บทที่ 3
แนวความคิดและระเบียบวิธีวิจัย

3.1 แนวความคิดเกี่ยวกับการพยากรณ์ข้อมูลอนุกรมเวลา

3.1.1 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลา

ข้อมูลอนุกรมเวลา (Time series data) คือค่าสังเกต (Observation) ชุดหนึ่งซึ่งถูกกำหนดขึ้น ณ เวลาต่างๆ ถ้าค่าสังเกตกระทำในเวลาที่ต่อเนื่องกันจะเรียกว่า ข้อมูลอนุกรมเวลาต่อเนื่อง แต่ถ้าค่าสังเกตกระทำ ณ จุดเวลาที่ไม่ต่อเนื่องกัน เรียกว่า ข้อมูลอนุกรมเวลาไม่ต่อเนื่อง ดังนั้นการวิเคราะห์ข้อมูลอนุกรมเวลาจึงเป็นการวิเคราะห์ค่าสังเกตที่มีการเปลี่ยนแปลงไปตามเวลาที่กระทำ

การพยากรณ์ข้อมูลอนุกรมเวลาต้องอยู่บนข้อสมมุติว่า ข้อมูลอนุกรมเวลาต้องมีลักษณะนิ่ง(Stationary) (Enders,1995) ซึ่งสามารถพิจารณาความนิ่งของข้อมูลได้จาก

ค่าเฉลี่ย	:	$E(X_t) = \mu =$ constant	(1)
ค่าความแปรปรวน	:	$V(X_t) = \sigma^2 =$ constant	(2)
ค่าความแปรปรวนร่วม	:	$cov(X_t, X_{t+k}) = E(X_t-\mu)(X_{t+k}-\mu)$	(3)

ข้อมูลที่มีลักษณะนิ่งแล้วจะต้องมีค่าเฉลี่ย ค่าความแปรปรวน ของทุกๆค่า ณ เวลา t ใดๆ คงที่ ในขณะที่ความแปรปรวนร่วมระหว่างสองคาบเวลาจะขึ้นอยู่กับช่องว่างระหว่างสองคาบเวลาเท่านั้น ไม่ขึ้นอยู่กับเวลาที่เปลี่ยนไป หากเงื่อนไขเงื่อนเงื่อนไขหนึ่งไม่เป็นดังที่กล่าวมาแสดงว่าข้อมูลมีลักษณะไม่นิ่ง (Nonstationary) (Charemza and Deadman,1992)

ดังนั้นจึงสรุปได้ว่า ข้อมูลจะมีลักษณะนิ่งจะต้องมีค่าเฉลี่ย ค่าความแปรปรวนร่วมคงที่ ทุกๆเวลา t ใดๆ

3.1.2 การทดสอบความนิ่งของข้อมูล (Unit Root Test)

การพยากรณ์ข้อมูลอนุกรมเวลาที่ให้ความแม่นยำนั้น จะต้องมีการทดสอบว่าข้อมูลนั้น ค่าเฉลี่ยและค่าความแปรปรวนของตัวแปรมีค่าคงที่หรือไม่ ซึ่งก็คือการทดสอบว่าข้อมูลมีลักษณะนิ่งหรือไม่ เรียกว่า การทดสอบ unit root โดยจะตั้งสมมุติฐานว่าง (Null hypothesis) ของการทดสอบคือ $H_0 : \rho = 1$ การทดสอบนี้เป็นการทดสอบ DF (Dickey and Fuller test) (Dickey and Fuller, 1981) และการทดสอบ ADF (Augmented Dickey and Fuller test) (Said and Dickey 1984)

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad (4)$$

จากสมการ (4) เป็นการนำไปสู่การทดสอบ unit root ถ้า $|\rho| < 1$ X_t จะมีลักษณะนิ่ง แต่เมื่อ $|\rho|=1$ X_t จะมีลักษณะไม่นิ่ง และจากสมการ (4)นั้นก็สามารถเปลี่ยนรูปแบบสมการได้เป็น

