การประยุกต์ใช้ ARIMA Model เพื่อการวิจัย

Application of ARIMA Model for Research

จินดามาส สุทธิชัยเมธี*
Jindamas Sutthichaimethee

*นักวิเคราะห์แผนและนโยบายกระทรวงวิทยาศาสตร์และเทคโนโลยี

*Lecturer Plan and Policy Analyst, Ministry of Science and Technology

บทคัดย่อ

บทความนี้ มีวัตถุประสงค์เพื่อสร้าง The Best Model โดยประยุกต์ ARIMA Model กับตัวแปรโครงสร้าง (Structure Variable) เรียกว่า ARIMAX Model ตามขั้นตอนของ Statistics Model และสามารถนำแบบจำลองดังกล่าวไปใช้สำหรับการพยากรณ์ให้เกิดประสิทธิภาพสูงสุดด้วย

จากการศึกษาข้อมูลในระบบเศรษฐกิจมหาภาคพบว่า ส่วนมากข้อมูลเหล่านี้จะมีลักษณะเป็น Non Stationary และเพื่อให้การพยากรณ์ให้เกิดประสิทธิภาพสูงสุด ในการศึกษาจึงได้ปรับข้อมูลให้มีลักษณะเป็น Stationary และถ้าข้อมูลมี Co-integration กัน จำเป็นอย่างยิ่งที่จะต้องนำค่า Error Correction Mechanism (ECM) ประกอบไว้ในแบบจำลองนั้นๆ และจากผลการศึกษาพบว่า การสร้างตัวแบบที่ดีที่สุดนั้น (The Best Model) จะต้องมีการเลือกแบบจำลองประมาณค่าให้ถูกต้องและเหมาะสมกับประเภทของข้อมูลนั้นๆ ซึ่งส่งผลให้การพยากรณ์เกิดความคลาดเคลื่อนต่ำและสามารถนำมาใช้ประโยชน์ได้อย่างถูกต้องต่อไป

คำสำคัญ : ตัวแปรโครงสร้าง/ข้อมูลอนุกรมเวลา/แบบจำลองที่ดีที่สุด

Abstract

This article is intended to create The Best Model ARIMA Model, which applies to the variable structure called ARIMAX Model. Statistics Model of steps and can take the model used for forecasting the maximum efficiency.

For information on the system economy, most will look a Non Stationary, so researchers need to be updated to look as Stationary and if the data is Co-integration parties is essential to introduce the Error Correction Mechanism assembly in that model and The Best Model to create a model estimating the correct and appropriate for that type of information will result in the forecast errors are low and can be used to accurately follow.

Keyword : Structure Variable / Time Series Data / The Best Model

1. บทนำ

การสร้างแบบจำลอง (Model) นั้นมีหลายวิธีด้วยกัน ซึ่งแต่ละวิธีนั้นๆ ย่อมมีความแตกต่างกัน สิ่งที่สำคัญซึ่งผู้สร้างหรือผู้วิจัยจะต้องคำนึงถึงคือ การเลือกใช้ให้ถูกต้องและเหมาะสมที่สุด และวิธีการสร้างแบบจำลองที่ได้รับความนิยมสูงวิธีหนึ่งคือ วิธี Box-Jenkins

การสร้างแบบจำลอง (Model) ด้วยวิธี Box–Jenkins โดย George E.P. Box และ Gwilym M.Jenkins ในปี ค.ศ. 1970 ได้นำเสนอรูปแบบ ARIMA Model และได้ทำการปรับปรุงในปี ค.ศ. 1994 ต่อมาได้มีการนำมาใช้กันอย่างมากในปัจจุบันนี้ การกำหนดแบบจำลอง และการพยากรณ์ด้วยวิธีของ Box–Jenkins เป็นวิธีวิเคราะห์อนุกรมเวลา (Time Series Data) โดยอาศัย Stochastic Process ที่มีลักษณะอนุกรมเวลาเป็น **Stationary Time Series** และ **Non–stationary Time Series** กล่าวคือ เป็นรูปแบบที่ใช้อธิบายการเคลื่อนไหวของข้อมูลที่อาศัยลักษณะที่มีสหสัมพันธ์และลักษณะนิ่ง (Stationary) หรือจะต้องปรับข้อมูลที่นำมาวิเคราะห์และทำการพยากรณ์ให้มีความเป็น Stationary ก่อนเสมอ เพื่อให้การกำหนดแบบจำลองได้อย่างถูกต้องมากที่สุด และสามารถนำไปใช้สำหรับ การพยากรณ์ให้เกิดความผิดพลาดจากค่าจริง (Actual Value) น้อยที่สุดนั่นเอง

2. แนวคิดการสร้างแบบจำลอง ARIMA Model

การตรวจสอบคุณสมบัติของข้อมูลที่นำมาสร้างแบบจำลองด้วยวิธี Box–Jenkins ซึ่งข้อมูลที่นำมาสร้างแบบจำลอง ARIMA Model นั้น ต้องมีการตรวจสอบคุณสมบัติ ได้แก่ (1) Stationary (2) Co–integration และ (3) Error Correction Mechanism (ECM) ซึ่งมีรายละเอียดดังต่อไปนี้

1.Stationary ถ้ากำหนดให้ Y_t เป็น Stochastic Variable Time Series และมีลักษณะเป็น Stationary จะต้องประกอบด้วยคุณสมบัติ 3 ประการ ดังนี้

Mean :

$$E(Y_t) = E(Y_{t-k}) = \mu$$

Variance :

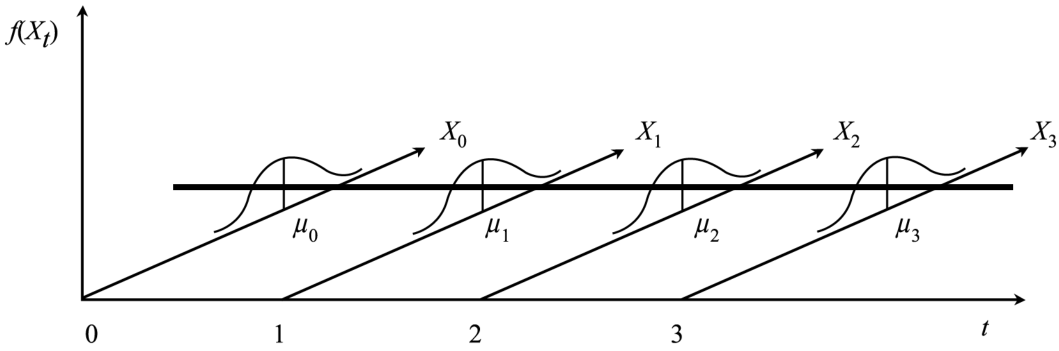
$$\mathrm{Var}(Y_t) = E(Y_t - \mu)^2 = E(Y_{t-k} - \mu)^2 = \sigma^2$$

Covariance :

$$\mathrm{E}\,[(Y_t - \mu)^2 (Y_{t-k} - \mu)^2] = \gamma_k$$

ความสูงหรือระยะห่างของค่า Covariance มีลักษณะเท่ากัน ซึ่งระยะห่างระหว่างค่า Y_t สองค่าไม่ได้ขึ้นอยู่กับเวลา (Time) ด้วยคุณสมบัติทั้ง 3 ข้อดังกล่าวจะเรียกข้อมูลนี้ว่าเป็น Stationary

Stationary Stochastic Process หรือ Stationary คือ ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ยหรือค่าความคาดหวัง (Mean or Expected Value) และค่าความแปรปรวน (Variance) และค่าความแปรปรวนร่วม (Covariance) คงที่ (Constant Over Time) ซึ่งไม่ได้ขึ้นอยู่กับเวลา แต่จะขึ้นอยู่กับระยะหรือช่วงห่างของช่วงเวลา (Distance or Lag) แสดงดังรูปต่อไปนี้



รูปที่ 1 แสดงคุณสมบัติความเป็น Stationary

ถ้าข้อมูลอนุกรมเวลาขาดคุณสมบัติข้อใดข้อหนึ่ง จะเรียกว่าเป็น Non–stationary ข้อมูลที่มีลักษณะเป็น Non–stationary มีลักษณะดังต่อไปนี้

Mean :

