บทที่ 3

กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการศึกษานี้จะทำการพยากรณ์ราคาสัญญาล่วงหน้าแป้งมันสำปะหลังประเภทสตาร์ชชั้นพิเศษโดยวิธีของ Box and Jenkins ซึ่งเป็นวิธีที่นิยมใช้ในการพยากรณ์ข้อมูลที่เป็นอนุกรมเวลา เนื่องจากมีความแม่นยำและเหมาะสมสูงกว่าวิธีอื่นในการพยากรณ์ข้อมูลระยะสั้นในอนาคตซึ่งให้ค่าพยากรณ์ที่ใกล้เคียงกับความเป็นจริงมากที่สุด เป็นที่ยอมรับว่าเมื่อใดที่ต้องการศึกษาผลที่เกิดขึ้นในอนาคตวิธีการนี้มักจะถูกนำมาใช้ ซึ่งต้องใช้ข้อมูลอนุกรมเวลาที่มีลักษณะนิ่ง (stationary) โดยการศึกษาต้องทำการทดสอบ unit root ก่อนที่จะทำการพยากรณ์โดยวิธีอารีมาเพื่อให้ค่าความคลาดเคลื่อนจากการพยากรณ์ลดลง เนื่องจากข้อมูลเป็นอนุกรมเวลา (time series data) ส่วนมากมีลักษณะไม่นิ่ง (non-stationary)

3.1.1 การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

การวิเคราะห์นี้ คือการแยกความเคลื่อนไหวต่างๆ ออกจากข้อมูลอนุกรมเวลามักจะประกอบไปด้วย 4 ความเคลื่อนไหว(วินัส ฤาชัย, 2543)

1) ค่าแนวโน้ม (Secular Trend หรือ Long-term Movement : T) เป็นความเคลื่อนไหวในเวลาที่ค่อนข้างจะยาวนานของข้อมูลอนุกรมเวลาว่าจะโน้มเอียงไปทางใดทางหนึ่ง กล่าวคือไปในทางที่สูงขึ้นหรือต่ำลง ค่าแนวโน้มนี้ปกติแสดงถึงทิศทางที่อนุกรมเวลาชุดนั้นๆ มุ่งไปสู่ ค่าแนวโน้มอาจมีลักษณะเป็นเส้นตรงหรือเส้นโค้งหรือลักษณะอื่นใดก็ได้

2) การเคลื่อนไหวตามฤดูกาล (Seasonal หรือ Periodic Movement : S) เป็นความเคลื่อนไหวของข้อมูลที่เกิดขึ้นเนื่องจากอิทธิพลของฤดูกาลซึ่งจะเคลื่อนไหวขึ้นๆ ลงๆ ซ้ำกันในช่วงเวลาเดียวกันของแต่ละปี ความเคลื่อนไหวตามฤดูกาลนี้พบในข้อมูลในช่วงเวลาที่น้อยกว่าหนึ่งปี ราย 4 เดือน ราย 3 เดือน รายเดือนหรือรายสัปดาห์

3) การเคลื่อนไหวตามวัฏจักร (Cyclical Movement : C) เป็นการเคลื่อนไหวของข้อมูลที่เกิดขึ้นซ้ำๆ กันคล้ายกับความเคลื่อนไหวตามฤดูกาล เพียงแต่ความเคลื่อนไหวนี้เกิดขึ้นเป็นวัฏจักรในระยะเวลายาวมากกว่าหนึ่งปี และพบในข้อมูลรายปี วัฏจักรเหล่านี้มีแบบแผนไม่แน่นอนจึงยากที่จะพยากรณ์ ความเคลื่อนไหวตามวัฏจักรมีสาเหตุเกิดขึ้นจากสภาพทางเศรษฐกิจโดยทั่วไป การเปลี่ยนแปลงในนโยบายของรัฐบาล หรือการเปลี่ยนแปลงในรสนิยมของผู้บริโภคและนิสัยการจับจ่ายใช้สอย