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \qquad (5)$$

จากสามการ (5) จะสามารถพิจารณาความนิ่งของข้อมูลอนุกรมเวลาwfhโดยที่ $\rho = (1 + \theta)$ ดังนั้นถ้า θ มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (1) จะมีค่าน้อยกว่า 1 ดังนั้นก็สามารถสรุปได้ว่า การปฏิเสธสมมติฐานว่าง ซึ่งก็คือ $H_0 : \theta = 0$ ถือว่าเป็นการยอมรับ $H_a : \theta < 0$ แสดงว่า $\rho < 1$ และ X_t จะมีลักษณะ integration of order zero (Charemza and Deadman, 1992) นั่นคือ X_t มีลักษณะนิ่ง ในทางตรงกันข้ามถ้ายอมรับ $H_0 : \theta = 0$ ได้ หมายความว่า X_t มีลักษณะไม่นิ่ง ถ้า X_t เป็นแนวคิดเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) ก็สามารถเขียนแบบจำลองได้ดังสมการ (6) และถ้า X_t เป็นแนวคิดเดินเชิงสุ่มมีความโน้มเอียงทั่วไปรวมอยู่ด้วย และมีแนวโน้มตามเวลาเชิงเส้น (linear time trend) สามารถเขียนเป็นแบบจำลองดังสมการ (7)

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \qquad (6)$$

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t \qquad (7)$$

ซึ่งการพิจารณาความนิ่งยังคงให้ความสำคัญต่อ θ ดังเช่นการพิจารณา สมการ (5) คือมีสมมติฐานว่าง คือ $H_0 : \theta = 0$ และ $H_a : \theta < 0$ ซึ่งจะเปรียบเทียบค่าสถิติ t (t-statistic)ที่คำนวณกับค่าวิกฤติที่อยู่ในตาราง Dickey – Fuller (Enders,1995) หรือกับค่าวิกฤติ MacKinnon (MacKinnon critical values) (Gujarati,1995)

เมื่อมีการแทนที่ของกระบวนการเชิงถดถอย (Autoregressive process) ลงในสมการ (5) (6) และ (7) ก็ไม่ทำให้ค่าวิกฤติเปลี่ยนแปลงไป ดังสมการต่อไปนี้

$$\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \qquad (8)$$

$$\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \qquad (9)$$

$$\Delta X_t = \alpha + Bt + \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_i \Delta X_{t-i} + \varepsilon_t \qquad (10)$$

ซึ่งเรียกว่าการทดสอบ ADF (Augmented Dickey-Fuller test) โดยที่จำนวนของ logged difference terms ที่จะนำมาใช้ในสมการที่นี้มีมากพอที่จะทำให้พจน์ค่าความคลาดเคลื่อนมีลักษณะเป็น Serially independent ซึ่งการทดสอบจะใช้ค่าสถิติทดสอบ ADF (ASF test statistic)

และมีการแจกแจงเชิงเส้นกำกับ ดังนั้นจึงสามารถใช้ค่าวิกฤติแบบเดียวกับการทดสอบ DF (Gujarati,1995)

สำหรับการเลือก lag length นั้น Walter Enders (1995) กล่าวว่าควรจะเริ่ม lag length ที่มีค่าที่มากพอ แล้วพิจารณาความมีนัยสำคัญทางสถิติ ที่ระดับนัยสำคัญต่างๆ (α = 0.01 0.05 และ0.1) เมื่อพบว่าที่ lag length ที่เลือกมีค่า t-Statistic ที่ไม่มีนัยสำคัญทางสถิติ ณ ระดับนัยสำคัญ 0.1 แล้ว จึงทำการลด lag length ลงทีละ 1 ช่วง จนกระทั่งทำการปฏิเสธสมมติฐานว่าง คือ ค่า t-Statistic มีนัยสำคัญทางสถิติ

3.1.3 การพยากรณ์โดยวิธี Box- Jenkins

วิธีของ Box-Jenkins เป็นการวิเคราะห์ข้อมูลอนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับข้อมูลอนุกรมเวลา (ทรงศิริ แต้สมบัติ, 2539) รูปแบบที่จะกำหนดให้กับข้อมูลอนุกรมเวลาจะเป็นรูปแบบในกลุ่มของรูปแบบ ARIMA (p,d,q) (Integrated Autoregressive-Moving Average order p and q) ซึ่งเป็นการรวมส่วนของรูปแบบ AR (p) และรูปแบบ MA (q) เข้าด้วยกัน ส่วนอันดับของ d คือจำนวนครั้งที่หาผลต่าง (Integrated) รูปแบบ AR (p) หมายถึงรูปแบบที่แสดงให้เห็นว่าค่าสังเกต Y_t ก็จะขึ้นอยู่กับค่าของ $Y_{t-1}, \dots, Y_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า และรูปแบบ MA (q) คือรูปแบบที่แสดงว่าค่าสังเกต Y_t จะขึ้นอยู่กับค่าของ ความคลาดเคลื่อน $\varepsilon_{t-1}, \dots, \varepsilon_{tq}$ หรือความคลาดเคลื่อนที่อยู่ก่อนหน้า q ค่า และรูปแบบ I(d) เกิดจากการหาผลต่างของอนุกรมเวลา ซึ่งรูปแบบ ARIMA (p,d,q) มีการกำหนดรูปแบบดังนี้