$E(Y_t) = E(Y_{t-k}) = t\mu$

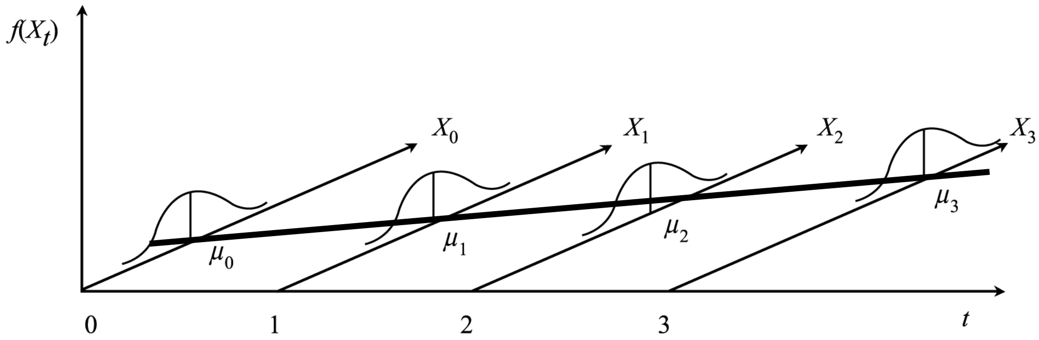
Variance :

$\mathrm{Var}(Y_t) = E(Y_t - \mu)^2 = E(Y_{t-k} - \mu)^2 = t\sigma^2$

Covariance :

$\mathrm{E}\,[(Y_t - \mu)^2\,(Y_{t-k} - \mu)^2] = t\gamma_k$

Non–stationary Stochastic Process หรือ Non–stationary นั้น จะมีค่าความคาดหวัง ค่าความแปรปรวน และค่าความแปรปรวนร่วมไม่คงที่ ซึ่งจะขึ้นอยู่กับเวลาหรือข้อมูลมีลักษณะที่เรียกว่า Random Walk แสดงได้ดังรูปต่อไปนี้



รูปที่ 2 แสดงคุณสมบัติความเป็น Non–stationary

การทดสอบความเป็น Stationary ของข้อมูลที่ใช้ในการศึกษามีด้วยกันหลายวิธี เช่น วิธี Dickey Fuller (DF), Augmented Dickey and Fuller (ADF) เป็นต้น ข้อมูลอนุกรมเวลาส่วนใหญ่จะมีคุณสมบัติเป็น Non–stationary หรือกล่าวได้ว่า ข้อมูลอนุกรมเวลาเหล่านั้นมีส่วนประกอบที่เรียกว่า Unit Root ถ้าหากนำข้อมูลที่มีคุณสมบัติ Non–stationary หรือมี Unit Root มาประมาณการโดยใช้ Regression Model ด้วยวิธี Ordinary Least Square (OLS) ค่าพารามิเตอร์ที่ได้จากการประมาณการแม้ว่ามีนัยสำคัญทางสถิติ (Significance) แต่จะขาดความน่าเชื่อถือ เรียกว่าเกิดปัญหา **Spurious Regression** (จินดามาส สุทธิชัยเมธี, 2549)

การปรับข้อมูลที่มีลักษณะ Non–stationary ให้เป็น Stationary หรือ Weak Stationary นั้น จะปรับข้อมูลเฉพาะ First Moment และ Second Moment แต่ถ้าในกรณีที่ข้อมูลเป็น Strictly Stationary ก็จะไม่พิจารณาแต่เพียง Moment ลำดับที่หนึ่งและสองเท่านั้นต้องพิจารณา Moment ในลำดับที่สูงขึ้นด้วย

การพิจารณาเรื่อง Moment นั้นประกอบด้วยหลักสำคัญ 2 ประการ ได้แก่ ประการแรก ตัวแปรที่มีลักษณะเป็น Stationary จะผันแปรในช่วงแคบ ๆ บริเวณค่าเฉลี่ยของตัวแปรนั้น แต่ถ้าข้อมูลที่มีลักษณะ Non–stationary จะมีการผันแปรมากกว่าเมื่อ Observation สูงขึ้น และประการที่สอง ถ้าในระบบเศรษฐกิจมีปัจจัยภายนอกมากระทบ (Shock) สำหรับข้อมูลที่เป็น Stationary จะเบี่ยงเบนออกจากค่าเฉลี่ยเพียงชั่วคราว และจะกลับเข้าสู่ดุลยภาพได้อย่างรวดเร็ว แต่ถ้าหากข้อมูลที่นำมาวิเคราะห์มีลักษณะเป็น Non–stationary จะให้ผลในทางตรงข้ามกัน โดยจะส่งผลกระทบต่อ Model หรือส่งผลกระทบต่อตัวแปรอื่นๆ ในแบบจำลอง และในเวลาช่วงอื่นๆ ด้วย แสดงให้เห็นว่าตัวแปรที่มีลักษณะเป็น Non–stationary จะไม่มีค่าเฉลี่ยระยะยาว (Long Run Mean Level) (ถวิล นิลใบ, 2544)

การทดสอบความเป็น Stationary ของข้อมูลที่ใช้ในการศึกษามีด้วยกันหลายวิธี เช่น วิธี Dickey Fuller (DF) และวิธี Augmented Dickey Fuller (ADF) เป็นต้น

การทดสอบคุณสมบัติของตัวแปรตามแนวคิดของ Dickey Fuller Test และ Augmented Dickey Fuller Test โดยใช้ Unit Root Test แนวคิดของ Dickey–Fuller ได้นำมาใช้เพื่อแก้ปัญหาข้อมูลที่เป็น Non–stationary เรียกว่าสมการถดถอยในรูปของผลต่าง (Difference Regression) โดยมีลักษณะตัวแบบสมการที่เรียกว่า First Order Autoregressive Process ตามรูปแบบใดรูปแบบหนึ่งใน 3 สมการต่อไปนี้

ก. $\Delta Y_t = \delta Y_{t-1} + \varepsilon_t$ (Random Walk Process หรือ Pure Random Walk)

ข. $\Delta Y_t = \alpha_1 + \delta Y_{t-1} + \varepsilon_t$ (Random Walk with Drift หรือเป็นการเพิ่มค่า Intercept)

ค. $\Delta Y_t = \alpha_1 + \alpha_2 T + \delta Y_{t-1} + \varepsilon_t$ (Random Walk with Drift และมี Linear Time Trend ใส่ค่า Drift Term และค่า T ซึ่ง T เป็นตัวแปรแนวโน้ม)

กำหนดให้

ΔY_t = ตัวแปรที่ต้องการศึกษา

δ = สัมประสิทธิ์ของความล่าช้า (Coefficient of Lagged)

ε_t = Error Term โดยที่ ε_t มีการกระจายแบบปกติ, Mean = 0, Variance = σ^2

สมมติฐาน (Hypothesis) ของ Unit Root Test

$H_0 : \delta = 0$, Non–stationary

$H_\alpha : \delta < 1$, Stationary

ΔY_t จะมีคุณสมบัติเป็น Non–stationary เมื่อ Accept H_0 แสดงให้เห็นว่า δ = 0 และความแปรปรวนจะเพิ่มขึ้นแบบ Exponential หรือ Explosive เมื่อเวลาเพิ่มขึ้น โดยรายละเอียดของแต่ละรูปแบบ แสดงได้ดังนี้

ก. $\Delta Y_t = \delta Y_{t-1} + \varepsilon_t$ (1)

จากสมการที่ (1) มีลักษณะเป็นแบบจำลองอย่างง่ายที่มีค่าเฉลี่ย (Mean) เท่ากับศูนย์ เหมาะสมสำหรับข้อมูลที่ค่าเฉลี่ยรวมทั้งหมดของข้อมูลเป็นศูนย์เท่านั้น จึงได้ปรับปรุงแบบจำลองโดยการใส่ค่าคงที่ (Drift Term) ไว้ในสมการดังนี้

ข. $\Delta Y_t = \alpha_1 + \delta Y_{t-1} + \varepsilon_t$ (2)

กำหนดให้

α_1 = ค่าคงที่ (Drift Term)

จากสมการที่ (2) ยังสามารถปรับปรุงการทดสอบ Unit Root Test เมื่อมีส่วนประกอบของกระบวนการ Trend Stationary (TS) และ Difference Stationary (DS) ได้ดังสมการต่อไปนี้

ค. $\Delta Y_t = \alpha_1 + \alpha_2 T + \delta Y_{t-1} + \varepsilon_t$

กำหนดให้

T = ความโน้มเอียงของเวลา (Time Trend)

α_2 = ค่าสัมประสิทธิ์ความโน้มเอียง

ε_t มีลักษณะ Stationary ค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเท่ากับ σ^2 หรือ $\varepsilon_t \sim IID, (0, \sigma^2)$