4) ความเคลื่อนไหวผิดปกติ (Irregular Movement : I) เป็นการเคลื่อนไหวที่ไม่แน่นอน ไม่สามารถพยากรณ์โดยใช้ข้อมูลในอดีตเนื่องจากสาเหตุจากปัจจัยต่างๆ เช่น ดินฟ้าอากาศเปลี่ยนแปลง น้ำท่วม เกิดสงคราม ปฏิวัติรัฐประหาร การนัดหยุดงานของคนงาน การเลือกตั้ง เป็นต้น การเปลี่ยนแปลงนี้ไม่ได้เกิดขึ้นเป็นปกติประจำ

3.1.2 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลา

วิธีการของ Box – Jenkins เป็นการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลาโดยการใช้ค่า autocorrelation function (ACF) และค่า partial autocorrelation function (PACF) เป็นหลักในการพิจารณา รูปแบบที่ใช้อยู่ในกลุ่มของรูปแบบ integrated autoregressive 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} ,…, X_{t-p} หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA(q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต X_t จะขึ้นอยู่กับค่าคลาดเคลื่อน e_{t-1} , e_{t-2} , e_{t-3} ,…, e_{t-p} หรือค่าคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่า ซึ่งในที่นี้ค่าสังเกต X_t ซึ่งรูปแบบ ARMA(p,q) มีการกำหนดรูปแบบดังนี้

AR(p) คือ $X_t = \theta_0 + \phi_t X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t$ (3.1)

MA(q) คือ $X_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (3.2)

ARMA(p, q) คือ $X_t = \theta_0 + \phi_t X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (3.3)

ARIMA(p, q) คือ $\Delta^d X_t = \theta_0 + \phi_t X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (3.4)

อนุกรมเวลาที่จะนำมาศึกษาเพื่อใช้ในการพยากรณ์นั้นการเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่างๆ ได้แก่ แนวโน้ม ตัวแปรฤดูกาล ตัวแปรวัฏจักร และเหตุการณ์ที่ผิดปกติ โดยวิธี Box – Jenkins จะสามารถแบ่งอนุกรมเวลาออกเป็น 2 ประเภทดังนี้

1) อนุกรมเวลาที่เป็น stationary series คืออนุกรมเวลา $\{X_t\}$ ที่มีค่าเฉลี่ยและค่าความแปรปรวนของ X_t คงที่

2) อนุกรมที่ไม่เป็น stationary series เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น stationary series การจะหารูปแบบ ARMA(p, q) ให้กับอนุกรมเวลาดังกล่าวได้จะต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ stationary series เสียก่อน

3.1.3 การทดสอบความนิ่งของข้อมูล (Unit Root Test)

แบบจำลองที่ใช้ในการศึกษาของข้อมูลอนุกรมเวลา สำหรับทดสอบว่าตัวแปรแต่ละตัวมีลักษณะนิ่งหรือไม่ คือการทดสอบ unit root (ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์, 2542) ได้อธิบายว่าการทดสอบนั้นสามารถทำได้หลายวิธี ได้แก่ การทดสอบ ADF (Augmented Dickey-Fuller Test) (Said and Dickey, 1984) และการทดสอบ DF (Dickey-Fuller Test) (Dickey and Fuller, 1981) โดยสมมุติฐานว่าง (null hypothesis) ของการทดสอบ DF คือ $H_0 : \rho = 1$ จากสมการ (3.5) ด้านล่าง

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad (3.5)$$

ซึ่งเรียกว่าการทดสอบ unit root โดยถ้า $|\rho| < 1$ X_t จะมีลักษณะนิ่ง (stationary) และถ้า $\rho = 1$ แล้ว X_t จะมีลักษณะไม่นิ่ง (nonstationary) อย่างไรก็ตามการทดสอบนี้สามารถทำได้อีกทางหนึ่งซึ่งเหมือนกับสมการ (3.5) กล่าวคือ

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \qquad (3.6)$$

ซึ่งคือ $X_t = (1+\theta)X_{t-1} + \varepsilon_t$ ซึ่งคือสมการที่ (3.5) นั่นเอง โดยที่ $\rho = (1+\theta)$ ถ้า θ ในสมการ (3.6) มีค่าเป็นลบ จะได้ว่า ρ ในสมการ (3.5) 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 และถ้าเราไม่สามารถปฏิเสธ $H_0 : \theta = 0$ ได้ ก็จะหมายความว่า X_t มีลักษณะไม่นิ่ง