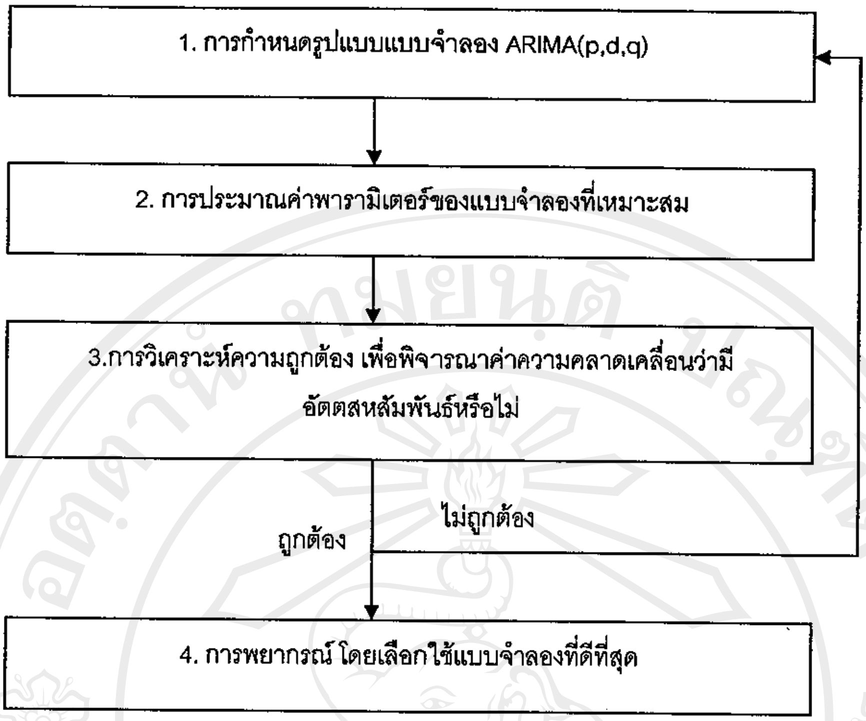
AR (p) คือ $Yt = \theta_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t$

MA (q) คือ $Yt = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$

ARMA (p,q) คือ $Yt = \theta_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$

ARIMA (p,d,q) คือ $\Delta^d Yt = \theta_0 + \phi_1 \Delta^d Y_{t-1} + \dots + \phi_p \Delta^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$

การพยากรณ์โดยใช้วิธี Box- Jenkins ประกบด้วย 4 ขั้นตอน คือ (1) กำหนดรูปแบบจำลอง (Identification) (2) การประมาณค่า (Estimation) (3) การวิเคราะห์ความถูกต้อง (Diagnostic checking) (4) การพยากรณ์ (Forecasting) ดังรูป 3.1



รูป 3.1 ขั้นตอนการพยากรณ์โดยวิธี Box- Jenkins
ที่มา : Gujarati, N. (2003)

1. การกำหนดแบบจำลอง (Identification)

การกำหนดแบบรูปแบบ ให้กับอนุกรมเวลาที่มีลักษณะนิ่ง เป็นการหารูปแบบ ARIMA (p,d,q) ที่คาดว่าจะเหมาะสมให้กับข้อมูลอนุกรมเวลา โดยจะใช้ค่าพารามิเตอร์สองลักษณะ ได้แก่ ค่าอัตตสหสัมพันธ์ (Autocorrelation Function: ACF) และค่าอัตตสหสัมพันธ์บางส่วน (Partial Autocorrelation : PACF)

อัตตสหสัมพันธ์ของประชากร (ρ_k) คือการวัดความสัมพันธ์ของแต่ละช่วงเวลาของค่าความคลาดเคลื่อน (Moving average) โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา และ ρ_k มีค่าเท่ากับ $-1 \leq \rho_k \leq 1$ ซึ่งมีสูตรดังนี้

$$\rho_k = \frac{\sum_{t=a}^{n-k}(Y_t-\bar{Y})(Y_{t+k}-\bar{Y})}{\sum_{t=a}^{n}(Y_t-\bar{Y})^2} \tag{11}$$