อย่างไรก็ตาม การนำข้อมูลอนุกรมเวลา (Time Series) ไปทำการ First Difference หรือการหาค่าความแตกต่างอันดับแรกหรือค่าความแตกต่างอันดับสูงกว่า ก็จะทำให้ข้อมูลอนุกรมนั้นมีคุณสมบัติความเป็น Stationary โดยแสดงวิธีการ Difference Stationary ได้ดังนี้

$Y_t = Y_{t-1}$

สามารถนำมาสร้างสมการได้ดังนี้

$$\Delta Y_t = \alpha_1 + \alpha_2 T + \delta Y_{t-1} + \varepsilon_t \quad (4)$$

จากสมการที่ (4) พบว่า ในระดับเวลาที่เพิ่มขึ้น ณ ระดับ Level ถ้า H_α หรือ Accept H_0 แสดงว่าข้อมูลนั้นมีลักษณะเป็น Non–stationary คือค่าสัมประสิทธิ์ $\delta = 0$ อย่างมีนัยสำคัญ แสดงให้เห็นว่า Tau Statistics ของสัมประสิทธิ์ที่คำนวณในรูป Absolute Term น้อยกว่าค่า DF Critical ในรูป Absolute Term ในกรณีดังกล่าวนี้แสดงให้เห็นว่า ε_t ขาดคุณสมบัติความเป็น White Noise หรือแสดงให้เห็นว่าข้อมูลที่นำมาวิเคราะห์มีลักษณะเป็น Autocorrelation จะต้องเลือกใช้การทดสอบแบบ Augmented Dickey Fuller (ADF) โดยมีสมมติฐาน และ Goodness of Fit ที่ใช้ในการทดสอบเหมือนการทดสอบแบบ Dickey Fuller (DF) ซึ่งได้มีการปรับรูปแบบสมการโดยการเพิ่มจำนวนตัวแปรล่าช้า (Lagged) ของตัวแปรตาม (Dependent Variable) ในลำดับที่สูงขึ้น เพื่อขจัดปัญหา Autocorrelation แสดงรูปแบบดังนี้

สมมติฐาน (Hypothesis) ของ Unit Root Test

$H_0 : \delta$ = 0, Non–stationary

$H_\alpha : \delta < 1$, Stationary

ณ ระดับ Level ถ้า Reject H_0 หรือ Accept H_α แสดงว่า ข้อมูลนั้นมีลักษณะเป็น Stationary กล่าวคือ ค่าสัมประสิทธิ์ $\delta \neq 0$ อย่างมีนัยสำคัญ แสดงให้เห็นว่า Tau Statistics ของสัมประสิทธิ์ที่คำนวณในรูป Absolute Term มากกว่าค่า ADF Critical ในรูป Absolute Term ในกรณีดังกล่าวนี้พบว่า ε_t มีคุณสมบัติความเป็น White Noise หรือข้อมูลอนุกรมเวลาของตัวแปรมีลักษณะ Stationary ซึ่งก็แสดงให้เห็นว่า ΔY_t Integrated ลำดับที่ d หรือเขียนในรูปสัญลักษณ์

ได้ว่า $\Delta Y_t \sim I(d)$

$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i+1} + \varepsilon_t \quad (5)$$

$$\Delta Y_t = \alpha_1 + \delta Y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i+1} + \varepsilon_t \quad (6)$$

$$\Delta Y_t = \alpha_1 + \alpha_2 T + \delta Y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i+1} + \varepsilon_t \quad (7)$$

กำหนดให้

p = ตัวแปรล่าช้าของผลต่างของตัวแปร (Lagged Values of First Difference of the Variable)

จากสมการที่ (5), สมการที่ (6) และสมการที่ (7) พบว่า รูปแบบสมการที่เหมาะสมและนำมาใช้กับข้อมูลจริงมากที่สุดคือ Augmented Dickey Fuller ในสมการที่ (6) นั่นเอง

$$\Delta Y_t = \alpha_1 + \alpha_2 T + \delta Y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i+1} + \varepsilon_t$$

อย่างไรก็ตาม พบว่ารูปแบบสมการของ DF และ ADF จะไม่แตกต่างกันมาก ซึ่งวิธีการของ ADF ก็เพื่อแก้ปัญหาค่า Error Term ให้มีคุณสมบัติความเป็น White Noise หรือเพื่อต้องการให้ Error Term มีค่า Mean เท่ากับศูนย์นั่นเอง

2. Co–integration

สำหรับแนวคิดของ Eagle and Granger มีข้อสรุปทางทฤษฎีในเรื่อง Co–integration ว่าเป็นข้อมูลอนุกรมเวลา (Time Series) ตั้งแต่ 2 ชุดที่มีความสัมพันธ์กันในเชิงเคลื่อนไหวไปพร้อม ๆ กัน ซึ่งอยู่ในสภาพที่แน่นอน (Steady State) แม้ว่าข้อมูลที่นำมาวิเคราะห์จะไม่ใช่ข้อมูลที่เป็น Stationary ก็ตาม

Engle – Granger มีหลักการตรวจสอบคุณสมบัติ Co–integration ได้โดยพิจารณาคุณสมบัติของค่าคลาดเคลื่อน (Error) ซึ่งสามารถคำนวณได้จากสมการถดถอย Co–integrating Regression แล้วตรวจสอบสมมติฐาน (Hypothesis) ว่าเป็นอย่างไร ถ้า Reject H_0 แสดงว่าข้อมูลที่นำมาวิเคราะห์มีลักษณะเป็น Stationary กล่าวคือ ข้อมูลที่นำมาวิเคราะห์เป็นส่วนประกอบเชิงเส้นตรง (Linear Combination)

หรือแสดงว่าตัวแปรที่นำมาพิจารณามีคุณสมบัติเป็น Co–integration ซึ่งการตรวจสอบก็จะใช้แนวคิดของ Dickey–Fuller (DF) หรือ Augmented Dickey–Fuller (ADF) ก็ได้

ขั้นตอนการทดสอบ Co–integration

ขั้นตอนที่ 1 นำข้อมูลที่ใช้วิเคราะห์มาตรวจสอบลำดับ Integrated ของตัวแปรตาม (Dependent Variable : Y_t) และตัวแปรอิสระ (Independent Variable : X_t) โดยวิธีการตรวจสอบ Unit Root Test ถ้าพบว่าลำดับ Integrated ของตัวแปรแต่ละตัวไม่เท่ากัน แสดงว่าข้อมูลนั้นไม่มีลักษณะเป็น Co–integration แน่นอน โดยไม่ต้องทดสอบต่อไปอีก โดยอาจจะต้องเปลี่ยนข้อมูลใหม่ หรือเปลี่ยนแปลงช่วงของเวลา แต่ในทางตรงกันข้าม ถ้าลำดับ Integrated ของค่าตัวแปรแต่ละตัวอยู่ในลำดับเดียวกัน ก็จะดำเนินการทดสอบต่อไป

ขั้นตอนที่ 2 หาค่า Co–integrating Parameter จากการคำนวณหาค่าตัวคลาดเคลื่อน (Error Term) โดยใช้วิธีการคำนวณ Ordinary Least Squares (OLS)

$$u_t = Y_t - \alpha - \beta X_t \quad (8)$$

ขั้นตอนที่ 3 ตรวจสอบว่าค่า u_t ที่ได้จากการคำนวณมีคุณสมบัติความเป็น Stationary หรือไม่ หรือเป็นการตรวจสอบว่า u_t มีส่วนประกอบเชิงเส้นตรง (Linear Combination) หรือเป็นตัว White Noise หรือไม่นั่นเอง โดยจะเลือกใช้วิธีการของ Augment Dickey

Fuller (ADF) เพื่อป้องกันไม่ให้ตัวรบกวนเกิด Autocorrelation

ถ้าผลการทดสอบพบว่า Reject H_0 หรือ Accept H_α โดยเปรียบเทียบ Tau Test ในรูปค่าสัมบูรณ์ (Absolute) ของค่าสัมประสิทธิ์เทียบกับค่า Tau Critical ของ MacKinnon ในรูปค่าสัมบูรณ์ แสดงว่า u_t มีคุณสมบัติที่เรียกว่า Stationary คือไม่มี Unit Root กล่าวคือ Y_t และ X_t มีความสัมพันธ์เชิงดุลยภาพระยะยาว (Co–integration) แต่ในทางตรงกันข้าม Reject H_αหรือ Accept H_0 แสดงว่า u_t มีคุณสมบัติที่เรียก Non–stationary คือมี Unit Root กล่าวคือ Y_tและ X_t ไม่มีความสัมพันธ์เชิงดุลยภาพระยะยาว (Non–Co–integration)

