# บทที่ 3

# กรอบแนวคิดและระเบียบวิธีวิจัย

## 3.1 กรอบทฤษฎีแนวคิดของการศึกษา

การศึกษานี้จะทำการพยากรณ์ราคาผลิตภัณฑ์มันสำปะหลังที่คาดว่าจะเกิดขึ้นในอนาคต โดยใช้วิธีของบอกซ์แอนด์เจนกินส์ โดยการพยากรณ์จะนำอนุกรมเวลาในอดีตมาพยากรณ์อนุกรมเวลาในอนาคต ซึ่งเป็นวิธีการพยากรณ์ที่ให้ค่าความถูกต้อง (Accuracy) สูงกว่าวิธีอื่นในการพยากรณ์ระยะสั้น (Short term forecasting) และต้องใช้ข้อมูลอนุกรมเวลา (Time series data) ที่มีลักษณะนิ่ง (Stationary) โดยในการศึกษานี้ทำการทดสอบ Unit Root ก่อนที่จะทำการพยากรณ์โดยวิธีอารีมา เพื่อให้ค่าความคลาดเคลื่อนจากการพยากรณ์ลดลง เนื่องจากข้อมูลสินค้าเกษตรที่เป็นอนุกรมเวลา (Time series data) ส่วนมากมักมีลักษณะไม่นิ่ง (Non- stationary)

### 3.1.1 การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (Unit Root Test)

การทดสอบ Unit Root ของตัวแปรที่ใช้ในการศึกษาเนื่องจากข้อมูลอนุกรมเวลา (Time series data) มักจะมีความไม่นิ่งของข้อมูล (Non-stationary) ซึ่งการนำข้อมูลที่ไม่นิ่ง (Non-stationary) มาใช้วิเคราะห์ในสมการถดถอยจะทำให้เกิดความสัมพันธ์ที่ไม่แท้จริง (Spurious Regression) กล่าวคือค่าสถิติ $R^2$ มีค่าสูงมากและได้ค่าสถิติ t (T-statistic) มีนัยสำคัญหรือสูงเกินกว่าความเป็นจริงและ F-statistic ที่ได้จากสมการถดถอยที่เกิดความสัมพันธ์ที่ไม่แท้จริง (Spurious Regression) จะไม่ถูกต้องและไม่ควรนำมาใช้ เพราะมีการกระจายที่ไม่ได้มาตรฐานและตัวประมาณค่าที่ได้จากวิธีการ OLS จะไม่สามารถเชื่อถือได้ (Enders, 1995) โดยตัวแปรที่มีลักษณะนิ่ง (Stationary) จะมีคุณสมบัติ 3 ประการดังนี้

| | |
|---|---|
| ค่าเฉลี่ย (Mean) | $E(Y_t) = \mu$ |
| ความแปรปรวน (Variance) | $Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$ |
| ความแปรปรวนร่วม (Covariance) | $E[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k$ |

โดยที่ $Y_t$ คือ ตัวแปรที่มีลักษณะนิ่ง (Stationary)

การทดสอบ Unit Root เพื่อดูความนิ่ง (Stationary) หรือความไม่นิ่ง (Non-stationary) ของข้อมูลจากการศึกษาส่วนใหญ่ที่ผ่านมาจะนิยมการทดสอบ Unit Root โดยเป็นที่รู้จักกันดีในชื่อของ Dickey-Fuller test ซึ่งสามารถแบ่งออกได้เป็น 2 วิธี คือ

**1) Dickey - Fuller Test (DF)**

ทำการทดสอบตัวแปรที่เคลื่อนไหวไปตามช่วงเวลามีลักษณะเป็น Autoregressive model โดยสามารถเขียนรูปแบบของสมการได้ออกเป็น 3 รูปแบบคือ

กรณีตัวแปรที่ไม่มีค่าคงที่ $$X_t = \rho X_{t-1} + \varepsilon_t \quad (1)$$

กรณีตัวแปรมีค่าคงที่ $$X_t = \alpha_0 + \rho X_{t-1} + \varepsilon_t \quad (2)$$

กรณีตัวแปรมีทั้งค่าคงที่และแนวโน้มเวลา $$X_t = \alpha_0 + \alpha_2 t + \rho X_{t-1} + \varepsilon_t \quad (3)$$

โดยที่ $X_t$ คือตัวแปรที่เราทำการศึกษา $\alpha_0$ , $\rho$ คือ ค่าคงที่ t คือ แนวโน้มเวลา $\varepsilon_t$ คือตัวแปรสุ่ม มีการแจกแจงแบบปกติที่เป็นอิสระต่อกันและเหมือนกัน (Independent and Identical distribution) โดยมีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนคงที่ เขียนแทนด้วยสัญญาลักษณ์ $\varepsilon_t \sim iid\ (0,\sigma_\varepsilon^2)\,(0,\sigma_\varepsilon^2)$

ในการทดสอบว่า $X_t$ มีลักษณะเป็น Stationary process [$X_t \sim I(0)$] หรือไม่ ทำการทดสอบโดยการแปลงสมการทั้งสามรูปแบบให้อยู่ในรูปของ Frist differencing ($\Delta X_t$) ได้ดังนี้

$$\Delta X_t = X_t - X_{t-1} = \theta X_{t-1} + \varepsilon_t \quad (4)$$

$$\Delta X_t = X_t - X_{t-1} = \alpha_0 + \theta X_{t-1} + \varepsilon_t \quad (5)$$

$$\Delta X_t = X_t - X_{t-1} = \alpha_0 + \alpha_2 t + \theta X_{t-1} + \varepsilon_t \quad (6)$$

