บทที่ 3

กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการศึกษานี้จะทำการพยากรณ์มูลค่าการส่งออกรถยนต์และชิ้นส่วนที่คาดว่าจะเกิดขึ้นในอนาคตโดยวิธีของ Box and Jenkins มูลค่ารายเดือนโดยการพยากรณ์จะนำอนุกรมเวลาในอดีตมาพยากรณ์อนุกรมเวลาในอนาคต ซึ่งเป็นการพยากรณ์ที่ให้ค่าความถูกต้อง (accuracy) สูงกว่าวิธีอื่นในการพยากรณ์ระยะเวลาสั้น และต้องใช้ข้อมูลอนุกรมเวลาที่มีลักษณะนิ่ง (stationary) โดยการศึกษาต้องทำการทดสอบ unit root ก่อนที่จะทำการพยากรณ์โดยวิธีอารีมาเพื่อให้ค่าความคลาดเคลื่อนจากการพยากรณ์ลดลง เนื่องจากข้อมูลเป็นอนุกรมเวลา (time series data) ส่วนมากมีลักษณะไม่นิ่ง (non-stationary)

3.1.1 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลา

การพยากรณ์มูลค่าการส่งออกรถยนต์และชิ้นส่วน โดยกำหนดรูปแบบให้กับอนุกรมเวลาเพื่อพิจารณามูลค่าการส่งออกรถยนต์และชิ้นส่วนล่วงหน้าและนำค่าดังกล่าวมาใช้ในการตัดสินใจวางแผนการผลิตและการส่งออกรถยนต์และชิ้นส่วน ให้มีประสิทธิภาพดียิ่งขึ้นและสอดคล้องกับความต้องการของตลาด ในการพยากรณ์มูลค่าการส่งออกรถยนต์และชิ้นส่วนรายเดือนของประเทศไทยเป็นการพรรณนาอธิบายและวิธีการวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองอนุกรมเวลาในรูปแบบ ARIMA Model วิธีการวิเคราะห์อนุกรมเวลาของ Box – Jenkins นั้นเป็นวิธีการพยากรณ์ที่มีความถูกต้องและเหมาะสมกว่าวิธีอื่นในการพยากรณ์ระยะสั้น หากต้องการพยากรณ์ในช่วงเวลาที่ยาวนานกว่านี้ควรใช้ข้อมูลที่ปรับค่าพยากรณ์แล้วเพื่อให้ความคลาดเคลื่อนจากการพยากรณ์ลดลง ซึ่งสามารถพิจารณาความนิ่ง (stationary) ของข้อมูลได้จาก

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 หากข้อมูลอนุกรมเวลานั้นนิ่ง (stationary) จะมีค่าเฉลี่ย ค่าความแปรปรวน และค่าความแปรปรวนร่วม (จากสมการที่ 3.1, 3.2 , 3.3) มีค่าที่คงที่ ณ ทุกๆ เวลาที่เปลี่ยนแปลงไปซึ่งสามารถทดสอบข้อมูลว่ามีลักษณะนิ่งหรือไม่จากจากการทดสอบ unit root

3.1.2 การทดสอบความนิ่งของข้อมูล (unit root test)

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

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

ซึ่งเรียกว่าการทดสอบ unit root โดยถ้า $|\rho| < 1$ X_t จะมีลักษณะนิ่ง (stationary) และถ้า $\rho = 1$ แล้ว X_t จะมีลักษณะไม่นิ่ง (nonstationary) อย่างไรก็ตามการทดสอบนี้สามารถทำได้อีกทางหนึ่งซึ่งเหมือนกับสมการ (1) กล่าวคือ

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \tag{2}$$

ซึ่งคือ $X_t = (1+\theta)X_{t-1} + \varepsilon_t$ ซึ่งคือสมการที่ (1) นั่นเอง โดยที่ $\rho = (1+\theta)$ ถ้า θ ในสมการ (2) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (1) 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 : 131) นั่นคือ X_t มีลักษณะนิ่ง (Stationary) และถ้าเราไม่สามารถปฏิเสธ $H_0 : \theta = 0$ ได้ ก็จะหมายความว่า X_t มีลักษณะไม่นิ่ง (nonstationary)