ถ้า X_t เป็นแนวเดินเชิงสุ่มซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย (random walk with drift) เราสามารถเขียนแบบจำลองได้ดังนี้

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

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

$$\Delta X_t = \alpha + \beta t + \theta X_{t-1} + \varepsilon_t \qquad (3.8)$$

ซึ่งก็จะทำการทดสอบ $H_0 : \theta = 0$ โดยมี $H_a : \theta < 0$ เช่นเดียวกับที่กล่าวมาข้างต้น โดยสรุปแล้ว Dickey and Fuller ได้พิจารณาสมการถดถอย 3 รูปแบบที่แตกต่างกันในการทดสอบว่ามี unit root หรือไม่ ซึ่ง 3 สมการดังกล่าว ได้แก่ สมการที่ (3.6), (3.7) และ (3.8) 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, 2003 : 769)

อย่างไรก็ตามค่าวิกฤติ (critical values) จะไม่เปลี่ยนแปลง ถ้าสมการ (3.6), (3.7), (3.8) ถูกแทนที่โดยกระบวนการเชิงอัตถดถอย (autoregressive processes)

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

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

$$\Delta X_{t-1} = \alpha + \beta t + \theta X_{t-1} + \sum_{i=1}^{\rho} \varphi \Delta X_{t-1} + \varepsilon_t \tag{3.11}$$

จำนวนของ lagged difference terms ที่จะนำเข้ามารวมในสมการนั้นมีมากพอที่จะทำให้พจน์ค่าความคลาดเคลื่อน (error terms) มีลักษณะเป็น serially independent และเมื่อนำเอาการทดสอบ DF มาใช้กับสมการ (3.9) – (3.11) เราจะเรียกว่าการทดสอบ ADF

ค่าสถิติทดสอบ ADF (ADF test statistic) มีการแจกแจงเชิงเส้นกำกับ (asymptotic distribution) เหมือนกับสถิติ DF ดังนั้นสามารถใช้ค่าวิกฤติ (critical values) แบบเดียวกัน(Gujarati, 2003 : 720) ในกรณีของการหา lag length ที่เหมาะสมนั้น (Enders, 1995: 227) ได้เสนอแนะว่าวิธีหนึ่งในการหา lag length ก็คือ เริ่มต้นด้วยการให้มี 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 นั้นมีค่าแตกต่างไปจากศูนย์ อย่างมีนัยสำคัญ

3.1.4 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 แล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วนถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมากจะสรุปได้ว่า $E(Y_t)$ คงที่

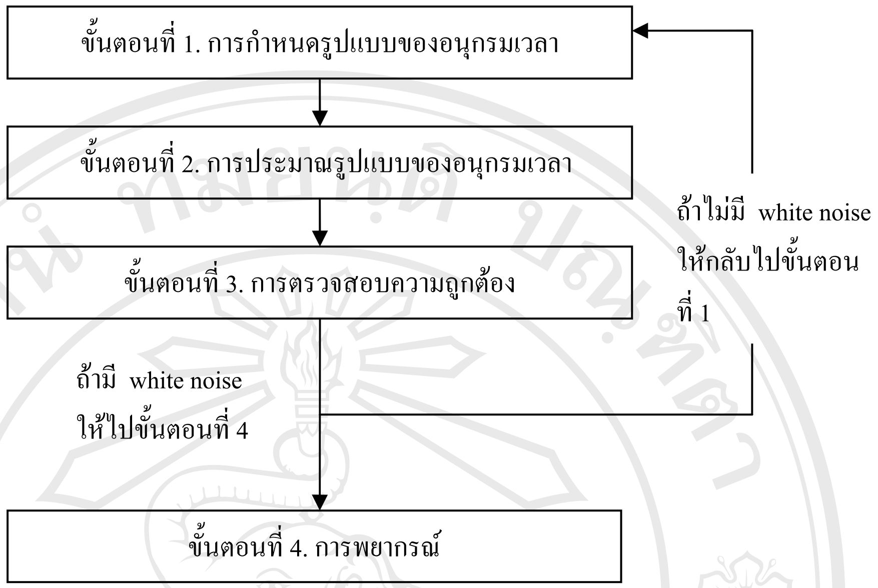
2) ค่าความแปรปรวน $V(Y_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆแล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วนถ้าค่าความแปรปรวนแต่ละส่วนย่อย ไม่แตกต่างกันมากนักจะสรุปได้ว่า $V(Y_t)$ คงที่

