

การเปรียบเทียบตัวแบบการประมาณค่า
สำหรับข้อมูลอนุกรมเวลา

A comparison of estimate model
for time series data

พฤทธ์สรศักดิ์ สุทธิไชยเมธี*

Pruethsan Suthichaimetee

* อาจารย์ประจำ วิทยาลัยเทคโนโลยีสยาม

* Lecturer Siam Technology College.

บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่อเปรียบเทียบโมเดลการประมาณค่า ระหว่างตัวแบบถดถอย (Regression Model) กับ ตัวแบบเวกเตอร์อัตถถดถอย (Vector Autoregressive Model) สำหรับข้อมูลอนุกรมเวลา โดยผลการเปรียบเทียบพบว่า

ข้อมูลทางเศรษฐกิจส่วนใหญ่ที่นำมาใช้จะมีลักษณะ Non Stationary จึงไม่เหมาะสมที่มีการเลือกใช้โมเดลประมาณค่าด้วย Regression Model เพราะจะส่งผลกระทบต่อการนำผลลัพธ์มาประยุกต์ใช้และส่งผลให้การพยากรณ์มีความคลาดเคลื่อนสูง จึงควรเลือก Vector Autoregressive Model เพราะมีการเลือกใช้เฉพาะข้อมูลที่เป็น Stationary ในระดับเดียวกัน (Level) และได้นำ Error Correction Mechanism (ECM) มาร่วมในโมเดล ซึ่งสามารถอธิบายผลลัพธ์ได้ชัดเจนมากยิ่งขึ้นด้วย

อย่างไรก็ตาม การสร้างตัวแบบที่ดีที่สุดนั้น (The Best Model) จะต้องมีการเลือกแบบจำลองประมาณค่าให้ถูกต้องและเหมาะสมกับประเภทของข้อมูลนั้นๆ ซึ่ง จะส่งผลให้การพยากรณ์เกิดความคลาดเคลื่อนต่ำ และสามารถนำมาใช้ประโยชน์ได้อย่างถูกต้องต่อไป

คำสำคัญ : ความคลาดเคลื่อน/ข้อมูลอนุกรมเวลา/การพยากรณ์

Abstract

This article aims to compare the estimated models for the Regression Model and Vector Autoregressive Model for time series data. The comparison is found that

For most economic data used to look Non Stationary is not appropriate to use a model Estimating Regression Model with because it will affect the implementation and results applying the results to the forecast error is high. Should select the Vector Autoregressive Model, because the information is used only for the Stationary same level (Level) and the Error Correction Mechanism (ECM) to join in the model. The results can be explained more clearly.

However, 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.

Keywords : Error Term / Time Series Data / Forecasting.

1. บทนำ

การสืบค้นหาคำตอบจากสิ่งที่เรากำลังศึกษา หรือจากสิ่งที่ปัญหาต้องแก้ นั่น โดยการประมาณการ หรือประมาณค่า (Estimation) ซึ่งการแสวงหาคำตอบดังกล่าวย่อมที่แตกต่างกันในแต่ละแบบจำลอง (Model) แต่สิ่งหนึ่งที่เราจะต้องปฏิบัติให้เหมือนกันในการประมาณค่า คือ การพยายามทำให้เกิดความคลาดเคลื่อนให้น้อยที่สุด เพื่อให้ผลจากการประมาณค่าสามารถนำไปใช้ให้เกิดประโยชน์สูงสุดต่อไปได้

การประมาณค่า (Estimation) ในแต่ละตัวแบบ (Model) มีด้วยกันหลายวิธี แต่ละวิธีย่อมนำมาใช้แตกต่างกันในแต่ละแบบจำลอง และแตกต่างกันในแต่ละประเภทของข้อมูล ซึ่งข้อมูลอนุกรมเวลา (Time Series Data) ก็เป็นข้อมูลประเภทหนึ่งที่จะต้องระมัดระวังในการประมาณค่า เพราะถ้าเลือกใช้ผิด จะส่งผลให้เกิดความเสียหายในการนำไปใช้ประโยชน์ ตัวอย่างเช่น ข้อมูลที่เกี่ยวข้องกับการวางแผนกำหนดกลยุทธ์ทางด้านเศรษฐกิจของประเทศ เป็นต้น อย่างไรก็ตาม จากการศึกษางานวิจัยต่างๆ ทั้งในประเทศและต่างประเทศ พบว่า มีงานวิจัยจำนวนมากที่ได้เลือกใช้แบบจำลอง และวิธีการประมาณค่าที่แตกต่างกัน สำหรับบทความวิชาการฉบับนี้ จะเลือกคัดสรรแบบจำลองเพื่อใช้ในการเปรียบเทียบกับกัน 2 แบบจำลอง คือ Regression Model และ Vector Autoregressive Model (VAR) ทั้งนี้เพื่อให้เห็นถึงข้อดีข้อเสียของแต่ละแบบจำลอง ตลอดจนความแตกต่างจากคำตอบที่ได้จากแต่ละแบบจำลอง เพื่อเป็นประโยชน์ในการนำไปใช้ต่อไป