3. Error – Correction Mechanisms (ECM)

เมื่อตรวจสอบว่าข้อมูลที่นำมาวิเคราะห์มีคุณสมบัติ Co–integration แล้ว ขั้นตอนต่อไปจำเป็นจะต้องสร้างแบบจำลองการปรับตัวในระยะสั้น หรือแบบจำลองกระบวนการการปรับตัวในระยะสั้น เพื่อให้ข้อมูลสามารถเข้าสู่ดุลยภาพในระยะยาวได้ หรือเพื่ออธิบายการปรับตัวในระยะสั้นแบบพลวัต (Short Run Dynamic Adjustment) ของตัวแปรในแบบจำลอง (Model) โดยเฉพาะนิยมนำมาใช้ในแบบจำลองเศรษฐกิจมหภาค (Macro Model) และพบบ่อยในงานวิจัยเกี่ยวกับเศรษฐกิจ ซึ่งการปรับตัวในระยะสั้นสามารถคำนวณได้จากค่าสัมประสิทธิ์ของตัวแปรที่อยู่ในสมการของ ECM ข้อดีของแบบจำลอง ECM คือ จะไม่ทำให้เกิดปัญหาความสัมพันธ์ที่แท้จริง และยังไม่เกิดผลเสียต่อลักษณะของข้อมูลที่นำมาวิเคราะห์ด้วย

หลักการสร้าง ECM Model เริ่มจากการนำข้อมูลมาวิเคราะห์ Co integration แล้ว ถ้าพบว่าเกิด Stationary แสดงได้ว่าข้อมูลนั้นมีลักษณะเป็น Co–integration ก็สามารถสร้างแบบจำลอง ECM เพื่ออธิบายกระบวนการปรับตัวในระยะสั้นของตัวแปรต่างๆ ให้เข้าสู่ดุลยภาพในระยะยาวได้ ซึ่งสามารถแสดงได้ดังนี้

$$\Delta Y_t = \alpha + \sum_{t=1}^{p}\left(\beta_{1i}\Delta X_{t-i} + \beta_{2i}\Delta Y_{t-i}\right) + \gamma u_{t-i} + \varepsilon_t \quad (9)$$

จากสมการที่ (9) พบว่า แบบจำลองการปรับตัวในระยะสั้น (ECM Model) ได้สร้างขึ้นโดยคำนึงถึงผลกระทบที่เกิดจากความคลาดเคลื่อน (Error Team : u_{t-i}) ของการปรับตัวของตัวแปรต่างๆ ในระยะยาว จุดเด่นของ Model คือ จะแสดงขนาดของความไม่สมดุลระหว่างค่า Y_t และ X_t ที่เกิดขึ้นในช่วงเวลาก่อนหน้านี้ ซึ่ง ECM Model ที่ประมาณค่าได้นั้น ถ้าหากค่าสัมประสิทธิ์ (γ) มีค่ามากเท่าไร แสดงว่าตัวแปร Y_t มีการปรับตัวเข้าสู่ดุลยภาพในระยะยาวได้อย่างรวดเร็วเท่านั้น แต่ถ้าค่าสัมประสิทธิ์ (γ) มีค่าน้อยเท่าไร ก็จะแสดงว่าตัวแปร Y_t จะมีการปรับตัวเข้าสู่ดุลยภาพในระยะยาวได้ช้าเท่านั้นด้วย

3. ขั้นตอนในการวิเคราะห์ ARIMA Model

ขั้นตอนการวิเคราะห์ ARIMA Model ด้วยวิธี Box–Jenkins มีขั้นตอนในการวิเคราะห์ 4 ขั้นตอน ได้แก่ (1) การกำหนดรูปแบบ (Identification) (2) การประมาณค่าพารามิเตอร์ (Parameter Estimator) (3) การตรวจสอบรูปแบบ (Diagnostic Checking) และ (4) การพยากรณ์ (Forecast) โดยมีรายละเอียดดังต่อไปนี้

1. การกำหนดรูปแบบ (Identification) รูปแบบของอนุกรมเวลาที่ใช้ในการพยากรณ์โดยวิธี Box – Jenkins เมื่ออนุกรมเวลามีคุณสมบัติเป็น Stationary หรือ invertible มีดังนี้

ก. รูปแบบ Autoregressive Model of Order p หรือ AR(p) มีรูปแบบดังนี้

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_1 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \varepsilon_t \quad (10)$$

จากสมการที่ (10) ประกอบด้วยรูปแบบดังต่อไปนี้

● AR (1) มีรูปแบบดังนี้

$$Y_t = \alpha + \beta_1 Y_{t-1} + \varepsilon_t \quad (11)$$

โดยกำหนดให้

$|\beta_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Stationary

● AR (2) มีรูปแบบดังนี้

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t \quad (12)$$

โดยกำหนดให้

β_1 และ β_2 และ $\beta_2 - \beta_1 < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Stationary

รูปแบบค่าเฉลี่ยเคลื่อนที่อันดับ q (Moving Average Model of Order q) หรือ MA(q) แสดงได้ดังนี้

$$Y_t = \delta + \varepsilon_t - \gamma_1 \varepsilon_t - 1 - \gamma_2 \varepsilon_t - 2 - \ldots - \gamma_2 \varepsilon_{t-1} \quad (13)$$

จากสมการที่ (13) ประกอบด้วยรูปแบบดังต่อไปนี้

● MA(1) มีรูปแบบดังนี้

$$Y_t = \delta + \varepsilon_t - \gamma_1 \varepsilon_{t-1} \quad (14)$$

โดยกำหนดให้

$|\gamma_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Invertible or Stationary

● MA(2) มีรูปแบบดังนี้

$$Y_t = \delta + \varepsilon_t - \gamma_1 \varepsilon_{t-1} - \gamma_2 \varepsilon_{t-2} \quad (15)$$

โดยกำหนดให้

สำหรับ $\gamma_1 + \gamma_2 < 1$, $\gamma_2 + \gamma_1 < 1$ และ $|\gamma_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Invertible or Stationary

ค. รูปแบบผสมค่าเฉลี่ยเคลื่อนที่ Autoregressive ที่มีอันดับ p และ q (Mixed Autoregressive and Moving – Average Model of Order p and q) หรือ ARMA (p, q) แสดงได้ดังนี้

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \varepsilon_t - \gamma_1 \varepsilon_{t-1} - \gamma_2 \varepsilon_{t-2} - \ldots - \gamma_q \varepsilon_{t-q}$$

● ARMA(1, 1) มีรูปแบบดังนี้

$$Y_t = \alpha + \beta_1 Y_{t-1} + \varepsilon_t \quad (16)$$

โดยกำหนดให้

สำหรับ $|\beta_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Stationary และ $|\gamma_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Invertible or Stationary

ง. รูปแบบ Integrated ค่าเฉลี่ยเคลื่อนที่ Autoregressive (Autoregressive Integrated Moving Average) หรือ ARIMA(p, d, q) โดย d เป็นอันดับที่ของผลต่าง (Different Term) ประกอบด้วยรูปแบบต่างๆ แสดงได้ดังนี้

● ARIMA(0,1,1) หรือ IMA(1,1) มีรูปแบบดังนี้

$$Y_t - Y_{t-1} = \delta + \varepsilon_t - \gamma_1 \varepsilon_{t-1} \quad (17)$$

โดยกำหนดให้

$|\gamma_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Invertible or Stationary

● ARIMA(1,1,0) หรือ ARI(1,1) มีรูปแบบดังนี้

$$Y_t - Y_{t-1} - \beta_1 (Y_{t-1} + Y_{t-1}) = \alpha + \varepsilon_t \quad (18)$$

โดยกำหนดให้

$|\beta_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Stationary

● ARIMA(1,1,1) มีรูปแบบดังนี้

$$Y_t - Y_{t-1} - \beta_1 (Y_{t-1} + Y_{t-1}) = \alpha + \varepsilon_t - \gamma_1 \varepsilon_{t-1} \quad (19)$$

โดยกำหนดให้

$|\gamma_1| < 1$, $|\beta_1| < 1$ แสดงให้เห็นว่าข้อมูลอนุกรมเวลามีคุณสมบัติเป็น Invertible or Stationary

