บทที่ 3

กรอบทฤษฎีและระเบียบวิธีวิจัย

3.1 ทฤษฎีที่เกี่ยวข้อง

วิธีการศึกษาพฤติกรรมการเคลื่อนไหวอนุกรมเวลาราคารายเดือนของสินค้าที่ศึกษา ใช้วิธีพรรณนาอธิบายและใช้วิธีวิเคราะห์เชิงปริมาณ โดยกำหนดแบบจำลองอารีมาตามวิธีของ Box — Jenkins พร้อมทั้งมีการอธิบายปัจจัยที่กำหนดการเคลื่อนไหวของราคาสินค้ารายเดือน

3.1.1 แนวคิดการพยากรณ์ข้อมูลอนุกรมเวลา

การศึกษาอนุกรมเวลาของราคาสินค้ารายเดือน ณ ตลาดต่างๆ โดยการกำหนดรูปแบบให้กับอนุกรมเวลา เพื่อพิจารณาความมีเสถียรภาพและลักษณะรูปแบบพฤติกรรมราคาว่ามีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้องกับสินค้าที่ต้องการศึกษาหรือไม่ รวมทั้งนำรูปแบบของสมการที่ได้ทำการวิเคราะห์อนุกรมเวลาแล้วนำมาทำนายราคาสินค้าที่ศึกษาไปล่วงหน้า เพื่อนำค่าพยากรณ์ดังกล่าวมาใช้ประโยชน์ในการจัดสรรทรัพยากรที่เหมาะสม

วิธีการวิเคราะห์อนุกรมเวลาของ Box — Jenkins นั้นเป็นวิธีการพยากรณ์ที่มีความถูกต้องและเหมาะสมกว่าวิธีอื่น ในการพยากรณ์ระยะสั้นในช่วงเวลา 1-4 เดือน หากต้องการพยากรณ์ในช่วงเวลาที่ยาวนานกว่านี้ควรใช้ข้อมูลที่ปรับค่าพยากรณ์แล้ว เพื่อให้ความคลาดเคลื่อนจากการพยากรณ์ลดลง

วิธีการของ Box — Jenkins เป็นการหารูปแบบที่เหมาะสมให้กับอนุกรมเวลา โดยการใช้ค่า Autocorrelation Fiction (ACF) และค่า Partial Autocorrelation Fiction (PACF) เป็นหลักในการพิจารณา รูปแบบที่เลือกใช้จะอยู่ในกลุ่มของรูปแบบ Integrated Autoregressive - Moving Average order p and q หรืออารีมา (p, d, q) ซึ่งเป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากการสังเกตหรือการพยากรณ์ล่วงหน้า และความคลาดเคลื่อนของการพยากรณ์ก่อนหน้าโดยเป็นการรวมส่วนของรูปแบบ AR (p) และ MA (q) เข้าด้วยกัน โดยที่รูปแบบ AR (p) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต SP_t จะขึ้นอยู่กับค่า SP_{t-1}, SP_{t-2}, SP_{t-3}, ..., SP_{t-p} หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ค่า ส่วนรูปแบบ MA (q) หมายถึงรูปแบบที่แสดงว่าค่าสังเกต SPt จะขึ้นอยู่กับค่าคลาดเคลื่อน e_{t-1}, e_{t-2}, e_{t-3} , ..., e_{t-q}หรือค่าความคลาดเคลื่อนที่เกิดขึ้นก่อนหน้า q ค่า ซึ่งในที่นี้ค่าสังเกต SP_t เป็นข้อมูลราคาส่งออกน้ำตาลทราย จำนวน 2 ชุดด้วยกัน จึงใช้สัญลักษณ์แทนข้อมูลทั้ง

2 ชุด ดังนี้ ราคาส่งออกน้ำตาลดิบใช้ SP_t และราคาน้ำตาลทรายขาวใช้ $SP1_t$ ซึ่งรูปแบบ ARMA (p, q) มีการกำหนดรูปแบบดังนี้

AR (p) คือ $SP_t = \theta_0 + \phi_t SP_{t-1} + ... + \phi_p SP_{t-p} + \varepsilon_t$

MA (q) คือ $SP_t = \theta_0 + \varepsilon_t - \theta_t \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

AR MA (p , q) คือ $SP_t = \theta_0 + \phi_t SP_{t-1} + ... + \phi_p SP_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$

ซึ่งในกรณีของข้อมูลน้ำตาลทรายขาว จะใช้การแทนด้วยสัญลักษณ์ $SP1_t$ ในลักษณะเดียวกันกับสมการข้างต้น