ถ้า X_t เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) เราสามารถเขียนแบบจำลองได้ดังนี้

$$\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t \tag{3}$$

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

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t \tag{4}$$

โดยที่ t = เวลา ซึ่งก็จะทำการทดสอบ $H_0 : \theta = 0$ โดยมี $H_a : \theta < 0$ เช่นเดียวกับที่กล่าวมาข้างต้นโดยสรุปแล้ว 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 $\theta = 0$; X_t จะมี unit root โดยการเปรียบเทียบค่าสถิติ t (t-statistic) ที่คำนวณได้กับค่าที่เหมาะสมที่อยู่ในตาราง Dickey-Fuller (Dickey-Fuller tables) (Enders,1995 : 221) หรือกับค่า วิกฤติ MacKinnon (MacKinnon critical values) (Gujarati,1995 : 769)

อย่างไรก็ตามค่าวิกฤติ (critical values) จะไม่เปลี่ยนแปลง ถ้าสมการ (2), (3), (4) ถูกแทนที่โดยกระบวนการเชิงอัตถดถอย (autoregressive processes)

$$\Delta X_{t-1} = \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{5}$$

$$\Delta X_{t-1} = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{6}$$

$$\Delta X_{t-1} = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{7}$$

(Enders, 1995 : 221 และ Gujarati, 1995 : 720) จำนวนของ lagged difference terms ที่จะนำเข้ามารวมในสมการนั้นมีมากพอที่จะทำให้พจน์ค่าความคลาดเคลื่อน (error terms) มีลักษณะเป็น serially independent และเมื่อนำเอาการทดสอบ DF (Dickey - Fuller (DF) Test) มาใช้กับสมการ (5) – (7) เราจะเรียกว่าการทดสอบ ADF (Augmented Dickey – Fuller (ADF) Test) ค่าสถิติทดสอบ ADF (ADF Test statistic) มีการแจกแจงเชิงเส้นกำกับ (asymptotic distribution) เหมือนกับสถิติ DF (DF statistic)

ดังนั้นก็สามารถใช้ค่าวิกฤติ (crititcal values) แบบเดียวกัน(Gujarati, 1995 : 720) ในกรณีของการหา lag length ที่เหมาะสมนั้น (Enders, 1995, 277) ได้เสนอแนะว่าวิธีหนึ่งในการหา lag length ที่เหมาะสมนั้น (Enders, 1995: 227) ได้เสนอแนะว่าวิธีหนึ่งในการหา lag length ก็คือเริ่มต้นด้วยการให้มี lag length ที่ยาวมากพอและก็ลดขนาดของ lag length ลงโดยใช้ค่าสถิติทดสอบ

t (t-test) หรือค่าสถิติทดสอบ F (F-test) สมมติว่าเราใช้ lag length เท่ากับ n* ถ้าสถิติ t (t – statistic) ของ lag n* ไม่มีนัยสำคัญ ณ ค่าวิกฤติ(critical value) ที่กำหนดให้ เราก็จะต้องทำการประมาณค่าการถดถอยใหม่ โดยใช้ lag length n*-1 ทำอย่างนี้ เรื่อยไปจนกระทั่ง lag นั้นมีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญ

3.1.3 แบบจำลองการพยากรณ์ โดยวิธี Box – Jenkins

การพยากรณ์อนุกรมเวลาโดยวิธี Box – Jenkins ในรูปแบบ อารีมา (p, d, q) ต้องพิจารณาว่าอนุกรมเวลาเป็น stationary series หรือไม่ (ทรงศิริ แต้สมบัติ, 2539) โดยพิจารณาจาก

1) ค่าเฉลี่ย $E(Y_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมากจะสรุปได้ว่า $E(Y_t)$ คงที่

2) ค่าความแปรปรวน $V(Y_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆแล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วนถ้าค่าความแปรปรวนแต่ละส่วนย่อย ไม่แตกต่างกันมากนักจะสรุปได้ว่า $V(Y_t)$ คงที่