2. ประเภทแบบจำลอง

ตัวแบบ (Model) ที่ใช้ในการอธิบายเปรียบเทียบ สำหรับข้อมูลที่เป็น Time Series จะเลือกใช้ 2 แบบจำลอง คือ Regression Model และ Vector Autoregressive Model (VAR) แสดงรายละเอียดดังนี้

2.1 Regression Model

การนำข้อมูล Time Series มาใช้ในการวิเคราะห์ข้อมูล โครงสร้างของโมเดลจะให้ความสำคัญกับค่าความคลาดเคลื่อน โดยรวมความคลาดเคลื่อนของแต่ละจุดเข้าด้วยกันและไม่ต้องคำนึงถึงเครื่องหมาย แสดงได้ดังนี้

$$\sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

จากสมการที่ (1) พบว่า ผลรวมของส่วนที่เบี่ยงเบนไปจากเส้นตรงนั้น จะเบี่ยงเบนไปไม่ว่าทางด้านบวกหรือทางด้านลบ แต่ผลรวมของการเบี่ยงเบนจะเข้าใกล้ศูนย์มากที่สุด ดังนั้น ในการคำนวณเส้นที่ดีที่สุดจะคำนวณโดยใช้วิธี กำลังสองน้อยที่สุด (Ordinary Least Square : OLS) โดยใช้โมเดลประเภท Single Regression Model ดังต่อไปนี้

การหาค่า α โดยการ Take Partial Derivative มุ่งตรงต่อ α และ β_1 แล้ว Set ให้เท่ากับศูนย์

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^n \hat{e}_i^2 = \frac{\partial}{\partial \alpha} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_i)^2$$

Normal Equation

$$n\hat{\alpha} + \sum_{i=1}^n \hat{\beta}_1 X_i = \sum_{i=1}^n Y_i$$

$$\hat{\alpha} \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

จากสมการ Normal Equations สามารถคำนวณหาคำตอบได้หลายวิธีด้วยกัน เช่น การแทนค่า (Substitution) การกำจัดตัวแปร (Systematic Elimination) และ Cramer's Rule เป็นต้น

การคำนวณหาค่า $\hat{\alpha}$ จะคำนวณโดยใช้สูตรหรือไม่ก็ได้ ถ้าไม่ใช้สูตรสามารถคำนวณได้จากการย้ายข้างสมการ Normal Equation ดังนี้

$$\frac{n\hat{\alpha}}{n} = \frac{\sum_{i=1}^n Y_i}{n} - \frac{\sum_{i=1}^n \hat{\beta}_1 X_i}{n}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (3)$$

สำหรับการคำนวณ $\hat{\beta}_1$ สามารถคำนวณได้จากหลายวิธีด้วยกัน เช่น การคำนวณด้วยวิธี Cramer's Rule แสดงได้ดังนี้

$$\begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \quad (4)$$

$$\hat{\beta}_1 = \frac{Cov(X_1 Y)}{Var(X_1)} \quad (5)$$

อย่างไรก็ตาม จากสมการที่ (1) ถึงสมการที่ (5) พบว่าในการประมาณค่าสมการถดถอย หรือคำนวณค่า $\hat{\alpha}$ และ $\hat{\beta}_1$ เพื่อจะได้