อนุกรมเวลาที่จะนำมาศึกษาเพื่อใช้ในการพยากรณ์นั้น การเคลื่อนไหวของอนุกรมเวลาขึ้นอยู่กับส่วนประกอบต่างๆ ได้แก่ แนวโน้ม (Trend) ตัวแปรฤดูกาล (Seasonal factor) ตัวแปรวัฏจักร (Cyclical factor) และเหตุการณ์ที่ผิดปกติ (Irregular Movement) โดยวิธี Box – Jenkins จะสามารถแบ่งอนุกรมเวลาออกเป็น 2 ประเภท ดังนี้

1. อนุกรมเวลาที่เป็น Stationary Series คืออนุกรมเวลา (SP_t) ที่มีค่าเฉลี่ยและค่าความแปรปรวนของ SP_t คงที่ กล่าวคือ ค่าเฉลี่ย E (SP_t) และค่าความแปรปรวน V (SP_t) มีค่าคงที่สำหรับอนุกรมแต่ละอนุกรมเวลา ซึ่งอนุกรมเวลาที่มีแนวโน้มและ/หรืออิทธิพลของฤดูกาลจะมีค่าเฉลี่ย E (SP_t) ไม่คงที่ และอนุกรมเวลาที่มีความแปรปรวนของ SP_t สูงจะเป็นลักษณะของอนุกรมเวลาที่ V (SP_t) มีค่าไม่คงที่ ซึ่งจะเรียกอนุกรมเวลาดังกล่าวนี้ว่า อนุกรมเวลาที่ไม่เป็น Stationary Series นอกจากอนุกรมเวลาที่เป็น Stationary Series จะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้วยังจะต้องมีค่าสัมประสิทธิ์สหสัมพันธ์ (Autocorrelation) ที่ lag k ขึ้นอยู่กับค่า k อย่างเดียว อนุกรมเวลาที่สามารถกำหนดรูปแบบ ARMA (p, q) ได้จะต้องเป็น Stationary Series แล้ว

2. อนุกรมเวลาที่ไม่เป็น Stationary Series เป็นอนุกรมที่ไม่มีคุณสมบัติเป็น Stationary Series (ข้อ1) การจะหารูปแบบ ARMA (p, q) ให้กับอนุกรมเวลาดังกล่าวได้จะต้องแปลงอนุกรมเวลาดังกล่าวให้เป็นอนุกรมเวลาใหม่ที่มีคุณสมบัติ Stationary Series เสียก่อน การแปลงอนุกรมเวลาที่ไม่เป็น Stationary Series ให้เป็นอนุกรมเวลาที่เป็น Stationary Series อาจทำได้ด้วยวิธีการต่าง ๆ ดังนี้

2.1 การหาผลต่างปกติของอนุกรมเวลาเพื่อกำจัดแนวโน้ม คืออนุกรมเวลา $(Y)_t$ ที่ มีแนวโน้มอยู่ในอนุกรมเวลา จะต้องแปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม (Z_t)โดย $Z_t = \Delta^d SP_t$ โดย d เป็นลำดับของการหาผลต่างปกติและ Δ คือผลต่างของตัวแปร เช่น เมื่อ d = 1 จะได้ $Z_t = \Delta SP_t = SP_t - SP_{t-1}$ เมื่อ d = 2 จะได้ $Z_t = \Delta^2 SPt = \Delta (SPt - SP_{t-1}) = \Delta SP_t - \Delta SP_{t-1} = SP_t - 2 SP_{t-1} + SP_{t-2}$ เป็นต้น จำนวนครั้งที่หาผลต่างจะขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่

เป็น Stationary Series หรือไม่ ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป โดยทั่วไปถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรงจะใช้ $d = 1$ อนุกรมเวลาที่มีแนวโน้มเป็นแบบควอดราติก จะใช้ $d = 2$

2.2 การหาผลต่างฤดูกาลของอนุกรมเวลา ถ้าอนุกรมเวลามีตัวแปรฤดูกาลเข้ามาเกี่ยวข้อง จะต้องแปลงอนุกรมเวลาเดิม (SP_t) ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล (Z_t) โดย $Z_t = \Delta_L^D\ SP_t$ โดย D เป็นลำดับของการหาผลต่างฤดูกาลและ L เป็นจำนวนฤดูกาลต่อปี เช่น สำหรับอนุกรมเวลารายเดือน (L = 12) เมื่อ D = 1 จะได้ $Z_t = \Delta_{12}\ SP_t$ หรือ $Z_t = SP_t - SP_{t-12}$ และเมื่อ D = 2 จะได้ $Z_t = \Delta_{12}^2\ SP_t$ หรือ $Z_t = \Delta^2\ (SP_t - SP_{t-12}) = SP_t - 2\ SP_{t-12} + SP_{t-24}$ เป็นต้น ผลต่างนี้จะทำกี่ครั้งขึ้นกับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็น Stationary Series หรือไม่ถ้ายังไม่เป็น Stationary Series ต้องหาผลต่างต่อไป