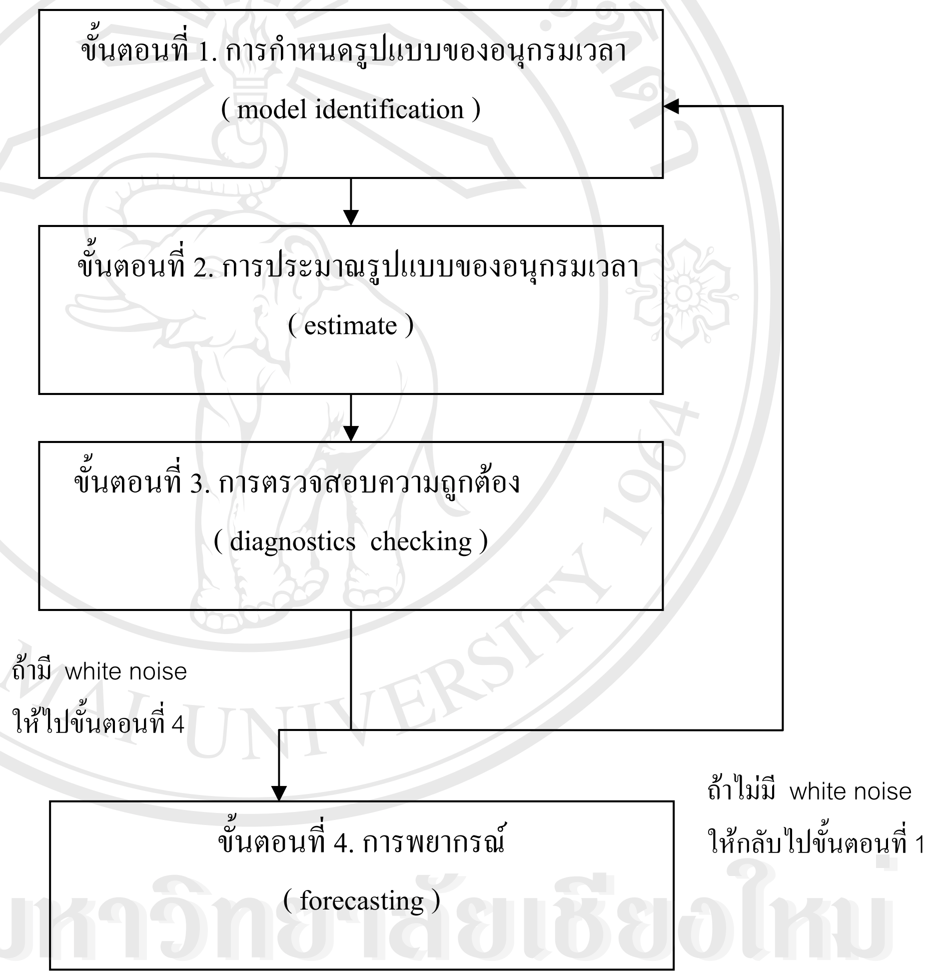
3) พิจารณาแนวโน้มและปัจจัยฤดูกาล ด้วยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและปัจจัยฤดูกาลมักจะเห็นชัดเจนได้จากรูปที่เรียกว่าคอเรลโรแกรม (correlogram)

4) พิจารณาคอเรลโลแกรม ของสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ stationary ค่าคอเรลโลแกรม ของสัมประสิทธิ์สหสัมพันธ์ (r_k) จะมีค่าลดลงค่อนข้างเร็ว เมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นถ้าค่าสัมประสิทธ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้าจะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้าและมีค่าค่อนข้างสูงที่ k = L, 2L, 3L จะเป็น ข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลของฤดูกาลและถ้าการเคลื่อนไหวของค่า คอเรลโลแกรมของสัมประสิทธิ์สหสัมพันธ์ (r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบภายใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วว่าอนุกรมเวลาที่ศึกษาไม่เป็น stationary ก่อนจะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่เป็น stationary จะต้องแปลงอนุกรมเวลาที่เป็น stationary เสียก่อนโดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลได้อนุกรมเวลาที่เป็น stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ ให้แปลงอนุกรมเวลาเดิมโดยการหา ลอการิทึม

