

เนื่องจากข้อมูลอนุกรมเวลาราคาไก่~~เนื้อ ชนิดเนื้ออกถอดกระดูกและเนื้อสันใน~~มีลักษณะแปรปรวนมาก ดังนั้นจึงได้แปลงอนุกรมราคานี้ ให้อยู่ในรูปของลอการิทึมฐานธรรมชาติ ทั้งนี้ในการศึกษาจะใช้ ตัวแปร SBR ~~LBB~~ แทนราคาเนื้ออกถอดกระดูกที่แปลงในรูปของลอกการิทึมฐานธรรมชาติ~~และตัวแปร LFL แทนราคาเนื้อสันในที่แปลงในรูปลอกการิทึมฐานธรรมชาติ~~ จากแนวคิดที่เกี่ยวข้อง พบว่าเนื่องจากข้อมูลที่จะสร้างแบบจำลองเพื่อใช้ในการพยากรณ์ จะต้องมีลักษณะที่นิ่ง กล่าวคือ ข้อมูลจะมีค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วม ไม่เปลี่ยนแปลง ในขณะที่เวลาเปลี่ยนแปลงไป ดังนั้นจึงต้องมีการทดสอบความนิ่ง (Stationary) ของข้อมูล โดยการศึกษานี้ใช้การทดสอบความนิ่งด้วย Unit root test โดยใช้ ADF-test ดังแสดงในสมการดังต่อไปนี้

เนื้ออกถอดกระดูก

$$\Delta LBB_t = \theta LBB_{t-1} + \sum_{i=1}^{4} \phi_i \Delta LBB_{t-i} + \varepsilon_t \tag{3.22}$$

$$\Delta LBB_t = \alpha + \theta LBB_{t-1} + \sum_{i=1}^{4} \phi_i \Delta LBB_{t-i} + \varepsilon_t \tag{3.23}$$

$$\Delta LBB_t = \alpha + \beta t + \theta LBB_{t-1} + \sum_{i=1}^{4} \phi_i \Delta LBB_{t-i} + \varepsilon_t \tag{3.24}$$

เนื้อสันใน

$$\Delta LFL_t = \theta LFL_{t-1} + \sum_{i=1}^{4} \phi_i \Delta LFL_{t-i} + \varepsilon_t \tag{3.25}$$

$$\Delta LFL_t = \alpha + \theta LFL_{t-1} + \sum_{i=1}^{4} \phi_i \Delta LFL_{t-i} + \varepsilon_t \tag{3.26}$$

$$\Delta LFL_t = \alpha + \beta t + \theta LFL_{t-1} + \sum_{i=1}^{4} \phi_i \Delta LFL_{t-i} + \varepsilon_t \tag{3.27}$$

จากสมการที่ (3.22) – (3.24) ซึ่งแสดงการทดสอบ ADF Test ด้วยวิธี Unit root test ของเนื้ออกถอดกระดูก และจากสมการที่ (3.25) - (3.27) แสดงการทดสอบ ADF Test ด้วยวิธี Unit root test ของเนื้อสันใน ทั้งนี้วิธีการหา Lag Length จะเริ่มที่ความยาว Lag ที่ 4 แล้วจะลดขนาดลงเรื่อยๆ หากว่าค่า T-statistic ไม่มีนัยสำคัญทางสถิต ณ ระดับ 10 % จนกระทั่งค่า T-statistic มีนัยสำคัญ โดยกำหนดข้อสมมุติฐานค่า θ ดังนี้

- $H_0$: $\theta = 0$ คือข้อมูลราคาจะมี unit root หรือกล่าวคือมีลักษณะที่ไม่นิ่ง ซึ่งจะต้องทำการหาผลต่างอันดับต่อไป

- $H_a$: $\theta < 0$ คือข้อมูลราคาปราศจาก Unit root หรือกล่าวคือมีลักษณะที่นิ่ง ณ อันดับนั้น

เมื่อทราบถึงข้อมูลว่าเป็นลักษณะข้อมูลอันดับใด (Order of Integrated) แล้ว ก็จะนำข้อมูลอนุกรมเวลานี้มาเพื่อที่จะหารูปแบบ ARIMA (p,d,q) ด้วยวิธีของ Box-Jenkins ดังนี้

1. การหารูปแบบที่เหมาะสม (Identification) โดยการเลือกรูปแบบจำลองที่เหมาะสมด้วยการพิจารณาจาก Correlogram และจากค่า RMSE

2. การประมาณค่า (Estimation) คือการนำเอารูปแบบ ARMA (p,q) ที่เลือกจากขั้นตอนการหารูปแบบที่เหมาะสมมา ประมาณค่าพารามิเตอร์ โดยพิจารณาความมีนัยสำคัญทางสถิตด้วย T-statistic

3. การตรวจสอบความถูกต้อง (Diagnostic checking) คือการตรวจสอบสหสัมพันธ์ (Autocorrelation) จากค่าความคลาดเคลื่อนที่ประมาณได้ว่ามีลักษณะเป็น White noise หรือไม่

4. การพยากรณ์ (Forecasting) คือการพยากรณ์ข้อมูลล่วงหน้า โดยในการศึกษาครั้งนี้จะแบ่งช่วงการพยากรณ์ดังนี้

- Ex-post forecast เพื่อเปรียบว่าแบบจำลองใดจะมีความสามารถในการพยากรณ์ที่ดีมากที่สุด โดยการลดจำนวนข้อมูลลงจาก 135 ค่าสังเกต เหลือ 131 ค่าสังเกตและทำการถดถอยข้อมูลใหม่ และพยากรณ์ราคา 4 คาบเวลาถัดไป กล่าวคือ พยากรณ์ช่วงเวลาที่ 132 จนกระทั่งช่วงเวลาที่ 135 และนำข้อมูลที่พยากรณ์ได้เปรียบเทียบกับข้อมูลจริง โดยพิจารณาจากค่าสถิติ RMSE ที่ต่ำที่สุด

- Ex-ante forecast เมื่อทราบแบบจำลองที่สามารถพยากรณ์ได้ดีที่สุด แล้วจึงนำเอาแบบจำลองนั้นไปพยากรณ์ช่วงเวลาถัดไปอีก 4 คาบเวลาถัดไป กล่าวคือ การพยากรณ์ ณ เวลาที่ 136 137 138 และ 139