● ARIMA(0,1,0) มีรูปแบบดังนี้

$$Y_t - Y_{t-1} = \varepsilon_t \quad (20)$$

จ. รูปแบบ Integrated ค่าเฉลี่ยเคลื่อนที่ Autoregressive เมื่อมีอิทธิพลฤดูกาล (Seasonal Autoregressive Integrated Moving Average) หรือ SARIMA(p, d, q)^L โดย d เป็นอันดับที่ของผลต่าง L เป็นช่วงระยะห่างของการเกิดฤดูกาล ถ้ารูปแบบฤดูกาลรายเดือน แสดงได้ดังนี้

$$Y_t - Y_{t-12} = \varepsilon_t - \gamma^* \varepsilon_{t-12}$$

โดยกำหนดให้

$|\gamma_1| < 1$

$Y_t - Y_{t-12}$ = ผลต่างของค่าสังเกตที่อยู่ห่างกัน 12 เดือน

γ^* = ค่าพารามิเตอร์ (Parameter) รูปแบบค่าเฉลี่ยเคลื่อนที่ฤดูกาล (Seasonal Moving Average Model)

2. การประมาณค่าพารามิเตอร์ (Parameter Estimation) การประมาณค่าพารามิเตอร์ (Parameter Estimation) จากรูปแบบที่พิจารณาแล้วว่าเหมาะสมกับอนุกรมเวลาตามข้อที่ 1 เป็นการหาความสัมพันธ์ของฟังก์ชันสหสัมพันธ์กับพารามิเตอร์ ซึ่งเป็นความสัมพันธ์สำหรับอนุกรมเวลาแต่ละรูปแบบ การประมาณค่าพารามิเตอร์โดยใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Square : OLS) และเพื่อความสะดวก สามารถใช้โปรแกรมคอมพิวเตอร์คำนวณจนกว่าจะได้ตัวประมาณที่ได้ค่าคลาดเคลื่อนน้อยที่สุด

3. การตรวจสอบรูปแบบ (Diagnostic Checking) การตรวจสอบว่ารูปแบบที่สร้างขึ้นเหมาะสมกับอนุกรมเวลาหรือไม่ ถ้าไม่เหมาะสม จะพิจารณากำหนดรูปแบบที่เหมาะสมใหม่ ซึ่งมีหลายวิธี โดยวิธีตรวจสอบที่ใช้บ่อยมี 2 วิธี แสดงได้ดังต่อไปนี้

ก. การทดสอบว่าพารามิเตอร์แต่ละตัวในรูปแบบเท่ากับ 0 หรือไม่ โดยการใช้การทดสอบ t–statistic ภายใต้สมมติฐาน $H_0 : \gamma = 0$ และ $H_\alpha : \gamma \neq 0$ เมื่อกำหนดให้ γ เป็นค่าพารามิเตอร์ ตัวทดสอบที่ใช้คือ

$$t = \gamma / S_\gamma \quad (21)$$

โดยกำหนดให้

γ = ค่าประมาณของพารามิเตอร์

S_γ = ค่าความคลาดเคลื่อนมาตรฐานของ γ

ข. การทดสอบของ Box – Pierce Chi – Square Test (Q) ซึ่ง Box – Pierce ได้เสนอวิธีตรวจสอบความเหมาะสมของรูปแบบภายใต้สมมติฐาน $H_0 : p_1(e_t) = ... = p_k(e_t) = 0$ โดยใช้ตัวทดสอบสถิติ Box – Pierce Chi –Square (Q) เพื่อตรวจสอบว่าความคลาดเคลื่อน ณ เวลา t คือ e_t, t = 1, 2,…, n มีความเป็นอิสระต่อกันหรือไม่ โดยเปรียบเทียบผลรวมของค่าสหสัมพันธ์ของ e_t ณ เวลาต่าง ๆ

$$Q = (n - d) \Sigma r_j^2(e_t)$$

n = จำนวนค่าสังเกตในอนุกรมเวลา

d = อันดับผลต่างของอนุกรมเวลาที่ทำให้อนุกรมเวลาเป็น Stationary

$r_j^2(e_t)$ = ฟังก์ชันสหสัมพันธ์ระหว่างค่าความคลาดเคลื่อนที่อยู่ห่างกัน j ช่วงเวลา

จากสมการที่ (22) พบว่า Q มีการแจกแจงแบบ Chi–Square โดยประมาณการที่ระดับองศาแห่งความเป็นอิสระ (Degree of Freedom) เท่ากับ $k - n_p$ ซึ่งจะยอมรับสมมติฐานหลัก เมื่อ Q มากกว่า หรือเท่ากับ $\chi^2_{\alpha/2,(k-n_p)}$ แสดงว่า Q ไม่เป็นอิสระต่อกัน ดังนั้นรูปแบบที่ใช้ยังไม่เหมาะสมจึงต้องกลับไปพิจารณาหารูปแบบที่เหมาะสมใหม่ต่อไป

4. การพยากรณ์ (Forecast) เมื่อได้รูปแบบที่ผ่านการทดสอบแล้วว่าเหมาะสม ก็จะใช้รูปแบบนั้นเพื่อการพยากรณ์ค่าในอนาคต ซึ่งสมการสำหรับการพยากรณ์ค่าในอนาคตนี้จะสร้างขึ้นจากรูปแบบที่เหมาะสมที่สุดเท่านั้น การพยากรณ์สามารถพยากรณ์ทั้งเป็นแบบจุด (Point Forecast) และแบบช่วง (Interval Forecast) ซึ่งจะพยากรณ์ล่วงหน้ากี่ช่วงเวลาก็ได้ แต่ปกติไม่นิยมพยากรณ์ล่วงหน้าหลายช่วงเวลา เพราะจะได้ค่าพยากรณ์ที่แตกต่างจากค่าจริงมาก ดังนั้นเมื่อได้ค่าจริง ณ ช่วงเวลาถัดไปแล้ว ควรนำค่า ณ เวลาดังกล่าวไปปรับสมการพยากรณ์เพื่อหาค่าที่เหมาะสมที่สุด

4. แนวทางการประยุกต์ใช้ ARIMA Model

สำหรับ ARIMA Model เป็น Statistics Model ที่มีความเหมาะสมในการนำไปใช้ในการกำหนดแบบจำลองต่างๆ และเหมาะสำหรับพยากรณ์ด้านต่างๆ ในอนาคตได้ แต่จุดอ่อนก็คือ จะมองเฉพาะข้อมูลของตัวเองในอนาคตเท่านั้น โดยมิได้มองถึงปัจจัยอื่นๆ ที่มีอิทธิพลหรือส่งผลให้เกิดการเปลี่ยนแปลง ดังนั้นรูปแบบสมการที่ใช้ได้ในสถานการณ์จริงจะต้องมีการผสม Structure Variables เข้าไปด้วย เรียกว่า **ARIMAX Model** แต่อย่างไรก็ตามพบว่า ARIMAX Model นั้นความยากในการวิเคราะห์มาก แต่ถ้านำไปใช้ก็จะเกิดประโยชน์ได้มากเช่นกันเพื่อให้เข้าใจ ARIMA Model และ ARIMAX Model มากยิ่งขึ้น แสดงได้ดังตัวอย่างที่ 1

ตัวอย่างที่ 1 จากงานวิจัย ของอาจารย์พฤทธ์สรรค์ สุทธิไชยเมธี เรื่อง **วิเคราะห์อุปสงค์การนำเข้ายางพาราของประเทศจีนจากประเทศไทย ได้แก่ ยางแผ่นรมควัน ยางแท่ง และน้ำยางข้น** ในช่วงปี 2538-2547 ด้วย ARIMA Model และพยากรณ์ล่วงหน้าช่วงไตรมาสที่ 1-4 ในปี 2548 ดังนี้

การวิเคราะห์อุปสงค์การนำเข้ายางพาราของประเทศจีนจากประเทศไทย ได้แก่ ยางแผ่น-รมควัน ยางแท่ง และน้ำยางข้น ในรูปแบบของ Autoregressive Integrated Moving Average – X (ARIMAX) ได้นำเอาวิธีการของ ARIMA และตัวแปรอื่นๆ ทางพฤติกรรมเข้ามาพิจารณาร่วมกับตัวแปรทางสถิติ ซึ่งแนวทางการศึกษาจะพิจารณาตามแนวคิดของ ARIMA แบ่งการวิเคราะห์ออกเป็น 3 ขั้นตอน โดยขั้นตอนแรกจะเป็นการทดสอบความเป็น Stationary ของข้อมูลทุกตัวในแบบจำลอง เพื่อกำหนดอันดับของความสัมพันธ์ (Determine Order of Integration) ของตัวแปร ขั้นตอนที่สองจะเป็นการพิจารณาทดสอบความสัมพันธ์ของตัวแปร Co–integration และขั้นตอนสุดท้ายจะเป็นการทดสอบแบบจำลอง ARIMAX เพื่อนำผลการศึกษาที่ได้มาวิเคราะห์อุปสงค์การนำเข้ายางพาราของประเทศจีนจากประเทศไทย