โดยที่ $\theta = (\rho - 1)$

**2) Augmented Dickey-Fuller Test (ADF)**

เป็นการทดสอบ Unit Root อีกวิธีหนึ่งที่พัฒนามาจาก DF Test เนื่องจากวิธี DF ไม่สามารถทำการทดสอบตัวแปรในกรณีที่เป็น Serial correlation ในค่า Error term ($\varepsilon_t$) ที่มีลักษณะความสัมพันธ์กันเองในระดับสูง ซึ่งจะมีการเพิ่ม Lagged change $\left[\sum_{j=1}^{p} \lambda_j \Delta X_{t-j}\right]$ เข้าไปในสมการทางด้านขวามือ จะได้ว่า

$$\Delta X_t = X_t - X_{t-1} = \theta X_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta X_{t-j} + \varepsilon_t \quad (7)$$

$$\Delta X_t = X_t - X_{t-1} = \alpha_0 + \theta X_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta X_{t-j} + \varepsilon_t \quad (8)$$

$$\Delta X_t = X_t - X_{t-1} = \alpha_0 + \alpha_2 t + \theta X_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta X_{t-j} + \varepsilon_t \quad (9)$$

ซึ่งพจน์ที่ใส่เข้าไปนั้น จำนวน Lagged term (p) ก็ขึ้นอยู่กับความเหมาะสมของข้อมูลอนุกรมเวลาของแต่ละงานวิจัย หรือต้องใส่จำนวน Lag เข้าไปกระทั่งไม่เกิดปัญหา Autocorrelation ในส่วนของ Error term (Pindyck and Rubinfeld, 1998)

สำหรับการเลือก Lag Length (P-lag) ที่เหมาะสมในการทดสอบ Unit Root ของตัวแปรนั้น Walter Enders (1995) กล่าวว่า ควรเริ่มต้นจาก Lag Length ที่สูงพอ เช่น P* แล้วดูว่าสัมประสิทธิ์ของ Lag Length P* แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติหรือไม่ โดยดูจากค่า t-statistic ถ้าพบว่าสัมประสิทธิ์ของ Lag Length P* นั้น ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติก็ทำการทดสอบ Unit Root ของตัวแปรนั้นโดยใช้ Lag Length P*-1 จนกระทั่ง Lag length ที่ใช้นั้นจะแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติ

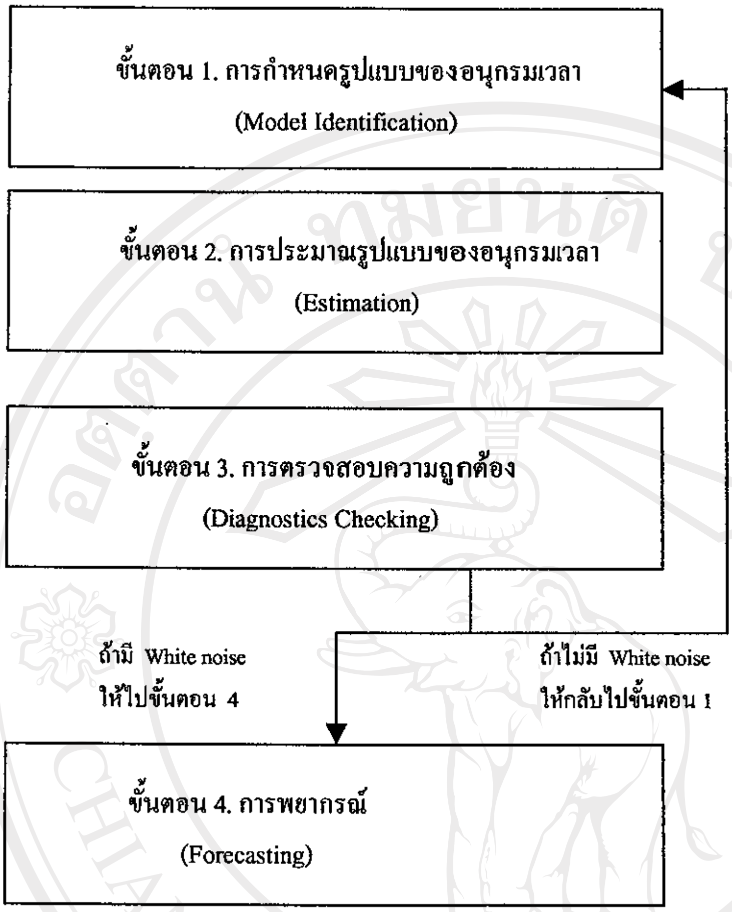
โดยในการทดสอบสมมุติฐานทั้งวิธี Dickey-Fuller test และวิธี Augmented Dickey-Fuller test ทดสอบว่าตัวแปรที่เราสนใจ ($X_t$) นั้นมี Unit Root หรือไม่ สามารถพิจารณาได้จากค่า $\theta$ ถ้าค่า $\theta$ มีค่าเท่ากับ 0 แสดงว่า $X_t$ นั้นมี Unit Root ซึ่งสามารถเขียนสมมติฐานในการทดสอบได้ดังนี้

$$H_0: \ \theta = 0$$
$$H_1: \ \theta < 0$$

จากสมมติฐาน ถ้ายอมรับ $H_0$ แสดงว่า ข้อมูลมีลักษณะไม่นิ่ง (Non-stationary) แต่ถ้าปฏิเสธ $H_0$ แสดงว่า ข้อมูลนั้นมีลักษณะนิ่ง (Stationary)

### 3.1.2 วิธีการพยากรณ์ โดยวิธีของ Box และ Jenkins

ซึ่งเมื่อข้อมูลมีลักษณะนิ่งแล้วก็จะทำการพยากรณ์ โดยวิธีของ Box และ Jenkins ซึ่งมี 4 ขั้นตอนสามารถแสดงได้ดังนี้




รูปที่ 3.1 แสดงขั้นตอนการพยากรณ์โดยวิธีของ Box และ Jenkins

**ขั้นตอน1 การกำหนดรูปแบบของอนุกรมเวลา (Identification)** ให้กับอนุกรมเวลาที่เป็น Stationary Series เป็นการหารูปแบบ ARIMA (p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลา โดยที่