ทราบขนาดของการเปลี่ยนแปลงและทิศทางของการเปลี่ยนแปลงของ X_i ที่มีผลต่อ Y_i แต่ผลลัพธ์ที่ได้จะไม่สามารถอธิบายหรือทราบคำตอบได้เลยถ้าหากว่า Regression Line นั้นขาดการทดสอบคุณสมบัติของตัวประมาณค่า หรือ คุณสมบัติของ OLS-Estimator หรือ คุณสมบัติความเป็น Best Linear Unbiased Estimator (BLUE) ตามทฤษฎีของ Gauss-Markov ซึ่งคุณสมบัติความเป็น BLUE ดังนี้

1. $\hat{\beta}_1$ และ $\hat{\alpha}$ เป็นตัวประมาณค่าที่ไม่เอนเอียง (Unbiased)

$$E(\hat{\beta}_1) = \beta_1$$

$$E(\hat{\alpha}) = \alpha$$

2. $\hat{\beta}_1$ และ $\hat{\alpha}$ มีความแปรปรวนต่ำที่สุด

(Minimum Variance) Variance ของ $\hat{\beta}_1$ แสดงได้ดังนี้

$$Var(\hat{\beta}_1) = E(\hat{\beta}_1 - \beta_1)^2$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

Variance ของ $\hat{\alpha}$ แสดงได้ดังนี้

$$Var(\hat{\alpha}) = E(\hat{\alpha} - \alpha)^2$$

$$Var(\hat{\alpha}) = \sigma^2 \left(\frac{\sum_{i=1}^n X_i^2}{n \sum_{i=1}^n x_i^2} \right)$$

$$Cov(\hat{\alpha}, \hat{\beta}_1) = E(\hat{\alpha} - \alpha)(\hat{\beta}_1 - \beta_1)$$

$$= E \left\{ \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} F_i \right) u_i \right\} \left\{ \sum_{i=1}^n F_i u_i \right\}$$

$$Cov(\hat{\alpha}, \hat{\beta}_1) = \frac{-\sigma^2 \bar{X}}{\sum_{i=1}^n x_i^2}$$

$$Cov(\hat{\alpha}, \hat{\beta}_1) \neq 0$$

ซึ่งตัวประมาณค่าของ $\sigma^2 = \hat{\sigma}^2$

$$Cov(\hat{\alpha}, \hat{\beta}_1) = \frac{-\hat{\sigma}^2 \bar{X}}{\sum_{i=1}^n x_i^2}$$

ดังนั้น จะพบว่า ตัวประมาณค่า $\hat{\alpha}, \hat{\beta}_1$ เป็นตัว OLS-Estimator ที่ดีที่สุด (Best Linear Unbiased Estimator : BLUE)

2.2 Vector Autoregressive Model (VAR)

รูปแบบมาตรฐานของแบบจำลอง VAR (Standard VAR) ประกอบด้วยแปรทางขวามือของทุกสมการมีจำนวนเท่ากันและอยู่ในรูปของตัวแปรล่าช้า (Lagged Variables) ของตัวแปรภายใน ดังนี้

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_n Y_{t-n} + u_t$$

Y_t = เวกเตอร์ของตัวแปรภายใน

Y_{t-1} = เวกเตอร์ของตัวแปรภายในล่าช้า

(Lagged Values of Endogenous Variables)

α_i = เมตริกซ์ของสัมประสิทธิ์หรือค่าพารามิเตอร์

มิเตอร์ของตัวแปรภายในในรูปล่าช้า

u_t = เวกเตอร์ของตัวรบกวน (Disturbance

Terms) หรือเรียกว่า Impulses หรือ Innovations หรือ Shocks

จากสมการที่ (1) ค่า $u_{1t} \sim N(0, \sigma_{11})$ และ $u_{2t} \sim N(0, \sigma_{22})$ ซึ่งค่าตัวรบกวนแต่ละตัวจะมีความสัมพันธ์กัน แต่จะไม่มีความสัมพันธ์กับตัวแปรล่าช้าของตัวแปรภายในทุกตัว และตัวแปรในสมการจะต้องมีความเป็นเหตุเป็นผลกัน ประกอบกับค่าสัมประสิทธิ์ที่ประมาณได้นั้นจะต้องมีคุณสมบัติความเป็น BLUE เพื่อไม่ให้เกิดปัญหา Spurious Estimate