รูปแบบสมการที่แสดงถึงอุปสงค์การนำเข้ายางพาราของประเทศจีนจากประเทศไทย ได้แก่ ยางแผ่นรมควัน ยางแท่ง และน้ำยางข้น ซึ่งมีรูปแบบดังนี้

แบบจำลองยางแผ่นรมควัน

$$\begin{aligned}\Delta \ln Q_{X_t} = {} & \alpha_1 \Delta \ln Q_{X_{t-1}} + \alpha_2 \Delta \ln Q_{X_{t-5}} + \alpha_3 MA_1 + \alpha_4 MA_2 \\ & + \alpha_5 MA_5 + \alpha_6 \Delta \ln p_{X_{t-1}} + \alpha_7 \Delta \ln q_{T_t} + \alpha_8 \Delta \ln E_{t-4} \\ & + \alpha_9 ECM + \alpha_{10} \Delta \ln r1_t + \varepsilon_t \qquad \text{(ก)}\end{aligned}$$

โดยกำหนดให้

Q_{x_t} = ปริมาณการนำเข้ายางแผ่นรมควันในช่วงเวลา t

$Q_{x_{t-i}}$ = ปริมาณการนำเข้ายางแผ่นรมควันในช่วงเวลาที่ $t - i$

$p_{x_{t-i}}$ = ราคายางแผ่นรมควันในช่วงเวลา $t - i$

$q_{T_{t-i}}$ = ปริมาณการผลิตยางรถยนต์และยางรถบรรทุกของประเทศจีนในช่วงเวลา $t - i$

E_{t-i} = อัตราแลกเปลี่ยนเงินบาทต่อเงินหยวนในช่วงเวลา $t - i$

ECM = Error Correction Mechanism

MA(i) = Moving Average ระยะเวลาที่ระดับ i

rl_{t-i} = รายได้ประชาชาติ (GDP) ของจีนในช่วงเวลาที่ $t - i$

ε_t = ค่าความคลาดเคลื่อนในช่วงเวลาที่ t

Δ = ผลต่างลำดับที่หนึ่ง (First Difference)

แบบจำลองยางแท่ง

$$\Delta \ln Q_{Y_t} = \beta_1 \Delta \ln Q_{y_{t-1}} + \beta_2 MA_4 + \beta_3 MA_4 + + \beta_3 MA_4 + \beta_3 \Delta \ln p_{y_{t-1}} + \beta_4 \Delta \ln q_{T_{t-2}} + \beta_5 \Delta \ln E_{t-5} + \beta_6 ECM + + \beta_7 \Delta \ln r1_{t-2} + \varepsilon_t \quad (ข)$$

โดยกำหนดให้

Q_{y_t} = ปริมาณการนำเข้ายางแท่งในช่วงเวลา t

$Q_{y_{t-i}}$ = ปริมาณการนำเข้ายางแท่งในช่วงเวลาที่ $t - i$

$p_{y_{t-i}}$ = ราคายางแท่งในช่วงเวลา $t - i$

$q_{T_{t-i}}$ = ปริมาณการผลิตยางรถยนต์และยางรถบรรทุกของประเทศจีนในช่วงเวลา $t - i$

E_{t-i} = อัตราแลกเปลี่ยนเงินบาทต่อเงินหยวนในช่วงเวลา t

ECM = Error Correction Mechanism

MA(i) = Moving Average ระยะเวลาที่ระดับ i

rl_{t-i} = รายได้ประชาชาติ (GDP) ของจีนในช่วงเวลาที่ $t - i$

ε_t = ค่าความคลาดเคลื่อนในช่วงเวลาที่ t

Δ = ผลต่างลำดับที่หนึ่ง (First Difference)

แบบจำลองน้ำยางข้น

$$\Delta \ln Q_{Z_t} = \mu_1 \Delta \ln Q_{Z_{t-5}} + \mu_2 MA_5 + \mu_3 \Delta \ln p_{Z_{t-3}} + \mu_4 \Delta \ln E_t + \mu_5 ECM + \mu_6 \Delta \ln r1_t + \varepsilon_t \quad (ค)$$

โดยกำหนดให้

Q_{Z_t} = ปริมาณการนำเข้าน้ำยางข้นในช่วงเวลา t

$Q_{Z_{t-i}}$ = ปริมาณการนำเข้าน้ำยางข้นในช่วงเวลาที่ $t - i$

$p_{Z_{t-i}}$ = ราคาน้ำยางข้นในช่วงเวลา $t - i$

E_{t-i} = อัตราแลกเปลี่ยนเงินบาทต่อเงินหยวนในช่วงเวลา $t - i$

ECM = Error Correction Mechanism

MA(i) = Moving Average ระยะเวลาที่ระดับ i

rl_{t-i} = รายได้ประชาชาติ (GDP) ของจีนในช่วงเวลาที่ $t - i$

ε_t = ค่าความคลาดเคลื่อนในช่วงเวลาที่ t

Δ = ผลต่างลำดับที่หนึ่ง (First Difference)

สำหรับการวิเคราะห์ผลลัพธ์ในการคำนวณจากสมการข้างต้นทั้ง 3 สมการนั้น จะนำมาซึ่งค่าสัมประสิทธิ์ของตัวแปรอิสระ ซึ่งจะแสดงในทิศทางของผลกระทบของตัวแปรอิสระที่มีผลต่อตัวแปรตามว่าสอดคล้องกับอุปสงค์ที่ตั้งไว้หรือไม่ จะช่วยให้ทราบถึงปัจจัยต่างๆ ที่มีอิทธิพลต่อการกำหนดอุปสงค์การนำเข้ายางพาราของประเทศจีนจากประเทศไทย

ผลการทดสอบความเป็น Stationary ของข้อมูล ซึ่งในการศึกษาครั้งนี้จะนำข้อมูลอนุกรม

E_t

เวลาจำนวน 9 ตัวแปร ได้แก่ ปริมาณการนำเข้ายางแผ่นรมควัน (Q_{x_t}), ปริมาณการนำเข้ายางแท่ง (Q_{y_t}), ปริมาณการนำเข้าน้ำยางข้น (Q_{Z_t}), ปริมาณการผลิตยางรถยนต์และยางรถบรรทุกของประเทศจีน (q_{T_t}), ราคายางแผ่นรมควัน (p_{x_t}),ราคายางแท่ง (p_{y_t}), ราคาน้ำยางข้น (p_{Z_t}), อัตราแลกเปลี่ยนเงินบาทต่อเงินหยวน (E_t) และรายได้ประชาชาติ (GDP) ของประเทศจีน (I_t) มาทำการศึกษาในสมการ ARIMAX ซึ่งจะทำการทดสอบคุณสมบัติความเป็น Stationary ด้วยวิธี Unit Root Test ตามวิธีของ Augmented Dickey Fuller Test (ADF) ซึ่งถ้าหากข้อมูลใดมีคุณสมบัติเป็น Non– stationary หรือมีคุณสมบัติเป็น Unit Root จะต้องทำการ Difference เพื่อให้มีคุณสมบัติเป็น Stationary ก่อนจะนำข้อมูลมาทำการศึกษาต่อไป แสดงผลการทดสอบดังต่อไปนี้

ตารางที่ 1 แสดงผลการทดสอบ Unit Root ของข้อมูลระดับปกติ (At Level)

ตัวแปร	Lag	ADF Test	MacKinnon Critical Value			Status
			1%	5%	10%	
Q_{x_t}	1	-3.2138	-4.2191	-3.5331	-3.1983	I(0)
Q_{y_t}	1	-2.4634	-4.2191	-3.5331	-3.1983	I(0)
Q_{Z_t}	1	-1.3101	-4.2191	-3.5331	-3.1983	I(0)
q_{T_t}	1	-3.2676	-4.2191	-3.5331	-3.1983	I(0)
p_{x_t}	1	-2.6385	-4.2191	-3.5331	-3.1983	I(0)
p_{y_t}	1	-1.4694	-4.2191	-3.5331	-3.1983	I(0)
p_{Z_t}	1	-1.7578	-4.2191	-3.5331	-3.1983	I(0)
E_t	1	-1.9339	-4.2191	-3.5331	-3.1983	I(0)
I_t	1	-8.9689	–4.2191	-3.5331	-3.1983	I(0)