2.3 การหาผลต่างปกติและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและตัวแปรฤดูกาล การปรับให้อนุกรมเวลาเป็น Stationary Series นั้นจะทำได้โดยการหาผลต่างปกติและผลต่างฤดูกาลควบคู่กันไป ซึ่งค่า d เป็นลำดับของการหาผลต่างปกติ และค่า D เป็นลำดับของการหาผลต่างฤดูกาล ค่า d และ D จะมีจำนวนครั้งเท่าใดนั้นขึ้นอยู่กับการหาผลต่างปกติและผลต่างฤดูกาลจนกว่าอนุกรมเวลาใหม่จะเป็น Stationary Series เช่น อนุกรมเวลารายเดือนที่มีทั้งแนวโน้มและฤดูกาล เมื่อ d = 1 และ D = 1 จะแปลงอนุกรมเวลาเดิม (SP_t) ให้เป็นอนุกรมเวลาใหม่ (Z_t) โดย $Z_t = \Delta\Delta_{12}\ SP_t = \Delta\ SP_t - SP_{t-12} = SP_t - SP_{t-1} - SP_{t-12} + SP_{t-13}$ เป็นต้น

2.4 การหาลอการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือแปลงอนุกรมเวลาเดิม (SP_t) ให้เป็นอนุกรมเวลาใหม่ (Z_t) โดย $Z_t = \ln(SP_t)$ การแปลงอนุกรมเวลาลักษณะนี้จะทำเมื่อความแปรปรวน $V(SP_t)$ ของอนุกรมเวลาไม่คงที่

3.1.2 การทดสอบความนิ่งของข้อมูล (Unit Root Test)

แบบจำลองที่ใช้ในการศึกษาของข้อมูลอนุกรมเวลา สำหรับทดสอบว่าตัวแปรแต่ละตัวมีลักษณะนิ่ง (Stationary) หรือไม่ คือการทดสอบ Unit Root ทรงศักดิ์ ศรีบุญจิตต์ และอารี วิบูลย์พงศ์ (2542) กล่าวว่า การประมาณค่าทางเศรษฐมิติโดยใช้ข้อมูลอนุกรมเวลานั้น มีข้อสมมุติ (Assumptions) เกี่ยวกับความนิ่ง (Stationary) ของข้อมูล สมมุติว่าแบบจำลองเป็นดังนี้

$$SP_t = \alpha + \beta X_t + u_{1t} \qquad (1)$$

$$X_t = X_{t-1} + u_{2t} \quad ; \quad u_{2t} \sim iid(0, \sigma^2_{u2}) \qquad (2)$$

โดยที่ u_{2t} เป็นอนุกรมเวลาของตัวแปรสุ่ม (Random Variables) ที่มีการแจกแจงปกติที่เหมือนกันและเป็นอิสระต่อกัน มีค่าเฉลี่ย (Mean) เท่ากับศูนย์ และค่าความแปรปรวน (variance) คงที่ ซึ่งตัวแปร X เป็นแนวเดินเชิงสุ่ม (Random Walk) และเป็น integrated of order one, I (1)

เพราะฉะนั้นตัวแปร SP_t ก็จะเป็น I (1) ด้วย ตามทฤษฎีเศรษฐมิตินั้น การถดถอยด้วยตัวแปรที่มีลักษณะไม่นิ่ง (Nonstationary) ค่าสถิติ t ที่ใช้โดยปกติจะมีการแจกแจงไม่มาตรฐาน (Nonstandard Distribution) ดังนั้นการใช้ตารางมาตรฐานสำหรับการทดสอบค่าสถิติแบบที่ใช้โดยทั่วไป อาจทำให้เกิดการสรุปที่ผิดพลาดและเกิดการถดถอยที่ไม่ถูกต้อง (Spurious Regression) (Jonhston and Dinardo, 1997)

การทดสอบ Unit Root นั้นทำได้หลายวิธี ได้แก่ การทดสอบ ADF (Augmented Dickey-Fuller (ADF) test) (Said and Dickey, 1984) และ DF (Dick-Fuller (DF) test) (Dickey and Fuller, 1981) โดยมีสมมุติฐานว่าง (Null Hypothesis) ของการทดสอบ DF (Dick-Fuller (DF) test) คือ $H_0 : \rho = 1$ จากสมการ (3)