Z = In (Y_t) จนกว่าจะได้อนุกรมเวลาใหม่ ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่เป็น stationary series แล้วจะทำตามขั้นตอนของ Box – Jenkins

ขั้นตอนของ Box-Jenkins ประกอบด้วย 4 ขั้นตอน ดังนี้ ขั้นตอนที่หนึ่ง คือการกำหนดรูปแบบจำลอง (identification) ขั้นตอนที่สอง คือ การประมาณค่า (estimation) ขั้นตอนที่สามคือ วิเคราะห์ความถูกต้อง (diagnostic checking) และขั้นตอนที่สี่ คือการพยากรณ์ (forecasting) ตามลำดับดังนี้



รูปที่ 3.1 แสดงขั้นตอนการพยากรณ์โดยวิธี Box and Jenkins

1) การกำหนดรูปแบบจำลอง (identification) ให้กับอนุกรมเวลาที่เป็น stationary series เป็นการหารูปแบบ ARMA (p.q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยที่ autocorrelation: ρ_k คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยที่ ρ_k มีค่าเท่ากับ $-1< \rho_k < 1$ โดยพิจารณาเปรียบเทียบค่าautocorrelation (r_k) ของอนุกรมเวลาตัวอย่างกับค่า

autocorrelation (ρ_k) ของอนุกรมเวลาของประชากรที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

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

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 ที่มาจากตัวอย่างโดยมีสูตรดังนี้

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

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

$$\rho_k = \frac{\sum_{t-a}^{n-k}(X_{t-q})(X_{t+k-q})}{\sum_{t-a}^{n}(X_{t-q})^2} \tag{1}$$

โดยที่ $X_t = \sum nt = a(X_t)$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตามเนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวตามแปรอิสระ ที่เป็นค่าความล่าช้า (Lag) ของตัวแปรตาม (autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (moving average) เนื่องจาก autocorrelation Function (ACF) จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อนแต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็น

ค่าความล่าช้าของตัวแปรตาม ซึ่ง patial autocorrelation function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule-walker (Pindyck and Rublinfeld, 1996) ดังนี้

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

ถ้า k มากกว่า p จะได้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p} \tag{3}$$

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 ซึ่งอาจจะใช้ตาราง 3.3 ดังต่อไปนี้พิจารณาร่วม

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 จะสามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วหายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็น ค่าที่ p ของ AR(p) ยกตัวอย่างเช่น เมื่อพิจารณา คอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรมเกิดขึ้น 1 แท่ง แปลได้ว่าแบบจำลองควรมีลักษณะเป็น AR (1) สำหรับ MA (q) นั้นก็จะมี ACF ที่

เกิดขึ้นมาไม่กี่ค่าแล้วหายไป 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: RMS) โดยจะเป็นการวัดค่าความคลาดเคลื่อนระหว่างค่าจริง และค่าที่ประมาณจากแบบจำลองมีความแตกต่างกันมากน้อยเพียงใด ซึ่งหากค่า RMSE มีค่าเท่ากับ 0 จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงใดก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากเพียงนั้น สามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s - X_t^a)} \qquad (4)$$

กำหนดให้ X_t^s คือค่าที่ประมาณจากแบบจำลอง

X_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

2. Theil's inequality coefficient โดยในหลักการเบื้องต้น พบว่าสมการที่ใช้นี้ยังคงมีหลักการที่คล้ายคลึงกันกับ RMSE โดยสิ่งที่ต่างออกไปจาก RMSE คือค่าสถิตินี้จะมีค่าอยู่ระหว่าง 0 และ 1 ทั้งนี้หากค่า U มีค่าเท่ากับศูนย์ นั่นก็หมายความว่าค่าที่ได้จากการประมาณมีค่าเท่ากันพอดีกับค่าที่เป็นข้อมูลจริงแสดงถึงแบบจำลองที่ประมาณได้เป็นแบบจำลองที่เป็นตัวแทนข้อมูลได้อย่างดีที่สุด ในขณะที่ถ้า U มีค่าเท่ากันกับหนึ่ง แปลว่าแบบจำลองทีประมาณได้เป็นแบบจำลองที่แย่ที่สุด ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่าUที่น้อยๆ ดังจะพิจารณาได้จากสมการที่ (5)