หมายเหตุ : ตัวแปรที่ใช้ทดสอบทุกตัวอยู่ในรูป Logarithm

I_t มีคุณสมบัติเป็น Trend Stationary (โดยที่ I_t คือรายได้ประชาชาติของจีน)

จากตารางที่ 1 พบว่าค่า ADF Test Statistic ของระดับ (Level) อนุกรมเวลานั้นมีคุณสมบัติ Non–stationary โดยค่าที่คำนวณได้จากวิธี ADF ทุกตัวมีค่าต่ำกว่าค่าวิกฤต (Critical) ที่ได้จากตาราง ณ ระดับความมีนัยสำคัญ 1% และ 5% ซึ่งจากข้อเสนอแนะของ Box – Jenkins ที่ว่า ข้อมูลอนุกรมเวลาใดก็ตามที่มีคุณสมบัติ Non–stationary สามารถปรับให้มีคุณสมบัติเป็น Stationary ได้โดยการหาผลต่าง (Differencing) ซึ่งปกติทั่วไปแล้ว ข้อมูลที่ผ่านการ First Differencing จะมีคุณสมบัติเป็น Stationary และเพื่อให้เห็นชัดเจนว่าจะไม่เกิดปัญหา Two–Unit Root การศึกษาจึงทำการทดสอบคุณสมบัติ Unit Root ของข้อมูลที่ผ่านการ First Differencing แล้ว ผลการทดสอบแสดงได้ดังนี้

ตารางที่ 2 แสดงผลการทดสอบ Unit Root ณ ระดับผลต่างลำดับแรก (At First Difference)

ตัวแปร	Lag	ADF Test	MacKinnon Critical Value			Status
			1%	5%	10%	
Q_{x_t}	1	- 6.2169	- 4.2268	- 3.5366	- 3.2003	I(1)
Q_{y_t}	1	- 6.0058	- 4.2268	- 3.5366	- 3.2003	I(1)
Q_{Z_t}	1	- 4.3705	- 4.2268	- 3.5366	- 3.2003	I(1)
q_{T_t}	1	- 5.2999	- 4.2268	- 3.5366	- 3.2003	I(1)
p_{x_t}	1	- 6.6846	- 4.2268	- 3.5366	- 3.2003	I(1)
p_{y_t}	1	- 4.8247	- 4.2268	- 3.5366	- 3.2003	I(1)
p_{Z_t}	1	- 4.6358	- 4.2268	- 3.5366	- 3.2003	I(1)
E_t	1	- 3.4325	- 4.2268	- 3.5366	- 3.2003	I(1)

หมายเหตุ : ตัวแปรที่ใช้ทดสอบทุกตัวอยู่ในรูป Logarithm

จากตารางที่ 2 เป็นผลการทดสอบ Stationary (Unit Root Test) ณ ระดับผลต่างลำดับแรก (At First Difference) ของตัวแปรทุกตัว พบว่า ค่าสัมบูรณ์ของ ADF T–Statistic ของข้อมูลทุกตัวแปรมีค่ามากกว่าค่าสัมบูรณ์ของค่าวิกฤต (MacKinnon Critical Value) แสดงว่าข้อมูลทุกตัวมีคุณสมบัติเป็น Stationary ที่ผลต่างลำดับแรก ที่ระดับนัยสำคัญ 1%, 5% และ 10% แสดงว่าข้อมูลที่ทำการ Differencing แล้วเหมาะสมในการนำไปสร้างสมการ ARIMAX Model เพื่อทำการวิเคราะห์อุปสงค์การนำเข้ายางพาราของประเทศจีนจากประเทศไทย ได้แก่ ยางแผ่นรมควัน ยางแท่ง และน้ำยางข้นต่อไป

ผลการทดสอบความสัมพันธ์ระยะยาว (Co–integration Test) เมื่อทดสอบคุณสมบัติ Stationary กับข้อมูลทุกตัวแล้ว จากนั้นนำข้อมูลมาทดสอบ Co–integration ซึ่งเป็นการทดสอบว่าตัวแปรแต่ละตัวมีความสัมพันธ์เชิงดุลยภาพในระยะยาวหรือไม่ ถ้าหากว่าตัวแปรนั้นๆ ผ่านการทดสอบและยอมรับว่ามี Co–integration กันแล้ว แสดงว่าแม้ในระยะสั้นจะเกิดภาวะที่ทำให้ตัวแปรเหล่านั้นออกจากดุลยภาพ แต่ในระยะยาว ตัวแปรต่างๆ เหล่านั้นจะปรับตัวเข้าสู่ดุลยภาพได้ โดยเงื่อนไขสำหรับทดสอบ Co–integration นี้คือ ตัวแปรที่ทดสอบนั้นจะต้องเป็น Stationary ที่ระดับเดียวกัน หรือ Integrated (I(d)) ในระดับเดียวกัน ซึ่งผลการทดสอบ Co–integration พบว่า ADF (τ–Statistic) มีค่ามากกว่าค่าวิกฤต (MacKinnon Critical Value) ทั้ง 3 สมการ ได้แก่ ยางแผ่นรมควัน ยางแท่ง และน้ำยางข้น ที่ระดับนัยสำคัญ 1%, 5% และ 10% กล่าวคือ ค่า Residual ที่ทดสอบเป็น Stationary แสดงว่าข้อมูลที่ใช้ในสมการอุปสงค์การนำเข้ายางพาราของประเทศจีนจากประเทศไทย ได้แก่ ยางแผ่น-รมควัน ยางแท่ง และน้ำยางข้น จะต้องนำตัวแปร Error Correction Mechanism ร่วมวิเคราะห์ด้วย โดยผลการทดสอบ Co–integration แสดงได้ดังนี้

ตารางที่ 3 แสดงผลการทดสอบ Co–integration โดย Engle – Granger

ตัวแปร	ADF Test Statistic	MacKinnon Critical Value			Status
		1%	5%	10%	
Residual x	-3.3441	- 2.6272	-1.9499	-1.6115	I(0)
Residual y	-3.5094	- 2.6272	-1.9499	-1.6115	I(0)
Residual z	-8.2431	- 2.6272	-1.9499	-1.6115	I(0)

หมายเหตุ : Residual x = Residual ของสมการปริมาณการนำเข้ายางแผ่นรมควัน
Residual y = Residual ของสมการปริมาณการนำเข้ายางแท่ง
Residual z = Residual ของสมการปริมาณการนำเข้าน้ำยางข้น

3. ผลการทดสอบแบบจำลอง ARIMAX จากสมการที่ (ก), สมการที่ (ข) และสมการที่ (ค) ดังนี้

แบบจำลองยางแผ่นรมควัน

$\Delta \ln Q_{x_t}$ = 0.86181 $\Delta \ln Q_{x_{t-1}}$ + 0.32296 $\Delta \ln Q_{x_{t-5}}$ – 0.95722 MA_1 +
(23.12164)***(5.08672) ***
(–42.50558) ***1.52574 MA_2 +
0.60955 MA_5 – 0.12175 $\Delta \ln p_{x_{t-1}}$
+ (23.76910)***(8.60323) ***
(–2.57628)** 0.93264 $\Delta \ln q_{T_t}$ +
0.78845 $\Delta \ln E_{t-4}$ – 0.68147 ECM
+ (3.88292)***(2.73838)***
(–0.68148)***0.000134 $\Delta \ln rl_t$
(9.10267)*** (ง)

R^2	= 0.811650
Adjust R^2	= 0.741019
LM – Statistic	= 7.53944
ARCH Test	= 0.123382
Ramsey RESET Test	= 0.001763
Jarque – Bera	= 0.341645

หมายเหตุ : ค่าในวงเล็บคือ t–statistic
*** มีนัยสำคัญที่ 1%
** มีนัยสำคัญที่ 5%
* มีนัยสำคัญที่ 10%

แบบจำลองยางแท่ง

$\Delta \ln Q_{y_t}$ = 0.34076 $\Delta \ln Q_{y_{t-1}}$ – 0.99002 $MA4$
–1.28113 $\Delta \ln p_{y_{t-1}}$ +(1.47138)
(–595634.7) ***(–3.65073) ***
1.04543 $\Delta \ln q_{T_{t-2}}$ + 1.30037
$\Delta \ln E_{t-5}$ – 0.86857 ECM +
(3.57268)*** (1.79248)*
(–3.48704) ***0.00003 $\Delta \ln rl_{t-2}$
(2.91406)*** (จ)