Autocorrelation : $P_k$ คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยที่ $P_k$ มีค่าเท่ากับ $-1 \leq P_k \leq 1$ โดยพิจารณาเปรียบเทียบค่า Autocorrelation ($r_k$) ของอนุกรมเวลาตัวอย่าง กับค่า Autocorrelation ($P_k$) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_k = \frac{\sum_{t=a}^{n-k}(Y_{t-q})(Y_{t+k+q})}{\sum_{t=a}^{n}(Y_{t-q})^2} \qquad (10)$$

โดยที่ $Y_t = \sum_{t=a}^{n} (Y_t)$

$q$ = จำนวนเวลาสุดท้ายที่ย้อนหลัง

Partial Autocorrelation : $p_{kk}$ คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยพิจารณาเปรียบเทียบค่า Partial Autocorrelation ($r_{kk}$) ของอนุกรมเวลาตัวอย่างกับค่า Partial Autocorrelation ($p_{kk}$) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_{kk} = \frac{\sum_{j=1}^{k-1} (r_{k-1,j})(r_{k-j})}{\sum_{j-1}^{k-1} (r_{k-1,j})(r_j)} \qquad (11)$$

การพิจารณาเปรียบเทียบแต่ละรูปแบบ ต้องพิจารณา $r_k, r_{kk}$ กับ $p_k$ และ $p_{kk}$ พร้อมกันหลาย ๆ ค่า จึงมักจะพิจารณาจากรูปที่เรียกว่าคอเรลโลเกรม (Correlogram) ที่ได้จากการพล๊อต $r_k, r_{kk}$, $p_k$ และ $p_{kk}$ ในช่วงเวลา k ดังนั้นการพิจารณาเปรียบเทียบ จะเป็นการเปรียบเทียบ Correlogram ของค่า Autocorrelation ของอนุกรมเวลาตัวอย่าง ($r_k$) กับค่า Autocorrelation ของอนุกรมเวลาของประชากร ($p_k$) และ Correlogram ของค่า Partial Autocorrelation ของอนุกรมเวลาตัวอย่าง ($r_{kk}$) กับค่า Partial Autocorrelation ของอนุกรมเวลาประชากร ($p_{kk}$) สำหรับแต่ละรูปแบบจะมี Correlogram ของ $p_k$ และ $p_{kk}$ ต่างกัน อนุกรมเวลาที่จะนำมากำหนดรูปแบบจะต้องเป็นอนุกรมเวลาที่ Stationary เท่านั้น หากไม่เป็น Stationary จะต้องแปลงให้เป็น Stationary เสียก่อน

วิธีการวิเคราะห์อนุกรมเวลาของ Box – Jenkins เป็นการวิเคราะห์อนุกรมเวลาโดยการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา โดยใช้ค่า Autocorrelation Function (ACF) และค่า Partial Autocorrelation Function (PACF) เป็นหลักในการพิจารณาและรูปแบบที่เลือกใช้จะอยู่ในกลุ่มของรูปแบบ ARIMA (p,d,q) หรือเรียก Integrated Autoregressive – Moving Average order p and q ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากการสังเกตหรือค่าพยากรณ์ก่อนหน้า และความคลาดเคลื่อนของการพยากรณ์ก่อนหน้า โดยเป็นการรวมส่วนของรูปแบบ AR(p) และรูปแบบ MA(q) เข้าด้วยกัน รูปแบบ AR(p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต $Y_t$ จะขึ้นอยู่กับค่า $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA (q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกตุ $Y_t$ จะต้องอยู่กับค่าของความคลาดเคลื่อน $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$ หรือค่าความคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่าซึ่งรูปแบบ AR MA ( p , q ) โดยมีการกำหนดรูปแบบดังนี้

AR (p) คือ $Y_t = \theta_0 + \varnothing_1 Y_{t-1} + \ldots + \varnothing_p Y_{t-p} + \varepsilon_t$

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

AR MA (p, q) คือ $Y_t = \theta_0 + \varnothing_1 Y_{t-1} + \ldots + \varnothing_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$

อนุกรมเวลาที่จะนำมาศึกษาเพื่อประโยชน์ในการพยากรณ์นั้น การเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่าง ๆ ได้แก่ แนวโน้ม (Trend) ตัวแปรฤดูกาล (Seasonal factor) ตัวแปรวัฏจักร (Cyclical factor) และเหตุการณ์ที่ผิดปกติ (Irregular Movement) โดยวิธี Box — Jenkins จะสามารถแบ่งอนุกรมเวลาออกเป็น 2 ประเภทดังนี้

1) อนุกรมเวลาที่เป็น Stationary Series คืออนุกรมเวลา ($Y_t$) ที่มีค่าเฉลี่ยและค่าความแปรปรวนของ $Y_t$ คงที่ นั่นคือค่าเฉลี่ย $E(Y_t)$ และค่าความแปรปรวน $V(Y_t)$ มีค่าคงที่สำหรับอนุกรมแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้มและ/หรืออิทธิพลฤดูกาลจะมีค่าเฉลี่ย $E(Y_t)$ ไม่คงที่และอนุกรมเวลาที่มีความแปรปรวนของ $Y_t$ สูงจะเป็นลักษณะของอนุกรมเวลาที่ $V(Y_t)$ มีค่าไม่คงที่ซึ่งเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่เป็น Stationary Series นอกจากนั้นอนุกรมเวลาที่เป็น Stationary Series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้วยังจะต้องมีค่าสัมประสิทธิ์สหสัมพันธ์แบบออโตที่ lag k ขึ้นอยู่กับค่า k อย่างเดียว อนุกรมเวลาที่กำหนดรูปแบบ ARMA(p,q) ได้จะต้องเป็นอนุกรมเวลาที่เป็น Stationary Series แล้ว