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{3}$$

โดยถ้าการทดสอบพบว่า $|\rho|<1$ แล้ว แสดงว่า X_t จะมีลักษณะนิ่ง แต่ถ้า $\rho = 1$ แล้ว X_t จะมีลักษณะไม่นิ่ง นอกจากนี้ยังสามารถทดสอบได้อีกทางหนึ่งซึ่งคล้ายกับสมการ (3)

$$\Delta X_t = \phi X_{t-1} + \varepsilon_t \tag{4}$$

สมมุติฐานว่าง (Null Hypothesis) ของการทดสอบ คือ $H_0 : \theta = 0$ และ $H_a: \theta < 0$ โดยที่ $\rho = (1 + \theta)$ ซึ่งถ้า θ ในสมการ (4) มีค่าเป็นลบ แสดงว่า $\rho < 1$ (ρ จากสมการ (3)) จึงสรุปได้ดังนี้ ถ้าปฏิเสธ $H_0 : \theta = 0$ หรือยอมรับ $H_a: \theta < 0$ หมายถึง X_t มีลักษณะนิ่ง หรือมี integration of order zero (Charemza and Deadman, 1992) ในทางตรงกันข้าม ถ้ายอมรับ H_0 จะหมายความว่า X_t มีลักษณะไม่นิ่ง โดยการพิจารณาค่าสถิติ DF เปรียบเทียบกับค่าวิกฤต MacKinnon (Gujarati, 2003: 815) หรือค่าวิกฤตจากตาราง Dick-Fuller (Ender, 1995: 221) โดยถ้า X_t มีแนวคิดเดินเชิงสุ่ม (random walk) ซึ่งมีความโน้มเอียงทั่วไปรวมอยู่ด้วย สามารถเขียนแบบจำลองได้ดังสมการที่ (5) และถ้ามีแนวโน้มตามเวลาเชิงเส้นรวมอยู่ด้วยจะได้แบบจำลองดังสมการที่ (6) ตามลำดับ

$$\Delta X_t = \alpha + \phi X_{t-1} + \varepsilon_t \tag{5}$$

$$\Delta X_t = \alpha + \beta_t + \phi X_{t-1} + \varepsilon_t \tag{6}$$

โดยที่ t = เวลา

ถ้าหากสมการที่ (4) (5) และ (6) มีกระบวนการเชิงอัตถดถอยเข้ามาเกี่ยวข้องแล้ว จะเรียกว่า การทดสอบโดยวิธี Augmented Dickey — Fuller test (ADF test) ได้สมการเป็นดังนี้

$$\Delta X_t = \phi X_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta X_{t-1} + \varepsilon_t \qquad (7)$$

$$\Delta X_t = \alpha + \phi X_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta X_{t-1} + \varepsilon_t \qquad (8)$$

$$\Delta X_t = \alpha + \beta_t + \phi X_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta X_{t-1} + \varepsilon_t \qquad (9)$$

ซึ่งการทดสอบแบบดังกล่าวมีการแจกแจงเหมือนกับการทดสอบ DF จึงสามารถใช้ค่าวิกฤตในการพิจารณาแบบเดียวกัน ส่วนการหา Lag Length ตามวิธีของ Enders (Enders, 1995: 227) นั้นให้เริ่มต้นที่ Lag Length ที่มากพอสมควรค่าหนึ่ง ในที่นี้ใช้ค่า Lag Length เริ่มตั้งแต่ 4 เนื่องจากข้อมูลที่ใช้เป็นข้อมูลรายเดือน ซึ่งในเวลา 1 ปีประกอบด้วย 4 ไตรมาส จากนั้นค่อยๆ ลดค่า Lag Length ลงเรื่อยๆ เมื่อพบว่าค่าสถิติ t-test หรือ F - test ที่ใช้ในการทดสอบนั้นไม่มีนัยสำคัญ ณ ระดับที่กำหนด จนกระทั่ง ค่าสถิติมีนัยสำคัญ จึงจะถือว่าค่า Lag นั้นมีความเหมาะสม

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 $E(SP_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าเฉลี่ยของอนุกรมเวลาแต่ละส่วน ถ้าค่าเฉลี่ยแต่ละส่วนย่อยไม่แตกต่างกันมากจะสรุปได้ว่า $E(SP_t)$ คงที่

2) ค่าความแปรปรวน $V(SP_t)$ คงที่ สำหรับทุกค่าของ t หรือไม่จะทำได้โดยการแบ่งอนุกรมเวลาออกเป็นส่วนๆ แล้วหาค่าความแปรปรวนของอนุกรมเวลาแต่ละส่วน ถ้าค่าความแปรปรวนแต่ละส่วนย่อย ไม่แตกต่างกันมากนักจะสรุปได้ว่า $V(SP_t)$ คงที่

3) พิจารณาแนวโน้มและปัจจัยฤดูกาล ด้วยการวาดกราฟอนุกรมเวลาในกรณีที่มีแนวโน้มและปัจจัยฤดูกาลมักจะเห็นชัดเจนได้จากรูปที่เรียกว่าคอเรลโลแกรม (Correlogram)