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s - X_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^a)^2}} \tag{5}$$

กำหนดให้ X_t^s คือค่าที่ประมาณจากแบบจำลอง

X_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

อย่างไรก็จามยังมีค่าสถิติอีกหลายอย่างที่สามารถนำมาพิจารณาประกอบร่วมกันกับRMSE และ Theil's inequality coefficient เพื่อใช้ในการเลือกแบบจำลองที่เหมาะสมที่สุด อาทิเช่น R^2,Adjusted R^2 และ Akike Information Criterion (AIC) ซึ่งสามารถอธิบายได้ดังต่อไปนี้

3. R^2 คือการวัดค่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดีเพียงใด หากค่านี้เท่ากับ 1 ก็หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ 100% ในทางกลับกัน หากค่านี้มีค่าเท่ากับ 0 แปลความหมายว่าตัวแปรอิสระไม่สามารถอธิบายตัวแปรตามได้เลย แต่อย่างไรก็ตามพบว่าหากมีการเพิ่มตัวแปรอิสระเข้าไปในสมการมาก ก็จะทำให้ค่า R มากขึ้นด้วย ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการที่ (6) ดังนั้นเพื่อปรับปรุงข้อจำกัดดังกล่าวข้างต้น จึงเกิดค่าสถิติใหม่ คือค่า Adjusted $R^2(\overline{R}^2)$ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการที่ (7)

$$R^2 = 1 - \frac{\sum \hat{u}_i}{\sum y_i^2} \tag{6}$$

$$\overline{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n-k)}{\sum y_i^2 / (n-1)} \tag{7}$$

4. Akaike Information Criterion (AIC) คือค่าสถิติที่ประยุกต์คล้ายกับ $\overline{R}^2$ แต่ใช้รูปแบบการใส่ค่าลอกาลิทึมฐานธรรมชาติ (Natural logarithm) โดยหากค่าสถิตินี้มีค่าน้อยเพียงใด นั้นก็แปลว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (Lag Length) ที่เหมาะสมอีกด้วย ซึ่งแสดงในสมการที่ (8)

$$\text{AIC} = \left(\frac{2k}{n}\right) + \log\left(\frac{\sum \hat{u}_i^2}{n}\right) \tag{8}$$

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

n คือค่าสังเกตทั้งหมด

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

5. Schwarz Criterion (SC) คือ วิธีการวัดปรับได้อย่างดี (Goodness of Fit) ของแบบจำลองได้ดีกว่า Akaike Information Criterion (AIC) เป็นวิธีที่ประยุกต์คล้ายกับ AIC และวิธีหนึ่งใน information criteria โดย information criteria จะประกอบด้วย Schwarz Criterion (SC) และ Akaike Information Criterion (AIC) สามารถเขียนในรูปสมการ SC ได้ดังสมการ (9)

$$\text{SC} = \log\left(\frac{\sum \hat{u}_i^2}{n}\right) + \left(\frac{2k \log n}{n}\right) \tag{9}$$

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

3) การตรวจสอบแบบจำลอง (diagnostics) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลองจะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่การตรวจสอบจะทำได้หลายวิธี ยกตัวอย่างเช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati (2003) ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลองโดยใช้การทดสอบของ (Box-pierce, 2003) ซึ่งจะแสดงได้โดยใช้ Q-statistic ดังในสมการที่ (10)