2) อนุกรมเวลาที่ไม่เป็น Stationary Series เป็นอนุกรมเวลาที่ไม่มีคุณสมบัติเป็น Stationary Series การจะหารูปแบบ ARMA (p,q) ให้กับอนุกรมเวลาดังกล่าวได้จะต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ Stationary Series เสียก่อน การแปลงอนุกรมเวลาที่ไม่เป็น Stationary Series ให้เป็นอนุกรมเวลาที่เป็น Stationary Series อาจทำได้ด้วยวิธีการต่าง ๆ ดังนี้

2.1) การหาผลต่างปกติ (Regular differencing) ของอนุกรมเวลาเพื่อกำจัดแนวโน้มนั่นคือถ้าอนุกรมเวลา ($Y_t$) มีแนวโน้มอยู่ในอนุกรมเวลาจะแปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม ($Z_t$) โดย $Z_t = \Delta^d Y_t$ โดย d เป็นลำดับของการหารผลต่างและ $\Delta$ คือผลต่างของของตัวแปร เช่นเมื่อ d = 1 จะได้ $Z_t = \Delta Y_t = Y_t - Y_{t-1}$ เมื่อ d = 2 จะได้ $Z_t = \Delta^2 Y_t = \Delta (Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$ เป็นต้นจำนวนครั้งที่หาผลต่างจะขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป โดยทั่วไปถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ d = 1 อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติก จะใช้ d = 2

2.2) การหาผลต่างฤดูกาล ของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรฤดูกาลเข้ามาเกี่ยวข้อง จะต้องแปลงอนุกรมเวลาเดิม ($Y_t$) ให้เป็นอนุกรมใหม่ที่ไม่มีฤดูกาล ($Z_t$) โดย $Z_t = \Delta_L^{D} Y_1$ โดย D เป็นลำดับของการหาผลต่างฤดูกาล และ L เป็นจำนวนฤดูกาลต่อปี เช่น สำหรับอนุกรมเวลารายเดือน (L = 12) เมื่อ D = 1 จะได้ $Z_t = \Delta_{12} Y_t$ หรือ $Z_t = Y_t - Y_{t-12}$ และเมื่อ D =2 จะได้ $Z_t = \Delta^2_{12} Y_t$ หรือ $Z_t = \Delta^2 (Y_t - Y_{t-12}) = Y_t - 2Y_{t-2} + Y_{t-24}$ เป็นต้น ผลต่างนี้จะทำกี่ครั้งขึ้นกับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป

2.3) การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น Stationary Series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูการ d และ D ควบคู่กันไปซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และค่า D เป็นลำดับของการหาผลต่างฤดูกาล โดยที่ค่า d และ D จะมีค่าเท่าไรนั้นขึ้นอยู่กับว่าเมื่อหาผลต่างและผลต่างฤดูกาลแล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไปเช่น อนุกรมเวลารายเดือน ที่มีทั้งแนวโน้มและฤดูกาล เมื่อ d = 1 และ D = 1 จะแปลงอนุกรมเวลาเดิม ($Y_t$)ให้เป็นอนุกรมเวลาใหม่ ($Z_t$)โดย $Z_t = \Delta\Delta_{12} Y_t = \Delta(Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$ เป็นต้น

2.4) การหาลอการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือแปลงอนุกรมเวลาเดิม ($Y_t$) ให้เป็นอนุกรมเวลาใหม่ ($Z_t$) โดย $Z_t = \ln(Y_t)$ การแปลงนี้ทำเมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ นั่นคือ $V(Y_t)$ สำหรับค่าเวลา t ต่าง ๆ

โดยการพิจารณาACF และ PACF ซึ่งสามารถระบุได้ว่าแบบจำลองจะมี Autoregressive (p), Differencing (d) และ Moving average (q) ที่ลำดับเท่าใด โดยจะพิจารณาจากตาราง 3.1

ตาราง 3.1 แสดงการพิจารณา ACF และ PACF

| ชนิดของแบบจำลอง | รูปแบบของ ACF | รูปแบบของ PACF |
|---|---|---|
| AR(p) | - โค้งลู่เข้าหาแกน | - เกิดค่าที่ชัดเจนเพียง p ค่าแล้วหายไป |
| MA(q) | - เกิดค่าที่ชัดเจนเพียง q ค่าแล้วหายไป | - โค้งลู่เข้าหาแกน |
| ARMA(p,q) | - โค้งลู่เข้าหาแกน | - โค้งลู่เข้าหาแกน |

ที่มา: Gujarati (2003)

จากตาราง 3.1 สามารถกำหนดรูปแบบของแบบจำลองได้ โดยหากคอเรลโลแกรมของ ACF มีลักษณะโค้งลู่เข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม PACF เกิดมีค่าขึ้นมาไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมา ให้นับเป็นค่าที่ p ของ AR (p) ตัวอย่างเช่น เมื่อพิจารณาคอเรลโลแกรมของ ACF ที่โค้งลู่เข้าแกนระนาบและ PACF ที่มีแท่งคอเรลโลแกรมเกิดขึ้นหนึ่งแท่ง แสดงว่าแบบจำลองมีลักษณะเป็น AR(1) สำหรับ MA(q) นั้นก็จะมี ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไปในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบนั้น ตัวอย่างเช่น หากค่า ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียงสองแท่งและหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ แสดงว่าแบบจำลองมีลักษณะเป็น MA(2) และถ้าหาก ACF และ PACF โค้งลู่เข้าหาแกนระนาบทั้งคู่ก็แสดงว่าแบบจำลองมีลักษณะเป็น ARMA(p,q) และเมื่อรวมกับการทดสอบความนิ่ง (Stationary) ในขั้นตอนแรกแล้วจะสามารถหาค่าของ Difference ได้ซึ่งผลจากการ Difference จำนวน d ครั้งนั้นก็จะได้แบบจำลองมีลักษณะเป็น ARIMA (p,d,q)