4) พิจารณาคอเรลโลแกรม ของสัมประสิทธิ์สหสัมพันธ์ ของตัวอย่าง (r_k) กรณีที่อนุกรมเวลาเป็นแบบ Stationary ค่าคอเรลโลแกรม ของสัมประสิทธิ์สหสัมพันธ์ (r_k)จะมีค่าลดลงค่อนข้างเร็ว เมื่อ k มีค่าเพิ่มขึ้นมาก ดังนั้นถ้าค่าสัมประสิทธิ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้าจะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้ม แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ (r_k) มีค่าลดลงค่อนข้างช้า และมีค่าค่อนข้างสูงที่ k = L , 2L , 3L จะเป็นข้อสังเกตว่าอนุกรมเวลาชุดนี้มีแนวโน้มและอิทธิพลของฤดูกาล และถ้าการเคลื่อนไหวของค่าคอเรลโลแกรมของสัมประสิทธิ์สหสัมพันธ์

(r_k) มีลักษณะคล้ายลูกคลื่น โดยคลื่นจะครบรอบใน 2 ช่วงเวลา แสดงว่าอนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง

เมื่อพิจารณาจากการตรวจสอบแล้วว่า อนุกรมเวลาที่ศึกษาไม่เป็น Stationary ก่อนที่จะทำการกำหนดรูปแบบให้กับอนุกรมเวลาที่ไม่เป็น Stationary จะต้องแปลงอนุกรมเวลาที่เป็น Stationary เสียก่อน โดยการหาผลต่างสำหรับอนุกรมเวลาที่มีแนวโน้ม ถ้าอนุกรมเวลาที่มีอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary ถ้าอนุกรมเวลามีทั้งแนวโน้มและอิทธิพลฤดูกาลให้หาผลต่างฤดูกาลจนได้อนุกรมเวลาที่เป็น Stationary แต่ถ้าอนุกรมเวลามีความแปรปรวนไม่คงที่ ให้แปลงอนุกรมเวลาเดิมโดยการหา ลอการิทึม $Z_t = \text{In}(SP_t)$ จนกว่าจะได้อนุกรมเวลาใหม่ ที่มีความแปรปรวนคงที่ จากอนุกรมเวลาใหม่เป็น Stationary Series แล้วจะทำตามขั้นตอนของ Box — Jenkins ดังนี้

ขั้นตอนการพยากรณ์โดยวิธีของ Box และ Jenkins ได้แก่

1. การกำหนดแบบจำลอง (identification) ให้กับอนุกรมเวลาที่เป็น Stationary Series เป็นการหารูปแบบ ARMA (p, q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยที่สัมประสิทธิ์สหสัมพันธ์ : P_k คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยที่ P_k มีค่าเท่ากับ $-1 \leq P_k \leq 1$ โดยพิจารณาเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ (r_k) ของอนุกรมเวลาตัวอย่าง กับค่าสัมประสิทธิ์สหสัมพันธ์ (P_k) ของอนุกรมเวลาของประชากร ที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

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

โดยที่ $Y_t = \sum_{t=a}^{n} (Y_t)$

q = จำนวนเวลาสุดท้ายที่ย้อนหลัง

สัมประสิทธิ์สหสัมพันธ์บางส่วน : P_{kk} คือการวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลา โดยพิจารณาเปรียบเทียบค่า สัมประสิทธิ์สหสัมพันธ์บางส่วน (r_{kk}) ของอนุกรมเวลาของตัวอย่างที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

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

การพิจารณาเปรียบเทียบแต่ละรูปแบบ ต้องพิจารณา r_k , r_{kk} กับ P_k และ P_{kk} พร้อมกันหลาย ๆ ค่า จึงมักจะพิจารณาจาก คอเรลโลแกรม ที่ได้จากการพล็อต r_k, r_{kk}, P_k และ P_{kk} ในช่วงเวลา k ดังนั้นการพิจารณาเปรียบเทียบ จะเป็นการเปรียบเทียบ คอเรลโลแกรม ของค่า สัมประสิทธิ์สหสัมพันธ์ของอนุกรมเวลาตัวอย่าง (r_k) กับค่าสัมประสิทธิ์สหสัมพันธ์ ของอนุกรมเวลาของประชากร (P_k) และ คอเรลโลแกรม ของค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนของอนุกรมเวลาตัวอย่าง (P_{kk}) สำหรับแต่ละรูปแบบจะมี คอเรลโลแกรม ของ P_k และ P_{kk} ต่างกัน อนุกรมเวลาที่จะนำมากำหนดรูปแบบจะต้องเป็นอนุกรมเวลาที่ Stationary เท่านั้นหากไม่เป็น Stationary จะต้องแปลงให้เป็น Stationary เสียก่อน