R^2	= 0.342569
Adjust R^2	= 0.236531
LM – Statistic	= 5.346580
ARCH Test	= 0.551461
Ramsey RESET Test	= 0.444123
Jarque – Bera	= 0.379525

แบบจำลองน้ำยางข้น

$\Delta \ln Q_{z_t}$ = 0.78347 $\Delta \ln Q_{z_{t-5}}$ – 0.88538 $MA5$
+1.03606 $\Delta \ln p_{z_{t-3}}$ +(9.25121)***
(–20.2410) ***(2.10172)**
1.21828 $\Delta \ln E_{t-4}$ – 0.30323 ECM
+ 0.00009 $\Delta \ln rl_t$ (2.12692)**
(–3.33231)*** (4.75533)*** (ฉ)

R^2	= 0.810356
Adjust R^2	= 0.776491
LM – Statistic	= 2.363755
ARCH Test	= 0.709357
Ramsey RESET Test	= 0.171932
Jarque – Bera	= 0.747478

เมื่อได้รูปแบบจำลองการพยากรณ์ที่เป็น The Best Model เรียบร้อยแล้ว ขั้นตอนต่อไปจะนำแบบจำลองที่ประมาณการได้มาพยากรณ์ล่วงหน้าช่วงไตรมาสที่ 1 - 4 ในปี 2548 และทำการตรวจสอบความถูกต้องการพยากรณ์โดยใช้ Root Mean Square Forecast Error แสดงดังตารางที่ 4

ตารางที่ 4 แสดงผลการพยากรณ์ล่วงหน้าช่วงไตรมาสที่ 1 - 4 ในปี 2548 ด้วย The Best Model

ปี พ.ศ.	ไตรมาส	ปริมาณการส่งออกยางพาราแต่ละประเภท			ผลจากการพยากรณ์		
		ยางแผ่นรมควัน	ยางแท่ง	น้ำยางข้น	ยางแผ่นรมควัน	ยางแท่ง	น้ำยางข้น
2548	1	52,011	55,413	48,040	52,000	54,955	49,000
	2	58,324	53,981	39,281	59,945	54,080	40,141
	3	53,215	45,008	25,423	55,084	44,978	23,504
	4	50,453	47,121	35,441	49,897	46,015	32,421
	Root Mean Square Forecast Error				0.05	0.02	0.01

จากตารางที่ 4 พบว่า เมื่อสร้างแบบจำลองให้เป็น The Best Model ได้แล้ว จะส่งผลต่อความถูกต้องสำหรับการพยากรณ์ได้ และการตรวจสอบความถูกต้องของการพยากรณ์วิธีนี้จะใช้ Root Mean Square Forecast Error โดยจะเก็บข้อมูลไว้ตรวจสอบ 1 ชุด นอกเหนือจากที่ใช้สร้างแบบจำลอง เพื่อตรวจสอบว่าแบบจำลองที่สร้างมานั้นเหมาะสมมากน้อยเพียงใด สำหรับการนำไปใช้พยากรณ์ต่อไปในอนาคตได้ แต่สิ่งที่ขาดไม่ได้คือ ในการสร้างแบบจำลอง ARIMA Model จะต้องทดสอบความถูกต้องจาก Correlogram เสมอ

5. สรุป

การสร้างแบบจำลองที่เรียกว่า The Best Model นั้น เป็นแบบจำลองที่เหมาะสมสำหรับนำไปใช้ในการวิเคราะห์ต่างๆ ซึ่งผลของการวิเคราะห์จะผิดพลาดน้อย และทำให้เกิดความน่าเชื่อถือสูง โดยเฉพาะทางด้านเศรษฐกิจ ซึ่งจำเป็นอย่างมากในการนำไปใช้สำหรับการวางแผนนโยบายต่างๆ ตลอดจนการกำหนดกลยุทธ์ทุกๆด้าน เพื่อให้ประเทศชาติเจริญเติบโต ดังนั้น การนำแนวคิดของ Box-Jinkins มาประยุกต์ใช้กับภาคจริง (Actual) จำเป็นอย่างยิ่งจะต้องผ่านขั้นตอนอย่างละเอียดและถูกต้อง และสิ่งที่จะขาดไม่ได้คือจะต้องมีการกำหนดตัวแปรอิสระให้ครบถ้วนและเหมาะสมกับแต่ละแบบจำลองนั้นๆ

ดังนั้น ถ้าผู้วิจัยหรือผู้สนใจศึกษาต้องการนำ ARIMA Model และ ARIMAX Model ไปใช้จำเป็นจะต้องศึกษาให้ละเอียดรอบคอบในแต่ละขั้นตอน เพื่อให้ได้ The Best Model ซึ่งสามารถนำไปประยุกต์ใช้ให้เกิดประโยชน์สูงสุดนั่นเอง

เอกสารอ้างอิง

จินดามาส สุทธิชัยเมธี. **วิเคราะห์ความได้เปรียบโดยเปรียบเทียบและอุปสงค์การนำเข้ายางพาราจากประเทศไทยในตลาดจีน.** งานวิจัย, 2549พฤทธ์สรรค์ สุทธิไชยเมธี. สถิติและการวิเคราะห์เชิงปริมาณ. พิมพ์ครั้งที่ 1, กรุงเทพฯ : บริษัทจามจุรี จำกัด, 2553.

พฤทธ์สรรค์ สุทธิไชยเมธี. **สถิติและการวิเคราะห์เชิงปริมาณขั้นสูง.** พิมพ์ครั้งที่ 1, กรุงเทพฯ : สำนักพิมพ์ดวงแก้ว , 2553.

พฤทธ์สรรค์ สุทธิไชยเมธี. **เศรษฐมิติประยุกต์เพื่อการวิจัย.** พิมพ์ครั้งที่ 1, กรุงเทพฯ : สำนักพิมพ์ดวงแก้ว , 2553.

พฤทธ์สรรค์ สุทธิไชยเมธี. (2553) **วิเคราะห์แนวโน้มการส่งออกยางพาราของประเทศไทยไปตลาดจีน งานวิจัย.** ทุนวิจัยวิทยาลัยเทคโนโลยีสยาม.

Anderson, O.D., “**Time Series Analysis and Forecasting – The Box – Jenkins Approach,**” Butterworths, London, 1975.

Box, George and D. Piece. “Distribution of Autocorrelations in Autoregressive Moving Average Time Series Models.” Journal of the American Statistical Association 65 (1970), 1509-26

Dickey, D. A., “**Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root**”, Econometric (March 1987), 1981, 251-76

Dickey, and W.A. Fuller (1979),”Distribution of the Estimators for Auto–Regressive Time Series with a Unit Root”, journal of American Statistical Association, 74 , pp.427-431.

Drapper, N.R, and Smith, H., “**Applied Regression Analysis,**” 2nd Edition, John Wiley & Sons, New York,1981.

Granger, Clive and P. Newbold. “**Spurious Regressions in Econometrics.**” Journal of Econometrics 2 1974, 111-20.

Granger, E.S., JR. and Mckenzie ED. “**Forecasting Trends in Time Series.**” Management Science Vol 31 . 10 (October 1985) : 1237-46

Johansen, S. and K. Juselius , “Maximum Likelihood Estimation and Inference on Co-integration: With Applications to the Demand for Money.” **Oxford Bulletin of Economics and Statistics 52 (February 1990)**”, 1990.

Kolb, R.A. and Stekler, H.O., “**Are Economic Forecasts Significantly Better Than Naïve Predictions ?** An Appropriate Test,” International Journal of Forecasting, Vol.9, 1993, pp. 117 – 120.

Makridakis, S., **The accuracy of major extrapolation (time series) method. J. of Forecasting**., 1: 1982, 111 – 153.

Montgomery, D.C., Johnson, L.A. and Gardiner, J.S., "**Forecasting and Time Series Analysis,**" 2nd Edition, McGraw – Hill Inc., New York, 1990.

Nelson, C.R., "**Applied Time Series Analysis for Managerial Forecasting**" Holden Day, San Francisco, 1973

Newbold, P. and Granger, C.W.J., "**Experience with Forecasting Univariate Time Series and the Combination of Forecast,**" Journal of Royal Statistical Society A,Vol,137, 1974, pp.131 – 146

Willeam W.S. Wei. **Time Series Analysis : Univariate and Multivariate method.** New York USA: Addison – Wesley Publishing company, 1990.