**ขั้นตอน 2 การประมาณรูปแบบของอนุกรมเวลา (estimation)** จะทำได้โดยการหาค่าประมาณแบบง่ายหรือค่าประมาณที่ได้จากการวิเคราะห์ตัวเลข (numerical analysis) สำหรับค่าประมาณ แบบง่ายจะทำโดยการสร้างสมการที่มาจากความสัมพันธ์ระหว่าง $p_k$ และ พารามิเตอร์ โดยสมการที่สร้างขึ้นจะมีจำนวนเท่ากับจำนวนพารามิเตอร์ที่ต้องการประมาณ ส่วนค่าประมาณที่ได้จากการวิเคราะห์ตัวเลขจะได้จากการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด ขั้นตอนของการวิเคราะห์ตัวเลขจะต้องมีการกำหนดค่าประมาณเริ่มต้น ซึ่งส่วนใหญ่จะใช้การประมาณแบบง่ายเป็นค่าประมาณเริ่มต้นเมื่อการวิเคราะห์สิ้นสุดจะได้ค่าประมาณสุดท้ายที่จะนำไปใช้ประโยชน์ในการสร้างสมการพยากรณ์

โดยพิจารณา

**- Adjusted $R^2$ ( $R^{-2}$)** เป็นการอธิบายว่ารูปแบบจำลองที่ได้สามารถอธิบายตัวแปรได้มากน้อยแค่ไหน

$$R^{-2} = 1-\frac{RSS/(n-k)}{TSS/(n-1)} = 1-(1-R^2)\frac{n-1}{n-k}$$

**- Durbin-Watson statistic (DW)** ใช้ทดสอบ Autocorrelation แสดงความมีสหสัมพันธ์ในตัวเองโดยค่า DW เข้าใกล้ 2 ถือว่าไม่มีสหสัมพันธ์ในตัวเอง

**- Akaike Information Criterion (AIC)** เป็นการอธิบายว่ารูปแบบจำลองที่ได้มีค่าความคลาดเคลื่อนของผลการพยากรณ์มากน้อยแค่ไหน

$$\text{AIC} = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{RSS}{n} = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{RSS}{n}$$

โดยที่ n = จำนวนของค่าสังเกต

k = จำนวนของพารามิเตอร์ที่ทำการประมาณค่า

**- F-Statistic** เป็นการดูความผันแปรของตัวมันเองผันแปรต่อตัวมันเอง

**ขั้นตอน 3 การตรวจสอบความถูกต้อง (Diagnostic checking)** เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลองแล้ว จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธีได้แก่ การพิจารณาคอเรลโลเกรมของ $r_k$ หรือโดยพิจารณา White noise ของค่าประมาณการความคลาดเคลื่อน(estimated residual : $e_t$ ) ของรูปแบบอนุกรมเวลา (Box-Pierce, 2003) โดยพิจารณาจาก

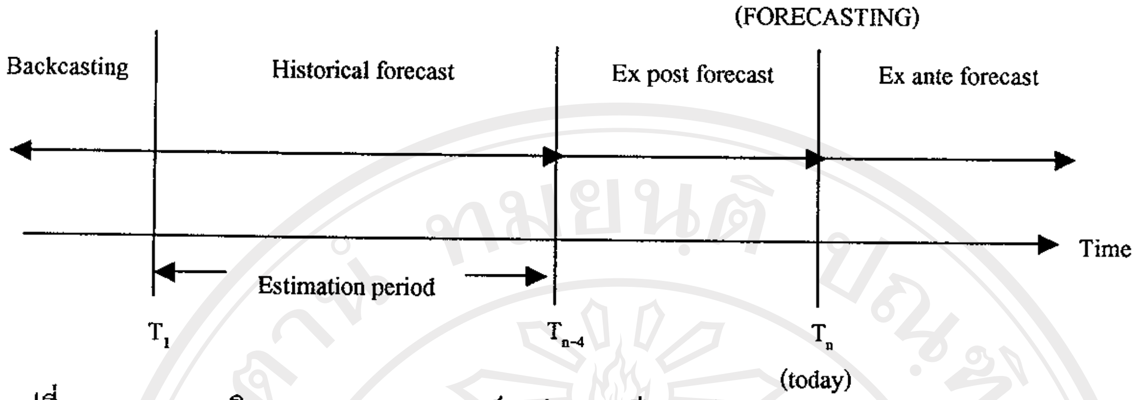
$$\text{Q-statistic} = n\sum_{k=1}^{m}\hat{\rho}_k^{2}$$

โดยที่ m = ค่า Lag Length

n = จำนวนของตัวอย่าง

หากพบว่า ค่า Q-statistic ของแบบจำลอง ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่า $e_t$ เป็น White noise หรือ $e_t$ มีการกระจายแบบปกติ (Normal distribution) มีค่าเฉลี่ย (mean) เท่ากับ 0 และค่าความแปรปรวนเท่ากับ $\sigma^2 I$ [$e_t \sim NID(0, \sigma^2 I)$] แสดงว่า $e_t$ ไม่มีสหสัมพันธ์ในตัวเอง (Autocorrelation) และไม่มีความแปรปรวนแตกต่างกัน หมายความว่า ตัวแปรอนุกรมเวลา ได้ผ่านการตรวจสอบความถูกต้อง (Diagnostics Checking) ดังนั้นรูปแบบจำลองที่คำนวณได้จึงมีความเหมาะสมต่อการนำไปใช้พยากรณ์ แต่หากพบว่าแบบจำลองที่กำหนดนั้นไม่เหมาะสมจะต้องทำตามขั้นตอน 1 เพื่อกำหนดรูปแบบของแบบจำลองใหม่

**ขั้นตอน 4 การพยากรณ์ (Forecasting)** จะพิจารณาการเลือกรูปแบบ (Model) ที่มีความเหมาะสมที่สุดเพื่อที่จะนำไปพยากรณ์นั้น โดยพิจารณา Root Mean Squared Error และค่า Theil's Inequality Coefficient ที่มีค่าต่ำสุด สามารถพิจารณาผลการพยากรณ์ออกเป็น 3 ช่วง ดังต่อไปนี้