3. การปรับข้อมูลให้มีลักษณะเป็น Stationary

ตัวแปรที่มีลักษณะเป็น NonStationary

จะมีค่าผันแปรไปตามเวลา ซึ่งอาจมีค่าเพิ่มขึ้นหรือลดลงก็ได้เมื่อเวลาเปลี่ยนแปลงไป กล่าวอีกนัยคือ ตัวแปรที่มีลักษณะ Non Stationary จะได้รับอิทธิพลจากเวลา ดังนั้น หลักการในการปรับตัวแปรให้เป็น Stationary ก็คือต้องตัดหรือแยกผลของตัวแปรเวลา (Time Variable) ออกจากข้อมูลที่สังเกต ซึ่งกระทำได้ 2 วิธี ดังนี้

1. การปรับข้อมูลให้อยู่ในรูปผลต่าง

ลำดับที่หนึ่งหรือสูงกว่า (Differencing the Series Once or More)

2. การปรับข้อมูลด้วยวิธี Co-integration

การทดสอบว่า ตัวแปรที่พิจารณามีคุณสมบัติเป็น Co-integration หรือไม่ มีวิธีการตรวจสอบที่นิยมใช้ 2 วิธีคือ Engle-Granger (EG) หรือ Augmented Engle - Granger (AEG) และวิธี Johansen and Juselius Test

1. การปรับข้อมูลให้อยู่ในรูปผลต่าง

การตรวจสอบคุณสมบัติความเป็น

Stationary โดยการตรวจสอบ “ A Unit Root Test” เพราะข้อมูลอนุกรมเวลา (Time - Series Data) มีลักษณะเป็นตัวแปรเชิงสุ่ม (Random Variables) เมื่อนำมาเรียงกันตามลำดับเวลาที่เกิดขึ้น (Order in Time) เรียกว่า Stochastic หรือ Random Process โดยที่การผันแปรของข้อมูลในลักษณะที่เป็น Stochastic มี 2 แบบ คือ Stationary Stochastic Process กับ Non Stationary Stochastic Process โดยคุณสมบัติของ Stationary และ Non Stationary แสดงได้ดังนี้

Stationary Stochastic Process คือ ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ย หรือค่าความคาดหวัง (Mean or Expect Value) และค่าความแปรปรวน (Variance) คงที่ตลอดช่วงเวลา (Constant Over Time) และค่าความแปรปรวนร่วม (Covariance) จะไม่ขึ้นอยู่กับเวลา แต่จะขึ้นอยู่กับระยะหรือช่วงห่างของช่วงเวลา

$$\text{Mean: } E(Y_t) = E(Y_{t+k}) = \mu$$

$$\begin{aligned} \text{Variance: } VAR(Y_t) &= E(Y_{t-\mu})^2 \\ &= E(Y_{t-k-\mu})^2 = \sigma^2 \end{aligned}$$

$$\text{Covariance: } E\{Y_{t-\mu})(Y_{t+k-\mu})\} = \gamma_k$$

Non Stationary Stochastic Process คือ ข้อมูลที่มีค่าความคาดหวังและความแปรปรวนไม่คงที่ แต่จะเปลี่ยนไปตามเวลา

$$\text{Mean: } E(Y_t) = E(Y_{t+k}) = t\mu$$

$$\begin{aligned} \text{Variance: } VAR(Y_t) &= E(Y_{t-\mu})^2 \\ &= E(Y_{t-k-\mu})^2 = t\sigma^2 \end{aligned}$$

$$\text{Covariance: } E\{Y_{t-\mu})(Y_{t+k-\mu})\} = \gamma_k$$

สำหรับการคำนวณ Stationary ของ ข้อมูลที่นำมาวิเคราะห์พิจารณาเฉพาะ First และ Second Moment สำหรับข้อมูลที่เป็น Stationary ค่าความคาดหวังหรือค่าเฉลี่ย (Mean) กับค่าความแปรปรวน (Variance) จะเป็นอิสระกับ “เวลา” และค่าความแปรปรวนร่วมจะขึ้นอยู่กับค่าระยะห่างของช่วงเวลา (Lag)

การเลือกรูปแบบการทดสอบ Stationary

1. Dickey-Fuller Test ซึ่งเป็นสมการถดถอยในรูปของผลต่าง (Difference Regression) ที่เรียกว่า First Order Autoregressive Process ภายใต้รูปแบบสมการดังนี้