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) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างสูงที่ k = L, 2L, 3L จะเป็น ข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลของฤดูกาลและถ้าการเคลื่อนไหวของค่า คอเรลโลแกรมของสัมประสิทธิ์สหสัมพันธ์ (r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบภายใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วพบว่าอนุกรมเวลาที่ศึกษามีลักษณะไม่นิ่ง ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลานั้นจะต้องแปลงอนุกรมเวลาให้มีลักษณะนิ่งเสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่มีลักษณะนิ่ง ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจะได้อนุกรมเวลาที่มีลักษณะนิ่ง แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลาเดิมโดยการหา ลอการิทึม Z = In (Y_t) จนกว่าจะได้อนุกรมเวลาใหม่ที่มีความแปรปรวนคงที่ 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



รูป 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

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{3.12}$$

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

$$r_k = \frac{C_k}{C_0} \quad (3.13)$$

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

$$\rho_k = \frac{\sum_{t=a}^{n-k}(X_{t-q})(X_{t+k-q})}{\sum_{t=a}^{n}(X_{t-q})^2} \quad (3.14)$$

โดยที่ $X_t = \sum_{t=a}^{n}(X_t)$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

อย่างไรก็ตาม เนื่องจากอนุกรมเวลาจะเผชิญกับปัญหาสหสัมพันธ์ทั้งที่เกิดจากตัวแปรอิสระ ที่เป็นค่าความล่าช้า (lag) ของตัวแปรตาม (autoregressive) และสหสัมพันธ์ของค่าความคลาดเคลื่อน (moving average) เนื่องจาก autocorrelation function (ACF) จะใช้ในการอธิบายสหสัมพันธ์ของค่าความคลาดเคลื่อนแต่ไม่สามารถใช้อธิบายความสัมพันธ์ของตัวแปรอิสระที่เป็นค่าความล่าช้าของตัวแปรตาม ซึ่ง partial autocorrelation function (PACF) จะใช้วัดความสัมพันธ์ดังกล่าว ดังจะสามารถพิจารณาได้จากสมการ Yule-walker (Pindyck and Rublinfeld, 1997) ดังนี้

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p\rho \quad (3.15)$$

ถ้า k มากกว่า p จะได้สมการดังนี้

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \quad (3.16)$$

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 ซึ่งอาจจะใช้ตารางดังต่อไปนี้พิจารณาร่วม

ตาราง 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 q)	ลู่โค้งเข้าหาแกน (tails off)
ARMA(p,q)	ลู่โค้งเข้าหาแกน (tails off)	ลู่โค้งเข้าหาแกน (tails off)

ที่มา : Gujarati (2003)

จากตารางจะสามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ 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 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 ซึ่งหากค่า RMSE มีค่าเท่ากับ 0 จะหมายถึงแบบจำลองที่ประมาณได้มีค่าเท่ากันกับค่าจริงพอดี ดังนั้นหากว่าค่า RMSE มีค่าน้อยเพียงใดก็แสดงว่าแบบจำลองนั้นสามารถเป็นตัวแทนค่าจริงได้ดีมากเพียงนั้น สามารถพิจารณาสมการค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง (RMSE) ได้ดังนี้

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s - X_t^a)} \qquad (3.17)$$

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 แปลว่าแบบจำลองที่ประมาณได้เป็นแบบจำลองที่แย่ที่สุด ดังนั้นวิธีการพิจารณาค่าสถิตินี้ให้เลือกจากแบบจำลองที่มีค่า U ที่น้อยๆ ดังจะพิจารณาได้จากสมการดังนี้

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s - X_t^a)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t^a)^2}} \qquad (3.18)$$