รูปที่ 3.2 แสดงการพิจารณาผลการพยากรณ์ แบ่งออกเป็น 3 ช่วง

ที่มา : Pindyck and Rubinfeld (1998)

**ช่วงที่ 1 Historical forecast** คือ ช่วง $T_1$ ถึง ช่วง $T_{n-4}$ (รูปที่ 3.2) เป็นการพยากรณ์เพื่อเปรียบเทียบกับค่าของข้อมูลจริง ในการเลือกรูปแบบที่มีความเหมาะสมที่ โดยการพิจารณาค่า Root Mean Squared Error และค่า Theil's Inequality Coefficient ที่มีค่าต่ำสุด

โดยที่ ค่า Root Mean Squared Error $= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2}$ (12)

ค่า Theil's Inequality Coefficient $= \dfrac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^a)^2}}$ (13)

กำหนดให้ T = จำนวนตัวอย่างจากรูปแบบจำลอง

$Y_t^s$ = ค่าจากรูปแบบจำลอง

$Y_t^a$ = ค่าจากข้อมูลจริง

**ช่วงที่ 2 Ex post forecast** คือ ช่วง $T_{n-4}$ ถึง ช่วง $T_n$ (รูปที่ 3.2) เป็นการพยากรณ์เพื่อเปรียบเทียบกับค่าของข้อมูลจริง โดยใช้รูปแบบจากช่วง Historical forecast ในการเลือกรูปแบบที่มีความเหมาะสมที่สุด โดยการพิจารณาค่า Root Mean Squared Error และค่า Theil's Inequality Coefficient ที่มีค่าต่ำสุด

**ช่วงที่ 3 Ex ante forecast** คือ ช่วง $T_n$ ที่พยากรณ์ต่อไปข้างหน้า (รูปที่ 3.2) โดยใช้รูปแบบจากช่วง Ex post forecast ที่มีความเหมาะสมที่สุดในการพยากรณ์

## 3.2 ระเบียบวิธีวิจัย

### 3.2.1 การวิเคราะห์แบบจำลองการพยากรณ์ราคาผลิตภัณฑ์มันสำปะหลัง โดยวิธี Box-Jenkins

การพยากรณ์อนุกรมเวลาราคาผลิตภัณฑ์มันสำปะหลัง โดยวิธี Box - Jenkins ในรูปแบบ ARIMA (p,d,q) นั้นต้องพิจารณาตัวแปรของราคาผลิตภัณฑ์มันสำปะหลัง (อนุกรม Y) ให้มีลักษณะนิ่ง (Stationary) เสียก่อน โดยการพิจารณาว่าจะมีลักษณะนิ่ง (Stationary) หรือไม่พิจารณาได้จาก

1) ค่าเฉลี่ย $E(Y_t)$ คงที่สำหรับทุกค่าของ t หรือไม่ จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วน ๆ แล้วทำการหาค่าเฉลี่ยในแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อย ไม่แตกต่างกันมาก สรุปได้ว่า $E(Y_t)$ คงที่

2) ค่าความแปรปรวน $E(Y_t)$ คงที่ สำหรับทุก ๆ ค่าของ t หรือทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วทำการหาค่าความแปรปรวนแต่ละส่วน ถ้าค่าความแปรปรวนในแต่ละส่วนไม่มีความแตกต่างกันนัก แสดงว่า $E(Y_t)$ คงที่

3) พิจารณาแนวโน้มและปัจจัยฤดูกาล ว่ามีผลต่ออนุกรมหรือไม่

4) พิจารณาคอเรลโลแกรมของค่าสัมประสิทธิ์สหสัมพันธ์ของออโตของตัวอย่าง ($R_k$) กรณีที่อนุกรมเวลามีลักษณะนิ่ง (Stationary) ค่าคอเรลโลแกรมของ Autocorrelation ค่า $R_k$ จะมีค่าลดลงค่อนข้างเร็วเมื่อ k เพิ่มมากขึ้น ดังนั้นถ้าค่า Autocorrelation มีค่าลดลงค่อนข้างช้าจะเป็นข้อสังเกตได้ว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ค่า Autocorrelation มีค่าลดลงค่อนข้างช้าและมีค่าค่อนข้างสูงที่ k = L, 2L, 3L จะเป็นที่สังเกตได้ว่าอนุกรมชุดนี้มีแนวโน้มและอิทธิพลฤดูกาล

เมื่อพิจารณาการตรวจสอบแล้วว่า อนุกรมเวลาผลิตภัณฑ์มันสำปะหลัง ที่ประกอบด้วยราคามันเม็ดแข็งและแป้งมันสำปะหลังที่ศึกษามีลักษณะไม่นิ่ง (Non-Stationary) จะต้องแปลงให้มีลักษณะนิ่ง (Stationary) เสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ ให้แปลงอนุกรมเวลาเดิมโดยการหาลอการิทึม ($Z_t = \ln Y_t$) จนกว่าจะได้อนุกรมเวลาใหม่ ที่มีความแปรปรวนคงที่ การแปลงอนุกรมเวลาโดยการหาลอการิทึมของราคามันเม็ดแข็งและแป้งมันสำปะหลัง ซึ่งก็คือ $\ln HP_t$ และ $\ln S_t$ ตามลำดับ จากอนุกรมเวลาใหม่ที่มีลักษณะนิ่ง (Stationary) แล้วก็ทำตามขั้นตอนของ Box-Jenkins ต่อไป

**ขั้นตอน1 การกำหนดรูปแบบของอนุกรมเวลา (Identification)** ให้กับอนุกรมเวลาที่เป็น Stationary Series เป็นการหารูปแบบ ARIMA (p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลา โดยที่