2. การประมาณค่าพารามิเตอร์ในรูปแบบ (Estimation) จะทำได้โดยการหาค่าประมาณแบบง่าย หรือค่าประมาณที่ได้จากการวิเคราะห์ตัวเลข (Numerical Analysis) สำหรับค่าประมาณแบบง่ายจะทำโดยการสร้างสมการที่มาจากความสัมพันธ์ระหว่าง P_k และพารามิเตอร์ โดยสมการที่สร้างขึ้นจะมีจำนวนเท่ากับจำนวนพารามิเตอร์ที่ต้องการประมาณ ส่วนค่าประมาณที่ได้จากการวิเคราะห์ตัวเลขจะได้จากการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด ขั้นตอนของการวิเคราะห์ตัวเลขจะต้องมีการกำหนดค่าประมาณเริ่มต้น ซึ่งส่วนใหญ่จะใช้การประมาณแบบง่ายเป็นค่าประมาณเริ่มต้นเมื่อการวิเคราะห์สิ้นสุด ซึ่งจะได้ค่าประมาณสุดท้ายที่จะนำมาใช้ประโยชน์ในการสร้างสมการพยากรณ์

3.2 ระเบียบวิธีวิจัย

การวิเคราะห์และพยากรณ์ราคาส่งออกน้ำตาลรายเดือนของประเทศไทย ด้วยแบบจำลองอารีมา ตามวิธีของ Box-Jenkins ใช้วิธีการพรรณนาและวิธีการวิเคราะห์เชิงปริมาณ สามารถสรุปการสร้างแบบจำลองด้วยวิธีอารีมา (Gujarati, 2003) ได้ดังนี้คือ

1. **การกำหนดแบบจำลอง (Identification)** เมื่ออนุกรมเวลานั้นมีลักษณะนิ่งแล้ว โดยการทำข้อมูลให้อยู่ในรูปของลอการิธึมธรรมชาติ (Natural Log) ก่อนจึงทดสอบความนิ่งของข้อมูล จากนั้นพิจารณาว่าแบบจำลองนี้ควรจะมี Autoregressive, p เท่าใด Differencing, d ที่ลำดับเท่าใด และ Moving average, q เท่าใด โดยพิจารณาจาก ACF และ PACF ซึ่งอาจจะใช้ตาราง 3.1 ในการพิจารณาร่วม

ตาราง 3.1 การพิจารณา 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)

จากตาราง 3.1 สามารถกำหนดรูปแบบของแบบจำลองได้ดังต่อไปนี้ หากคอเรลโลแกรมของ ACF มีลักษณะลู่โค้งเข้าหาแกนในระนาบ ในขณะที่คอเรลโลแกรม ของ PACF เกิดค่าขึ้นมาเพียงไม่กี่ค่าแล้วก็หายไป จำนวนของแท่งของค่าที่เกิดขึ้นมาให้นับเป็น p ค่าของแบบจำลอง AR(p) เช่น AR(1) จะเกิดขึ้นเมื่อคอเรลโลแกรม ของ ACF ที่โค้งลู่เข้าแกนระนาบ และ PACF จะมีแท่ง คอเรลโลแกรมเกิดขึ้น 1 แท่ง หากคอเรลโลแกรมของ ACF ที่เกิดขึ้นมาไม่กี่ค่าแล้วหายไป ในขณะที่ PACF จะลู่โค้งเข้าหาแกนระนาบนั้น เป็นแบบจำลอง MA(q) เช่น คอเรลโลแกรม ของ ACF เกิดแท่งคอเรลโลแกรมขึ้นเพียง 2 แท่ง และหลังจากนั้นก็หายไป ในขณะที่ PACF โค้งลู่เข้าหาแกนระนาบ แสดงว่าแบบจำลองจะมีลักษณะเป็น MA(2) และหาก ACF และ PACF โค้งเข้าหาแกนระนาบทั้งคู่ แบบจำลองที่ควรจะเป็นคือ ARMA(p, q) เมื่อรวมกันกับการทดสอบความนิ่งในข้อ 1 แล้ว จึงจะสามารถหาค่าผลต่างได้ ซึ่งผลจากการหาผลต่างจำนวน d ครั้งนั้น จะได้แบบจำลองอารีมา (p, d, q)

ในการหาแบบจำลองอารีมาที่เหมาะสม นอกจากจะใช้หลักการดังกล่าวแล้ว ยังต้องมีการพิจารณาค่าสถิติอื่นร่วมด้วย เนื่องจากการพยากรณ์ด้วยวิธีอารีมาไม่มีกฎเกณฑ์ตายตัวว่าแบบจำลองใดจะดีที่สุด ทั้งนี้ขึ้นอยู่กับทักษะความชำนาญของผู้เลือกแบบจำลองซึ่งค่าสถิติต่างๆ ที่สำคัญได้แก่