กำหนดให้ X_t^s คือค่าที่ประมาณจากแบบจำลอง

X_t^a คือค่าข้อมูลจริง

T คือจำนวนของคาบเวลาที่ใช้ในการประมาณแบบจำลอง

อย่างไรก็ตาม ยังมีค่าสถิติอีกหลายอย่างที่สามารถนำมาพิจารณาประกอบร่วมกันกับ RMSE 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^2 มากขึ้นด้วย ซึ่งนับเป็นข้อจำกัดของค่าสถิตินี้ โดยสามารถพิจารณารูปแบบสมการได้จากสมการที่ (3.19)

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \tag{3.19}$$

ดังนั้นเพื่อปรับปรุงข้อจำกัดดังกล่าวข้างต้น จึงเกิดค่าสถิติใหม่ คือค่า adjusted R^2 ($\bar{R}^2$) ซึ่งจะมีการผกผันกันระหว่างตัวแปรที่เพิ่มเข้าไปกับค่า R^2 ที่ได้เพิ่มขึ้นมา ดังแสดงในสมการที่ (3.20)

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n - k)}{\sum y_i^2 / (n - 1)} \tag{3.20}$$

4) Akaike Information Criterion (AIC) คือค่าสถิติที่ประยุกต์คล้ายกับ $\bar{R}^2$ แต่ใช้รูปแบบการใส่ค่าลอกาลิทึมฐานธรรมชาติ (natural logarithm) โดยหากค่าสถิตินี้มีค่าน้อยเพียงใดนั้นก็แปลว่าแบบจำลองที่ประมาณได้นั้นสามารถเป็นตัวแทนข้อมูลจริงได้ดีเพียงนั้น ทั้งนี้ค่าสถิตินี้เหมาะที่จะนำไปใช้ในการหาค่าย้อนหลัง (lag length) ที่เหมาะสมอีกด้วย ซึ่งแสดงในสมการดังนี้

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

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

n คือค่าสังเกตทั้งหมด

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

$$SC = \log\left(\frac{\sum \hat{u}_i^2}{n}\right) + \left(\frac{2k \log n}{n}\right) \tag{3.22}$$

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

2. การประมาณค่าพารามิเตอร์ในรูปแบบ (Estimation) คือจะทำการประมาณค่าสัมประสิทธิ์ที่มาจากรูปแบบการถดถอยในตัวเอง (autoregressive; AR: p) และรูปแบบการเคลื่อนที่ของค่าคลาดเคลื่อน (moving average; MA: q) โดยสามารถเลือกใช้วิธีการถดถอยเชิงเส้นอย่างง่าย (simple least square) แต่สามารถที่จะใช้วิธีการถดถอยแบบไม่เป็นเชิงเส้น (nonlinear) ได้เพื่อสร้างความสัมพันธ์ของสมการที่จะสามารถนำไปใช้ในการพยากรณ์หากรูปแบบความสัมพันธ์นั้นเป็นรูปแบบที่มีความเหมาะสมที่สุด

3. การตรวจสอบแบบจำลอง (Diagnostics) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลองจะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบสามารถจะทำได้หลายวิธี ยกตัวอย่างเช่น การพิจารณาคอเรลโลแกรมของอัตสหสัมพันธ์ของกลุ่มตัวอย่าง (ρ_k) แต่อย่างไรก็ตาม Gujarati ได้เสนอการทดสอบวิเคราะห์ความเหมาะสมของแบบจำลองโดยใช้การทดสอบของ Box-pierce ซึ่งแสดงได้โดยใช้ Q-statistic ดังในสมการ