Autocorrelation : $P_k$ คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยที่ $P_k$ มีค่าเท่ากับ $-1 \leq P_k \leq 1$ โดยพิจารณาเปรียบเทียบค่า Autocorrelation ($r_k$) ของอนุกรมเวลาตัวอย่าง กับค่า Autocorrelation ($P_k$) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_k = \frac{\sum_{t=a}^{n-k}(Y_{t-q})(Y_{t+k+q})}{\sum_{t=a}^{n}(Y_{t-q})^2} \qquad (14)$$

โดยที่ $Y_t = \sum_{t=a}^{n} (Y_t)$

$q$ = จำนวนเวลาสุดท้ายที่ย้อนหลัง

Partial Autocorrelation : $p_{kk}$ คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยพิจารณาเปรียบเทียบค่า Partial Autocorrelation ($r_{kk}$) ของอนุกรมเวลาตัวอย่างกับค่า Partial Autocorrelation ($p_{kk}$) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_{kk} = \frac{\sum_{j=1}^{k-1}(r_{k-1,j})(r_{k-j})}{\sum_{j-1}^{k-1}(r_{k-1,j})(r_j)} \qquad (15)$$

การพิจารณาเปรียบเทียบแต่ละรูปแบบ ต้องพิจารณา $r_k, r_{kk}$ กับ $p_k$ และ $p_{kk}$ พร้อมกันหลาย ๆ ค่า จึงมักจะพิจารณาจากรูปที่เรียกว่าคอเรลโลเกรม (Correlogram) ที่ได้จากการพล๊อต $r_k, r_{kk}, p_k$ และ $p_{kk}$ ในช่วงเวลา k ดังนั้นการพิจารณาเปรียบเทียบ จะเป็นการเปรียบเทียบ Correlogram ของค่า Autocorrelation ของอนุกรมเวลาตัวอย่าง ($r_k$) กับค่า Autocorrelation ของอนุกรมเวลาของประชากร ($p_k$) และ Correlogram ของค่า Partial Autocorrelation ของอนุกรมเวลาตัวอย่าง ($r_{kk}$) กับค่า Partial Autocorrelation ของอนุกรมเวลาประชากร ($p_{kk}$) สำหรับแต่ละรูปแบบจะมี Correlogram ของ $p_k$ และ $p_{kk}$ ต่างกัน อนุกรมเวลาที่จะนำมากำหนดรูปแบบจะต้องเป็นอนุกรมเวลาที่ Stationary เท่านั้น หากไม่เป็น Stationary จะต้องแปลงให้เป็น Stationary เสียก่อน

โดยการพิจารณาACF และ PACF ซึ่งสามารถระบุได้ว่าแบบจำลองจะมี Autoregressive (p), Differencing (d) และ Moving average (q) ที่ลำดับเท่าใด โดยจะพิจารณาจากตาราง 3.1

**ขั้นตอน 2 การประมาณรูปแบบของอนุกรมเวลา (estimation)** จะทำได้โดยการหาค่าประมาณแบบง่ายหรือค่าประมาณที่ได้จากการวิเคราะห์ตัวเลข (numerical analysis) สำหรับค่าประมาณ แบบง่ายจะทำโดยการสร้างสมการที่มาจากความสัมพันธ์ระหว่าง $p_k$ และ พารามิเตอร์ โดยสมการที่สร้างขึ้นจะมีจำนวนเท่ากับจำนวนพารามิเตอร์ที่ต้องการประมาณ ส่วนค่าประมาณที่ได้จากการวิเคราะห์ตัวเลขจะได้จากการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด ขั้นตอนของการวิเคราะห์ตัวเลขจะต้องมีการกำหนดค่าประมาณเริ่มต้น ซึ่งส่วนใหญ่จะใช้การประมาณแบบง่ายเป็นค่าประมาณเริ่มต้นเมื่อการวิเคราะห์สิ้นสุดจะได้ค่าประมาณสุดท้ายที่จะนำไปใช้ประโยชน์ในการสร้างสมการพยากรณ์

โดยพิจารณา

**- Adjusted $R^2$ ( $R^{-2}$)** เป็นการอธิบายว่ารูปแบบจำลองที่ได้สามารถอธิบายตัวแปรได้มากน้อยแค่ไหน

$$R^{-2} = 1-\frac{RSS/(n-k)}{TSS/(n-1)} = 1-(1-R^2)\frac{n-1}{n-k}$$

**- Durbin-Watson statistic (DW)** ใช้ทดสอบ Autocorrelation แสดงความมีสหสัมพันธ์ในตัวเองของค่าความคลาดเคลื่อน (Error term) โดยค่า DW เข้าใกล้ 2 ถือว่าไม่มีสหสัมพันธ์ในตัวเอง แสดงว่าแบบจำลองที่ใช้ในการศึกษาอยู่ในระดับที่น่าเชื่อถือได้

**- Akaike Information Criterion (AIC)** เป็นการอธิบายว่ารูปแบบจำลองที่ได้มีค่าความคลาดเคลื่อนของผลการพยากรณ์มากน้อยแค่ไหน ถ้ายิ่งมีค่าน้อยยิ่งดี

$$\text{AIC} = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{RSS}{n} = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{RSS}{n}$$

**- F-Statistic** เป็นการดูความผันแปรของตัวมันเองผันแปรต่อตัวมันเอง

**ขั้นตอน 3 การตรวจสอบความถูกต้อง (Diagnostic checking)** เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลองแล้ว จะต้อตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความ

เหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธีได้แก่ การพิจารณาคอเรลโลแกรมของ $r_k$ หรือโดยพิจารณา White noise ของค่าประมาณการความคลาดเคลื่อน(estimated residual : $e_t$ ) ของรูปแบบอนุกรมเวลา โดยพิจารณาจาก