$$\text{Q-statistic} = n\sum_{k=1}^{m}\rho_k^2 \tag{10}$$

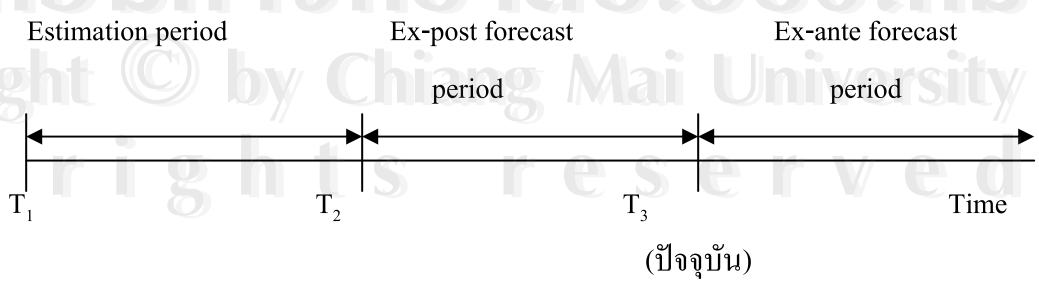
กำหนดให้ n คือ จำนวนของข้อมูล

M คือ ค่า Lag Length

จากสมการ ค่า Q-statistic ของแบบจำลองไม่แตกต่างกัน นั้นจะพบว่ามีการแจกแจงเป็นแบบ 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 (aucorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นปราศจากอัตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

4) การพยากรณ์ (forcecasting) เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากการวิเคราะห์ความถูกต้องแล้ว ก็สามารถนำแบบจำลองใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณีข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด

ดังนั้นการพยากรณ์จึงต้องมีการทดสอบแบบจำลองโดยการแบ่งการพยากรณ์ออกเป็น 3 ช่วง คือ ช่วง Historical forecast อันเป็นการพยากรณ์ตั้งแต่อดีตจนถึงช่วงเวลาที่พิจารณา (T_2) การพยากรณ์ช่วง Ex-post forecast คือการพยากรณ์โดยการตัดข้อมูลออกมาส่วนหนึ่งแล้วทำการพยากรณ์แล้วเปรียบเทียบข้อมูลจริงกับข้อมูลที่ได้จากการพยากรณ์ โดยพิจารณาค่า root mean square error (RMSE) และ Theil’s inequality coefficient (TIC) และค่า Akaike information criterion (AIC) จะพิจารณาค่าสถิติทั้งสามค่าที่มีค่าน้อยที่สุดซึ่งได้จากการทำการพยากรณ์เมื่อเลือกรูปแบบจำลองที่ดีที่สุดแล้วจึงนำแบบจำลองนั้นมาทำการพยากรณ์แบบ Ex-ante forecast ซึ่งเป็นการพยากรณ์ข้อมูลไปข้างหน้า ดังรูปที่ 3.2



รูป 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา : Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ราคาการส่งออกรถยนต์และส่วนประกอบในครั้งนี้จะใช้วิธีการพรรณนาอธิบายและวิธีการวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองให้กับอนุกรมเวลาในรูปแบบ ARIMA Model โดยวิธีของ Box Jenkins สามารถสรุปการสร้างแบบจำลองด้วย ARIMA (Gujarati, 2003) ได้ดังนี้คือ

1. การนำข้อมูลมาแจกแจง (plotting data)

การนำข้อมูลมาแจกแจง ด้วยวิธีการวาดกราฟ ระหว่าง Xt กับ t เพื่อที่จะพิจารณาแนวโน้มว่าข้อมูลมีเสถียรภาพหรือไม่ (stationary or nonstationary)

2. ขบวนการเปลี่ยนรูปแบบ (possibly transforming data)

ขบวนการเปลี่ยนรูปแบบหากข้อมูลที่วาดกราฟนั้นมีค่าเพิ่มขึ้นเรื่อยๆตามเวลาที่เปลี่ยนแปลงเพิ่มมากขึ้นหรือการทดสอบ unit root ของข้อมูล ถ้าข้อมูลนั้นพบว่ามี unit root แสดงว่าข้อมูลมีลักษณะไม่มีเสถียรภาพ ก็จะทำการเปลี่ยนแปลงรูปแบบโดยการทำ difference แต่ถ้าหากตรวจสอบแล้ว พบว่ามีเสถียรภาพ ก็ให้ทำในขั้นตอนถัดไป