$$\text{Q-statistic} = n\sum_{k=1}^{m}\rho_k^2 \tag{3.23}$$

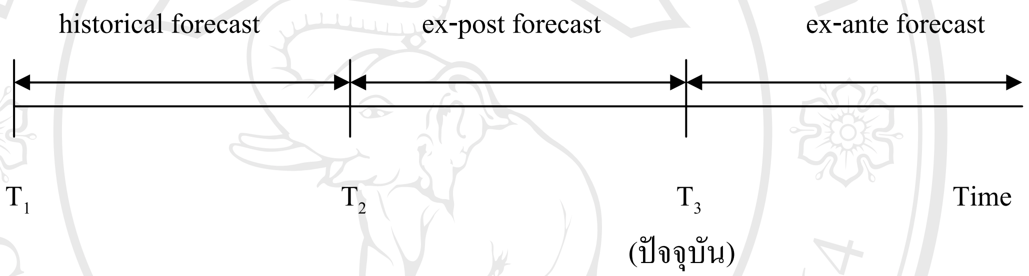
กำหนดให้ n คือ จำนวนของข้อมูล

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 (autocorrelation) ดังนั้นหากตรวจสอบพบว่าแบบจำลองนั้นปราศจากอัตสหสัมพันธ์แล้ว จะใช้แบบจำลองนั้นในการพยากรณ์ต่อไป แต่หากแบบจำลองนั้นไม่เหมาะสมต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบจำลองใหม่

4. การพยากรณ์ (Forecasting) เมื่อได้แบบจำลองที่เหมาะสมภายหลังจากการวิเคราะห์ความถูกต้องแล้ว ก็สามารถนำแบบจำลองใช้ในการพยากรณ์ แต่เนื่องจากการพยากรณ์ข้อมูลไปข้างหน้านั้นจะต้องใช้แบบจำลองที่ให้ค่าประมาณที่แม่นยำที่สุด

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) จะพิจารณาค่าสถิติทั้งสามค่าที่มีค่าน้อยที่สุดซึ่งได้จาการทำการพยากรณ์เมื่อเลือกรูปแบบจำลองที่ดีที่สุดแล้วจึงนำแบบจำลองนั้นมาทำการพยากรณ์แบบ ex-ante forecast ซึ่งเป็นการพยากรณ์ข้อมูลไปข้างหน้าดังรูป



รูป 3.2 แสดงช่วงเวลาที่ใช้ในการพยากรณ์

ที่มา : Pindyck and Rubinfeld (1997)

3.2 ระเบียบวิธีวิจัย

การพยากรณ์ราคาสัญญาล่วงหน้าแป้งมันสำปะหลังประเภทสตาร์ช ชั้นพิเศษโดยวิธีอารีมา เป็นการนำเอาข้อมูลอนุกรมเวลามาหารูปแบบแบบจำลองที่ดีที่สุดมีขั้นตอนดังนี้

1. การทดสอบความนิ่งของข้อมูล (unit root test) เพื่อทดสอบว่าราคาสัญญาล่วงหน้ามีลักษณะนิ่งหรือไม่

2. การกำหนดลำดับขั้น P,Q ในแบบจำลอง (identifying the dependence order of model) ขั้นตอนนี้คือการระบุว่าแบบจำลองนี้ควรจะมี autoregressive, p เท่าใด differencing, d ที่ลำดับเท่าใด และ moving average, q เท่าใด โดยพิจารณาจาก ACF และ PACF

3. การประมาณค่าพารามิเตอร์ (parameter estimation) เมื่อพิจารณา ACF และ PACF แล้วให้สร้างสมการแบบจำลองที่มีความเหมาะสม ARIMA(p,d,q)

4. การตรวจสอบความถูกต้อง (diagnostics) การตรวจสอบความถูกต้องเพื่อหาแบบจำลองที่เหมาะสมโดยการนำแบบจำลองที่เลือกไว้มาพิจารณาค่า Q-Statistic ค่า AIC และค่า SBC ว่าสมการแบบจำลองใดสามารถแทนข้อมูลเบื้องต้นได้ดีที่สุด

5. การพยากรณ์ (forecasting) การพยากรณ์จะทำได้ทั้งแบบจุด (point forecast) และแบบช่วง (interval forecast) ซึ่งจะทำการแบ่งการพยากรณ์ออกเป็น 3 ช่วงได้แก่ ช่วง historical forecast ช่วง ex-post forecast และช่วง ex-ante forecast

เมื่อได้พิจารณาแบบจำลองโดยเลือกเอาแบบจำลองที่ดีที่สุดที่สามารถเป็นตัวแทนข้อมูลเบื้องต้นได้แล้วก็นำแบบจำลองที่ได้ไปใช้ในการประมาณค่า ณ เวลาต่อไปได้