โดยที่ $Y_t = \sum_{t=a}^{n}(Y_t)$

ในขณะที่การพิจารณาความสัมพันธ์ตัวแปรตามกับตัวแปรอิสระจะใช้ ค่าอัตตสหสัมพันธ์บางส่วน เป็นตัววัดความสัมพันธ์ของตัวแปรทั้งสอง โดยพิจารณาจากสมการ Yule-walker (Pindyck and Rubinfely, 1996) ดังนั้น

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p \tag{12}$$

ถ้า k มากกว่า p จะได้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_{k-p} \tag{13}$$

เมื่อทราบคอเรลโลแกรมของ ACF และ PACF แล้ว สามารถนำมาหารูปแบบแบบจำลองโดยการพิจารณาลักษณะของ คอเรลโลแกรมของ ACFและPACF ดังตาราง 3.1

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, N. (2003)

จากตาราง 3.1 สามารถกำหนดรูปแบบจำลองได้โดยการเปรียบเทียบระหว่าง คอเรลโลแกรม ACF และ PACF ถ้าหากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็น ค่าที่ p ของ AR(p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ

ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรม เกิดขึ้น 1 แท่ง แปลได้ว่าแบบจำลองควรมีลักษณะเป็น AR(1) 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) และเมื่อรวมกันกับการทดสอบความนิ่งของข้อมูลอนุกรมเวลาแล้ว จะสามารถหาค่าผลต่างปกติและผลต่างฤดูกาลได้ ซึ่งผลจากการ หาค่าผลต่างปกติและผลต่างฤดูกาล จำนวน d ครั้งนั้น ก็จะได้แบบจำลอง ARIMA (p,d,q)

อย่างไรก็ตามการพิจารณาดังกล่าวเป็นเพียงการพิจารณาตามหลักการเท่านั้น เพราะการพิจารณารูปแบบ ARIMA (p ,d, q) นั้นเป็นศิลป์อย่างหนึ่ง จึงควรมีการพิจารณาค่าสถิติเพื่อประกอบการตัดสินใจ เช่น ค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root Mean Square Error) ค่า Theil' s inequality coefficient ค่า Adjusted R^2 และค่า Akaike Information Criterion (AIC)

1.1 ค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (Root Mean Squae Eorror)

ค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง คือการวัดค่าความแตกต่างระหว่างค่าความจริงและค่าที่ถูกประมาณจากแบบจำลอง หากค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง มีค่าน้อยแสดงว่าแบบจำลองสามารถประมาณค่าประมาณได้ใกล้เคียงกับค่าจริง (Pindyck,1998) ดังนั้นหากค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสองนั้นมีค่าเท่ากับศูนย์แล้ว จึงหมายความว่าไม่เกิดความคลาดเคลื่อนในแบบจำลอง ดังนั้นจึงสามารถพิจารณาค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง ได้ดังนี้

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^s - Y_t^a\right)} \tag{14}$$

กำหนดให้ Y_t^s = ค่าประมาณจากแบบจำลอง

Y_t^a = ค่าที่แท้จริง

T = จำนวนคาบเวลาที่ใช้ในการประมาณแบบจำลอง

1.2 ค่า Theil' s Ineqvality Coeffient

ค่า Theil' s ineqvality coeffient (U) นั้นมีที่มาคล้ายกับค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง แต่ค่า U นั้นจะมีค่าอยู่ระหว่าง 0 ถึง 1 ถ้า U มีค่าเท่ากับศูนย์ แสดงว่าค่าที่ได้จากการประมาณเท่ากับค่าจริงของทุกๆเวลา t และแบบจำลองที่ประมาณได้นั้นเป็นแบบจำลองที่ดี แต่ในทางตรงกันข้ามถ้าค่า U มีค่าเท่ากับหนึ่ง แสดงว่าแบบจำลองที่ประมาณเป็นแบบจำลองที่ไม่ดีที่สุด (Pindyck,1998) ดังนั้นสามารถพิจารณาค่า U ได้ดังนี้

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^s - Y_t^a\right)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^s\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t^a\right)^2}} \tag{15}$$

กำหนดให้ Y_t^s = ค่าประมาณจากแบบจำลอง

Y_t^a = ค่าที่แท้จริง

T = จำนวนคาบเวลาที่ใช้ในการประมาณแบบจำลอง

1.3 ค่า Adjusted R square (Adjusted R^2)