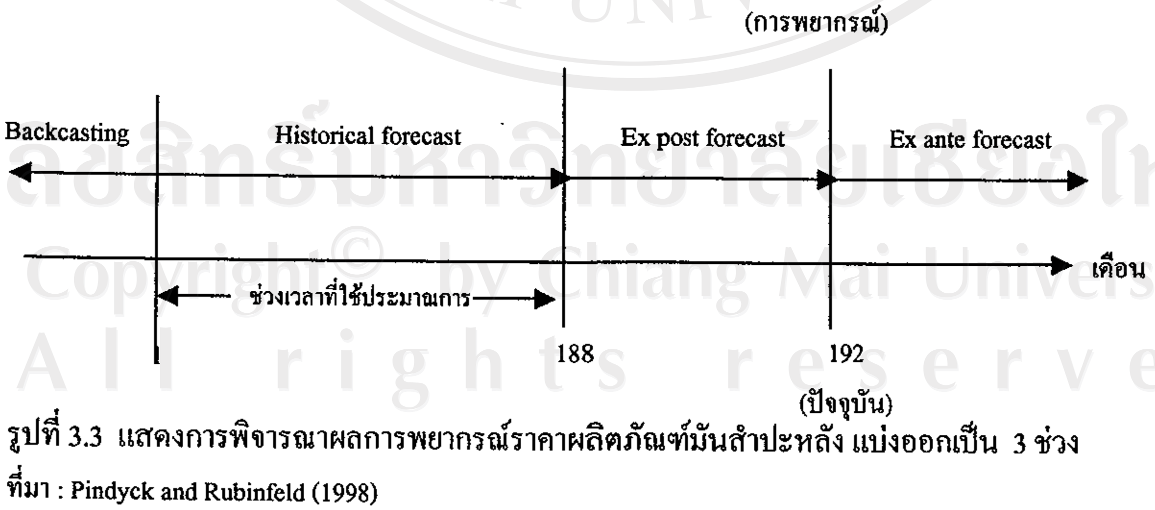
$$\text{Q-statistic} = n\sum_{k=1}^{m} \hat{\rho}_k^{2}$$

โดยที่ m = ค่า Lag Length

n = จำนวนของตัวอย่าง

หากพบว่า ค่า Q-statistic ของแบบจำลองไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญ แสดงว่า $e_t$ เป็น White noise หรือ $e_t$ มีการกระจายแบบปกติ (Normal distribution) มีค่าเฉลี่ย (mean) เท่ากับ 0 และค่าความแปรปรวนเท่ากับ $e_t$ $\sigma^2$ I [$e_t$ ~NID(0, $\sigma^2$ I)] แสดงว่า $e_t$ ไม่มีสหสัมพันธ์ในตัวเอง (Autocorrelation) และไม่มีความแปรปรวนแตกต่างกัน หมายความว่า ตัวแปรอนุกรมเวลา ได้ผ่านการตรวจสอบความถูกต้อง (Diagnostics Checking) ดังนั้นรูปแบบจำลองที่คำนวณได้จึงมีความเหมาะสมต่อการนำไปใช้พยากรณ์ แต่หากพบว่าแบบจำลองที่กำหนดนั้นไม่เหมาะสมจะต้องทำตามขั้นตอน 1 เพื่อกำหนดรูปแบบของแบบจำลองใหม่

**ขั้นตอน 4 การพยากรณ์ (Forecasting)** จะพิจารณาการเลือกรูปแบบ (Model) ที่มีความเหมาะสมที่สุดเพื่อที่จะนำไปพยากรณ์นั้น โดยพิจารณา Root Mean Squared Error และค่า Theil's Inequality Coefficient ที่มีค่าต่ำสุด สามารถพิจารณาผลการพยากรณ์ออกเป็น 3 ช่วง ดังต่อไปนี้




รูปที่ 3.3 แสดงการพิจารณาผลการพยากรณ์ราคาผลิตภัณฑ์มันสำปะหลัง แบ่งออกเป็น 3 ช่วง

ที่มา : Pindyck and Rubinfeld (1998)

**ช่วงที่ 1 Historical forecast** คือ ช่วง 1 ถึง ช่วง 188 (รูปที่ 3.3) เป็นการพยากรณ์เพื่อเปรียบเทียบกับค่าของข้อมูลจริง ในการเลือกรูปแบบที่มีความเหมาะสมที่ โดยการพิจารณาค่า Root Mean Squared Error และค่า Theil's Inequality Coefficient ที่มีค่าต่ำสุด

โดยที่ ค่า Root Mean Squared Error $= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2}$ (16)

ค่า Theil's Inequality Coefficient (U) $= \dfrac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^a)^2}}$ (17)

กำหนดให้ T = จำนวน 192 ตัวอย่าง

$Y_t^s$ = ค่าจากรูปแบบจำลอง

$Y_t^a$ = ค่าจากข้อมูลจริง

ค่า Theil's Inequality Coefficient (U) มักจะอยู่ระหว่างศูนย์ถึงหนึ่ง ถ้ารูปแบบจำลองมีค่าใกล้เคียงหรือเท่ากับค่าจากข้อมูลจริง ( $Y_t^s = Y_t^a$) จะทำให้ค่าค่า Theil's Inequality Coefficient (U) เข้าใกล้ศูนย์หรือเท่ากับศูนย์ หมายความว่า จะได้ Scale ที่มีความเหมาะสม

**ช่วงที่ 2 Ex post forecast** คือ ช่วง 189 ถึง ช่วง 192 (รูปที่ 3.3) เป็นการพยากรณ์เพื่อเปรียบเทียบกับค่าของข้อมูลจริง โดยใช้รูปแบบจากช่วง Historical forecast ในการเลือกรูปแบบที่มีความเหมาะสมที่สุด โดยการพิจารณาค่า Root Mean Squared Error และค่า Theil's Inequality Coefficient ที่มีค่าต่ำสุด

**ช่วงที่ 3 Ex ante forecast** คือ ช่วง 193 ที่พยากรณ์ต่อไปข้างหน้า (รูปที่ 3.3) โดยใช้รูปแบบจากช่วง Ex post forecast ที่มีความเหมาะสมที่สุดในการพยากรณ์