สมมติฐาน (Hypothesis)

$$H_0 : \delta = 0, \text{ Non Stationary}$$

$$H_1 : \delta < 1, \text{ Stationary}$$

$\Delta Y_t = \delta Y_{t-1} + \epsilon_t$ เรียกว่า Random Walk Process หรือ Pure Random Walk

$\Delta Y_t = \alpha_1 + \delta Y_{t-1} + \epsilon_t$ เรียกว่า Random Walk with Drift หรือ เป็นการเพิ่มค่า Intercept

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

2. Augment Dickey-Fuller Test เป็น

รูปแบบที่นิยมนำมาใช้ในกรณี ϵ_t ขาดคุณสมบัติความเป็น White Noise หรือเกิดปัญหาที่เรียกว่า Autocorrelation คือมีความสัมพันธ์ในอันดับสูงขึ้น (Higher Order Autoregressive Process) ดังนั้นเพื่อแก้ปัญหาดังกล่าวจึงกำหนดรูปแบบสมการใหม่ดังนี้

$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=2}^p \beta_i \Delta Y_{t-i+1} + \epsilon_t$$

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

$$\begin{aligned} \Delta Y_t &= \alpha_1 + \alpha_2 T + \delta Y_{t-1} + \sum_{i=2}^p \beta_i \Delta Y_{t-i+1} \\ &+ \epsilon_t \end{aligned}$$

2. การปรับข้อมูลด้วยวิธี Co-integration

Model Approach

การประมาณแบบจำลองที่มีตัวแปร
ลักษณะที่เป็น Non Stationary ด้วยวิธี Co-
integration จะช่วยแก้ปัญหา Spurious
Regression ซึ่งประกอบด้วยวิธี Engle-
Granger (EG) หรือ Augmented Engle-
Granger (AEG) Test และวิธี Johansen
and Juselius เป็นการตรวจสอบโดยพิจารณา
คุณสมบัติของค่าตัวแปรที่คำนวณได้จาก
สมการถดถอย Co-integrating Regression
ว่าเป็น Stationary หรือไม่ ถ้าพบว่าค่าตัว
แปร ซึ่งเป็นส่วนประกอบเชิงเส้นตรง
(Linear Combination) ของตัวแปรมีลักษณะ
เป็น Stationary แสดงว่า ตัวแปรที่พิจารณา
มีคุณสมบัติเป็น Co-integration วิธีการทำ
คือการทดสอบ Stationary ของค่าความคลาด
เคลื่อน (Error Term)

$$u_t = Y_t - \alpha - \beta X_t$$

สมมติฐาน (Hypothesis)

$$H_0 : \hat{u}_t \sim I(1) \text{ NonStationary}$$

$$H_1 : \hat{u}_t \sim I(0) , \text{ Stationary}$$

$$|\tau_{value}| > |\tau_{critical}| \text{ จะ Reject และ } H_0$$

Accept H_1 ที่ระดับนัยสำคัญ หรือแสดงว่า
ข้อมูลที่น่ามาวิเคราะห์มีความเป็น Stationary

กำหนด VAR ในรูปแบบมาตรฐาน
(Standard VAR) ดังนี้

$$\begin{bmatrix} Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \bar{Y} \\ \bar{Z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

VAR model มีขั้นตอนในการวิเคราะห์

4 ขั้นตอนดังต่อไปนี้

1. กำหนดรูปแบบ (Identification)

2. ประมาณค่าพารามิเตอร์ (Parameter
Estimation)

3. การประเมินผลการคำนวณของตัวแบบ

VAR

การประเมินตัวแบบที่คำนวณได้ของ

ตัวแบบ VAR จะมีความแตกต่างจากผลการ
ประเมินทางด้านสถิติทั่วไป ภายใต้ตัว VAR
ประเมินผลโดยใช้ F - Test คือการทดสอบ
ระดับความมีนัยสำคัญของทั้งกลุ่มของตัวแปร
ล่าช้า ส่วนการประเมินโดยใช้ t-statistic เพื่อ
ทดสอบนัยสำคัญของตัวแปรแต่ละตัวจะไม่มี
นัยสำคัญ ทั้งนี้เพราะว่าตัวแปรที่แสดงในรูป
ของตัวแปรล่าช้าในหลายๆ คาบเวลา จะทำให้
เกิดปัญหา Multicollinearity และในการนำ
VAR มาใช้ในทางสถิตินั้นจำเป็นต้องใช้
Impulse Response Function เพื่อพิจารณา
การสนองตอบของตัวแปรต่อการเปลี่ยนแปลง
ของตัวแปรด้วย