Adjusted R^2 คือการพิจารณาว่าตัวแปรอิสระสามารถที่จะอธิบายถึงการเปลี่ยนแปลงตัวแปรตามได้มากน้อยเพียงใด ถ้าหากค่า Adjusted R^2 มีค่าเท่ากับ 1 แสดงว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% แต่ถ้าหากค่า Adjusted R^2 มีค่าเท่ากับ 0 หมายความว่า ตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย ซึ่งค่า Adjusted R^2 นี้เป็นค่าสถิติที่เกิดจากการประยุกต์มาจากค่า R^2 ซึ่งถ้ามีการเพิ่มตัวแปรอิสระเข้าไปในสมการมากขึ้นก็จะทำให้ค่า R^2 สูงขึ้น ดังนั้นจึงมีการเพิ่มระดับความเป็นอิสระในสมการ เรียกว่า Adjusted R^2 ($\bar{R}^2$) (Gujarati,2003) โดยสามารถพิจารณาความสัมพันธ์ของ R^2 และ Adjusted R^2 ได้ดังสมการ

$$R^2 = 1 - \frac{\sum \hat{u}_i}{\sum y_i^2} \tag{16}$$

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2/(n-k)}{\sum y_i^2/(n-1)} \tag{17}$$

1.4 ค่า Akaike Information Criterion (AIC)

ค่า AIC เป็นค่าสถิติที่มีการประยุกต์คล้ายกับค่า Adjusted R^2 แต่มีการถ่วงน้ำหนักมากกว่า Adjusted R^2 และยังมีการลอกกาลิทึมฐานธรรมชาติ (Natural logarithm : ln) ดังนั้นถ้าค่า AIC น้อยจึงหมายถึงแบบจำลองสามารถเป็นตัวแทนข้อมูลจริงได้ดี (Gujarati,2003) และยังนำค่า AIC ไปใช้ในการหาค่าย้อนหลัง (lag length) ที่เหมาะสมได้อีก สามารถคำนวณค่า AIC ได้ดังนี้

$$\ln AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{\sum \hat{u}_i^2}{n}\right) \tag{18}$$

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

n = ค่าสังเกตทั้งหมด

2. การประมาณค่า (Estimation)

การประมาณค่า คือ การประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเองและรูปแบบการเคลื่อนที่ของค่าความคลาดเคลื่อน ซึ่งสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (Simple least square) หรือใช้วิธีการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear) เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์ได้ เมื่อรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3. การวิเคราะห์ความถูกต้อง (Diagnostic Checking)

การวิเคราะห์ความถูกต้อง หมายถึงการตรวจสอบรูปแบบแบบจำลองว่าเป็นแบบจำลองที่กำหนดขึ้นมานั้นมีความเหมาะสมหรือไม่ โดยการพิจารณาจากคอเรลโลแกรมของอัตตสหสัมพันธ์ของกลุ่มตัวอย่าง ซึ่งเป็นการทดสอบโดยใช้การทดสอบของ Box and Pierce ซึ่งพิจารณาจากค่า Q-statistic (Gujarati, 2003) ดังสมการ (19) ซึ่งมีการแจกแจงแบบ Chi-square และมีดีกรีเท่ากับ m โดยมีสมมติฐานว่าง คือ พจน์ความคลาดเคลื่อนที่ได้จากการประมาณมีลักษณะเป็น white noise ซึ่งหมายถึง แบบจำลองไม่มีอัตตสหสัมพันธ์ ถ้าหากแบบจำลองมีอัตตสหสัมพันธ์ จะต้องกลับไปทำการกำหนดรูปแบบแบบจำลองใหม่ แต่หากแบบจำลองไม่มีอัตตสหสัมพันธ์ ก็จะใช้แบบจำลองนั้นทำการพยากรณ์ต่อไป

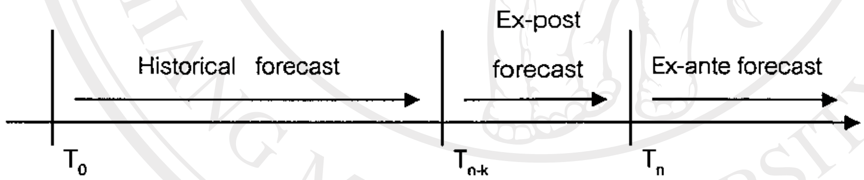
$$Q = n\sum_{k-1}^{m}\hat{\rho}_k^2 \qquad (19)$$