3. การกำหนดลำดับขั้น P,Q ในแบบจำลอง (identiftying the dependence order of model) ขั้นตอนนี้คือการระบุว่าแบบจำลองนี้ควรจะมี autoregressive,p เท่าใด differenceing, d ที่ลำดับเท่าใด และ moving average , q เท่าใด โดยพิจารณาจาก ACF และ PACF ซึ่งอาจจะใช้ตาราง 3 ดังต่อไปนี้พิจารณาร่วม

ตาราง 3.2 ตารางแสดงการพิจารณา ACF และ PACF

ชนิดของแบบจำลอง	รูปแบบของ ACF	รูปแบบของ PACF
AR(q)	ลู่โค้งเข้าหาแกน(tails off)	เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป (cut off after lag p)
MA(q)	เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป(cut off after lag p)	ลู่ค้าเข้าหาแกน (tails off)
ARMA(p,q)	ลู่โค้งเข้าหาแกน(tails off)	ลู่โค้งเข้าหาแกน(tails off)

ที่มา: Gujarati (2003)

จากตารางข้างต้น จะสามารถกำหนดรูปแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็นค่าที่ p ของ AR (p) ยกตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF ที่มีแท่งคอเรลโลแกรม เกิดขึ้น

1 แท่ง แปลได้ว่าแบบจำลองควรมีลักษณะเป็น AR(1) สำหรับ MA(q) นั้นก็จะมี AFC 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) และเมื่อรวมกันกับการทดสอบความนิ่ง (stationary) ในขั้นตอนที่ 1 แล้ว จะสามารถหาค่าของ difference ได้ ซึ่งผลจากการ difference จำนวน d ครั้งนั้น ก็จะได้แบบจำลอง ARIMA(p,d,q)

4. การประมาณค่าพารามิเตอร์ (parameter estimation)

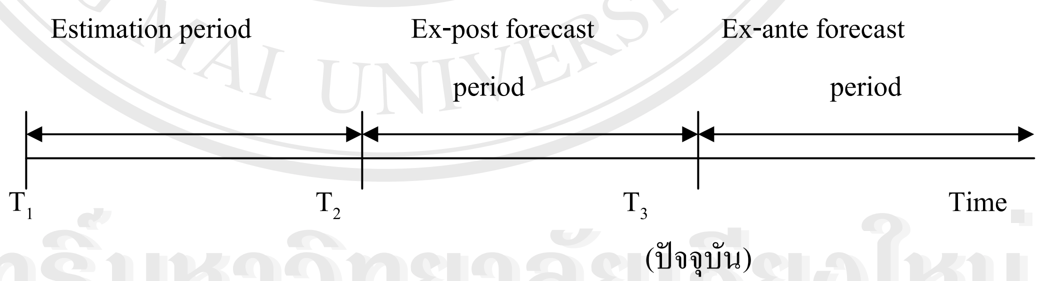
เมื่อพิจารณา ACF และ PACF แล้วให้สร้างสมการแบบจำลองที่มีความเหมาะสม (p,d,q) โดยเลือกสร้างไว้ ประมาณ 2-3 แบบจำลอง

5. การตรวจสอบความถูกต้อง (diagnostics)

การตรวจสอบความถูกต้องโดยการนำแบบจำลองที่เลือกไว้มาวาดกราฟเพื่อดูความเหมาะสม (Fit) กับข้อมูลเบื้องต้น ว่าสมการแบบจำลองใดสามารถแทนข้อมูลเบื้องต้นได้ดีที่สุด

6. การพยากรณ์ (forecasting)

การพยากรณ์จะทำได้ทั้งแบบจุด (point forecast) และแบบช่วง (interval forecast) ซึ่งข้อมูลที่ใช้ในการพยากรณ์เป็นข้อมูลรายเดือนดังรูป 7.1



รูป 7.1 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา : Pindyck and Rubinfeld (1997)

เมื่อได้พิจารณาแบบจำลองโดยเลือกเอาแบบจำลองที่ดีที่สุดที่สามารถเป็นตัวแทนข้อมูลเบื้องต้นได้แล้วก็นำแบบจำลองที่ได้ไปใช้ในการประมาณค่า ณ เวลาต่อไปได้