1.1 ค่า Root Mean Square Error (RMSE) เป็นค่าที่แสดงถึงความคลาดเคลื่อนระหว่างค่าที่ประมาณได้จากแบบจำลองกับข้อมูลจริง ถ้าค่า RMSE มีค่าน้อยยิ่งแสดงว่าแบบจำลองที่ใช้สามารถใช้เป็นตัวแทนของข้อมูลจริงได้ดี (Pindyck and Robinfeld, 1996) มีสมการการหาค่า RMSE ดังนี้

$$RMSE = \sqrt{1/T\sum_{t=1}^{T}(Y_t^s - Y_t^a)}$$

โดยที่ Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ข้อมูลจริง

T คือ คาบเวลาที่ใช้ในแบบจำลอง

1.2 ค่า Theil's Inequality Coefficient มีหลักในการพิจารณาคล้ายกับ RMSE แต่จะมีค่าระหว่าง 0 ถึง 1 ซึ่งถ้า Theil's Inequality Coefficient มีค่าเท่ากับศูนย์ หมายความว่า ค่าที่ประมาณจากแบบจำลองสามารถประมาณได้เท่ากับข้อมูลที่มีอยู่จริง หมายความว่าแบบจำลองนั้นสามารถใช้เป็นตัวแทนของข้อมูลได้เป็นอย่างดี มีสมการการหาดังนี้

$$\textbf{Theil's Inequality Coefficient} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)} \Bigg/ \left\{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^a)^2}\right\}$$

โดยที่ Y_t^s คือ ค่าที่ประมาณจากแบบจำลอง

Y_t^a คือ ข้อมูลจริง

T คือ คาบเวลาที่ใช้ในแบบจำลอง

1.3 ค่า Adjusted R – Square หมายถึงค่า R – Square ที่ปรับปรุงแล้วโดยการขจัดอิทธิพลของตัวแปรอิสระต่างๆ ที่มากเกินไปจนทำให้ ค่า R – Square มีค่ามากผิดปกติ ซึ่งอาจจะทำให้การตีความหมายผิดพลาดเนื่องจาก R – Square แสดงความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม ถ้า R – Square มีค่ามากแสดงว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้ดี ดังนั้นเพื่อหลีกเลี่ยงปัญหาดังกล่าว จึงพิจารณาค่า Adjusted R – Square แทน

$$\textbf{Adjusted R – Square} = 1 - \frac{\sum u_i^2 \ /(n-k)}{\sum y_i^2 \ /(n-1)}$$

1.4 ค่า Akaike Information Criterion (AIC) เป็นค่าสถิติที่อยู่ในรูปของลอการิทึมธรรมชาติ มีสมการการหาดังนี้

$$\textbf{ln AIC} = \left[\frac{2k}{n}\right] + \ln\left[\frac{\sum u_i^2}{n}\right]$$

โดยที่ n คือ จำนวนค่าสังเกตทั้งหมด

$\sum u_i^2$ คือ ผลรวมค่าความคลาดเคลื่อน

หลักในการพิจาณาค่าสถิติ AIC คือ ถ้าค่า AIC มีค่าน้อยเท่าใด หมายความว่าแบบจำลองที่ใช้สามารถเป็นตัวแทนของข้อมูลได้ดีมากเท่านั้น

ค่าสถิติที่ได้กล่าวถึงข้างต้นจะนำมาใช้ในการพิจารณาประกอบกันเพื่อเลือกแบบจำลองที่ดีไว้ประมาณ 3 – 4 แบบจำลอง และจะทำการเลือกอีกครั้งหนึ่งในขั้นการพยากรณ์ (Forecasting) ซึ่งเมื่อถึงขั้นตอนนั้นจะเลือกแบบจำลองที่ดีที่สุดเพียงแบบจำลองเดียว

2. การประมาณค่าพารามิเตอร์ (Parameter estimation) คือการพิจารณาค่าสัมประสิทธิ์ของ AR (autoregressive) และ MA (Moving Average) เพื่อนำมาสร้างสมการความสัมพันธ์ที่จะนำมาใช้ในการพยากรณ์ต่อไป

3. การตรวจสอบความถูกต้อง (Diagnostics Checking) คือการตรวจสอบรูปแบบที่นำมาใช้ว่ามีความถูกต้องเหมาะสมหรือไม่ ในที่นี้ใช้ Box – Pierce Q statistic ในการพิจารณา

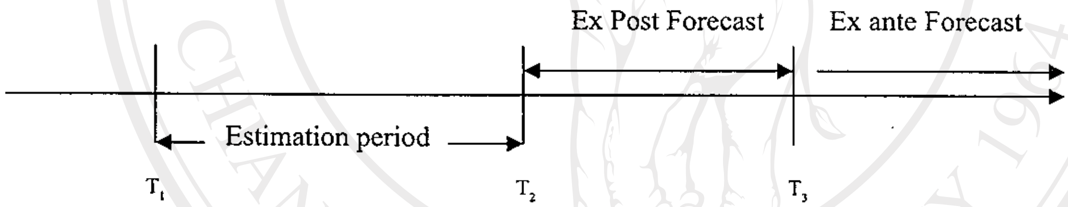
$$\text{Q statistic} = n\sum_{k=1}^{m} \rho_k^2$$