กำหนดให้ n = จำนวนของข้อมูล

M = ค่า lag length

4. การพยากรณ์ (Forecasting)

เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากการวิเคราะห์ความถูกต้องแล้ว ก็สามารถนำแบบจำลองใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลองโดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง Historical forecast อันเป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา (T_{n-k}) การพยากรณ์ช่วง Ex-post forecast คือการพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์แล้วเปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า Root Mean Square Error (RMSE) และ Theil's Inequality Coefficient (TIC) และค่า Akaike Information Criterion (AIC) จะพิจารณาค่าสถิติทั้งสามค่าที่มีค่าน้อยที่สุด ซึ่งได้จากการทำการพยากรณ์ เมื่อเลือกรูปแบบจำลองที่ดีที่สุดได้แล้ว จึงนำแบบจำลองนั้นมาทำการพยากรณ์แบบ Ex-ante forecast ซึ่งเป็นการพยากรณ์ข้อมูลไปข้างหน้า ดังรูป 3.2



รูป 3.2 ช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา : Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ราคาวัสดุก่อสร้าง โดยวิธีARIMA เป็นการนำเอาข้อมูลอนุกรมเวลามาหารูปแบบแบบจำลองที่เหมาะสมที่สุด ซึ่งมีขั้นตอนต่อไปนี้

1. การลอกกาลิทึมฐานธรรมชาติ (Natural logarithm : ln) ข้อมูลอนุกรมเวลาราคาวัสดุก่อสร้างเพื่อให้ข้อมูลไม่อยู่ภายใต้อิทธิพลของเวลาและข้อมูลไม่มีการแปรปรวนมาก และ ตัวแปร X_t ได้ถูกเปลี่ยนเป็น MP_t เพื่อความเหมาะสมของความหมายของข้อมูล ดังนั้น ตัวแปรที่จะใช้ในสมการต่างๆจึงถูกเปลี่ยนเป็น $\ln MP_t$

2. การทดสอบความนิ่งของข้อมูล เป็นการพิจารณาว่าข้อมูลอนุกรมเวลามีลักษณะนิ่งหรือไม่โดยการทดสอบ unit root ดังสมการต่อไปนี้

$$\Delta \ln MP_t = \theta \ln MP_{t-1} + \sum_{i=1}^{3} \phi_i \Delta \ln MP_{t-i} + \varepsilon_t \tag{20}$$

$$\Delta \ln MP_t = \alpha + \theta \ln MP_{t-1} + \sum_{i=1}^{3} \phi_i \Delta \ln MP_{t-i} + \varepsilon_t \tag{21}$$

$$\Delta \ln MP_t = \alpha + Bt + \theta \ln MP_{t-1} + \sum_{i=1}^{3} \phi_i \Delta \ln MP_{t-i} + \varepsilon_t \tag{22}$$

3. การกำหนดรูปแบบแบบจำลอง ARIMA(p,d,q) โดยการพิจารณาคอเรลโลแกรม Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF) เพื่อจะสามารถระบุว่าแบบจำลองควรจะมี Autoregressive(p) เท่าใด และ Moving average (q) เท่าใด โดยเลือกสร้างไว้หลายแบบจำลองเพื่อหาแบบจำลองที่เหมาะสมที่สุด

4. การประมาณค่าพารามิเตอร์ เพื่อนำค่าพารามิเตอร์ที่ได้นั้นนำไปทำการพยากรณ์ราคาต่อไปได้

5. การตรวจสอบความถูกต้อง เมื่อทำการหาแบบจำลองที่เหมาะสมและประมาณค่าพารามิเตอร์แล้ว จึงทำการทดสอบแบบจำลองโดยพิจารณาจากค่า Q-statistic จากคอเรลโลแกรมของอัตตสหสัมพันธ์ของกลุ่มตัวอย่าง

6. การพยากรณ์ จะทำการพยากรณ์ดัชนีราคาวัสดุก่อสร้าง ซึ่งจะทำการแบ่งการพยากรณ์ออกเป็น 3 ช่วง ได้แก่ ช่วง Historical forecast ช่วง Ex-post forecast และ ช่วง Ex-ante forecast ดังรูป 3.3

รูป 3.3 ช่วงเวลาที่ใช้ในการพยากรณ์จริง

ที่มา : Pindyck and Rubinfeld (1997)