โดยที่ n คือ จำนวนข้อมูล

m คือ จำนวน Lag

หลักการใช้ค่า Q statistic ในการตรวจสอบความถูกต้องคือ ค่า Q จะมีการแจกแจงแบบไคสแควร์ โดยมีสมมุติฐานว่าง คือ ความคลาดเคลื่อนที่ประมาณได้นั้นไม่เกิดอัตสหสัมพันธ์ (Autocorrelation) ที่ระดับความเชื่อมั่นที่กำหนดหรือมีลักษณะเป็นWhite Noise ถ้าตรวจสอบแล้วพบว่าไม่เกิดอัตสหสัมพันธ์ ถือว่าแบบจำลองนั้นมีความเหมาะสมที่จะนำไปสู่ขั้นตอนการพยากรณ์ต่อไป แต่ถ้าหากพบว่าเกิดอัตสหสัมพันธ์ ถือว่าแบบจำลองนั้นไม่เหมาะสมที่จะใช้ในการพยากรณ์จะต้องกำหนดแบบจำลองใหม่

4. การพยากรณ์ (Forecasting) การพยากรณ์โดยวิธีอารีมาควรจะเป็นการพยากรณ์ในช่วงเวลาสั้นๆ ซึ่งแบ่งช่วงการพยากรณ์ออกเป็น 3 ช่วงดังนี้



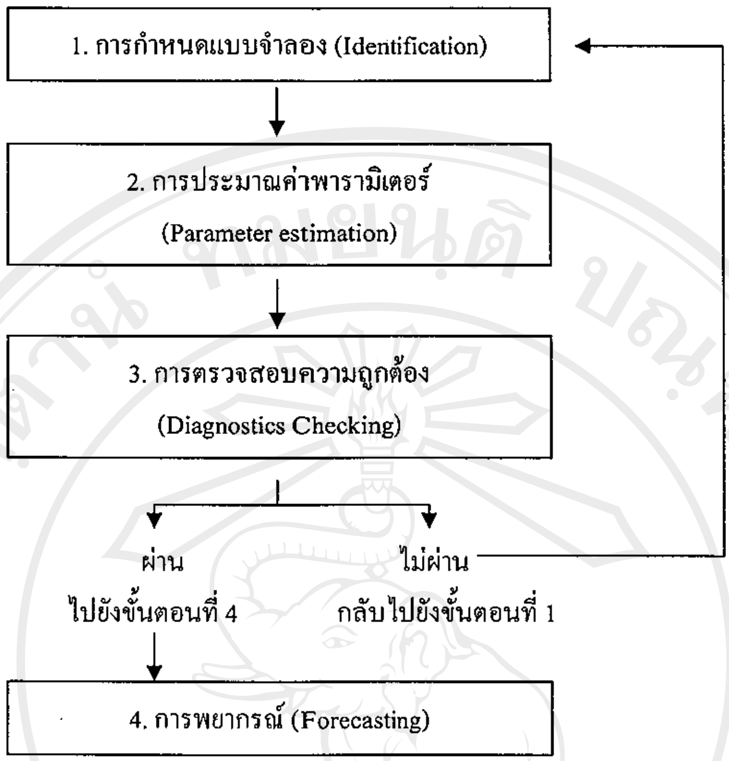
แผนภูมิ 3.1 แสดงขั้นตอนในการพยากรณ์

4.1 Estimation period หมายถึง การใช้แบบจำลองเพื่อคาดคะเนในช่วงข้อมูลตั้งแต่ T_1 ถึง T_2 จากข้อมูลทั้งหมดที่มีถึง T_2

4.2 Ex Post Forecast หมายถึง การพยากรณ์โดยใช้ข้อมูล T_2 ถึง T_3 เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ของแต่ละแบบจำลอง โดยการนำค่าพยากรณ์ที่ได้เปรียบเทียบกับกับข้อมูลที่มีอยู่จริง และดูค่า Root Mean Squared Error และ Theil's Inequality Coefficient ประกอบ

4.3 Ex ante Forecast หมายถึง การพยากรณ์ในช่วงเวลาต่อไปถัดจากข้อมูล T_3 เป็นต้นไป และดูค่า Root Mean Squared Error และ Theil's Inequality Coefficient ประกอบอีกครั้งเพื่อเลือกแบบจำลองที่เหมาะสม

ขั้นตอนต่างๆ ทั้งหมดสามารถสรุปเป็นแผนภูมิการวิจัยตามวิธี Box-Jenkins ได้ดังนี้



แผนภูมิ 3.2 แสดงขั้นตอนการวิจัยโดยวิธี Box-Jenkins