4. การพยากรณ์ (Forecast)

ตาม Model ที่ได้มีการกำหนดขึ้นเมื่อ

ได้รูปแบบที่ได้ทำการทดสอบว่าเหมาะสม ก็จะใช้
รูปแบบนั้นเพื่อการพยากรณ์ค่าในอนาคต ดัง
นั้นเมื่อได้ค่าจริง ณ ช่วงเวลาที่ัดไปแล้วควรนำ
ค่า ณ เวลาดังกล่าวไปปรับสมการพยากรณ์
เพื่อหาค่าที่เหมาะสมที่สุดและใกล้เคียงกับค่า
พยากรณ์มากที่สุด

ตัวอย่างโมเดลประมาณค่า

ตัวอย่างงานวิจัยที่ได้เลือกใช้ Regression

Model และ Vector Autoregressive Model
คือ โมเดลดัชนีผลผลิตอุตสาหกรรมของ

ประเทศไทย (ตัวแปรตาม) ที่ขึ้นอยู่กับปัจจัยที่มีอิทธิพลต่อการเปลี่ยนแปลงคือ มูลค่าการนำเข้าสินค้าวัตถุดิบและทุน (ตัวแปรอิสระ) โดยใช้ข้อมูลทุติยภูมิ (Secondary Data)

รายไตรมาสตั้งแต่ไตรมาสที่ 1 พ.ศ. 2535 – ไตรมาสที่ 4 พ.ศ. 2552 รวมทั้งหมด 72 ข้อมูล (Observation) โดยแบ่งการวิเคราะห์แต่ละโมเดลแสดงได้ดังนี้

กำหนดให้

x แทน ดัชนีผลผลิตอุตสาหกรรมของประเทศไทย

mc แทน มูลค่าการนำเข้าสินค้าวัตถุดิบและทุน

ผลการวิเคราะห์จาก Regression Model

$$lx = 0.04 - 0.43lmc + 0.022AR(1) \quad (1)$$

(0.20) (-7.54)** (2.03)

$$F\text{-Statistic} = 12.79 \quad p\text{-value} = 0.00$$

$$R^2 = 0.89 \quad \text{Adjust } R^2 = 0.81$$

$$\text{Durbin Watson} = 2.31$$

ผลการวิเคราะห์จาก Vector Autoregressive Model

$$\Delta lx = 0.007 - 0.60 \Delta lx_{t-1} + 0.03 \Delta lmc_{t-1} + 0.03ECM \quad (2)$$

(3.50) (-7.72) (2.53) (0.40)

$$F\text{-Statistic} = 16.25$$

$$\Delta lmc = 0.25 - 1.55 \Delta lmc_{t-1} - 2.68 \Delta lx_{t-1} + 0.25ECM$$

(1.37) (-6.67) (-2.49) (2.01)

$$F\text{-Statistic} = 27.25$$

จากผลการวิเคราะห์ในสมการที่ (1) และสมการที่ (2) พบว่า รูปแบบสมการทั้งสอง เป็นรูปแบบสมการที่เรียกว่า The Best Model ดังนั้น ในการเปรียบเทียบประสิทธิภาพของ โมเดลสามารถกระทำได้โดยการพยากรณ์ (Forecasting)

อย่างไรก็ตามในการเลือกรูปแบบของการพยากรณ์จะต้องคำนึงถึงความแม่นยำที่ได้จากการใช้รูปแบบต่างๆ ด้วย ซึ่งการพยากรณ์ จะต้องให้ความคลาดเคลื่อนน้อยที่สุด ดังนั้น

รูปแบบการพยากรณ์ที่ควรเลือกก็ควรเป็นรูปแบบที่มีความคลาดเคลื่อนน้อยที่สุด หรืออาจวิเคราะห์โดยใช้ค่าสถิติวัดความถูกต้องก็ได้ สถิติที่ใช้วัดความถูกต้อง Root Mean Square Error หรือกล่าวคือ หากรูปแบบสมการของการพยากรณ์ใด มีค่า Root Mean Square Error ต่ำจึงน่าจะเป็นหนทางหนึ่งในการเลือกสมการ ดังกล่าวมาใช้ในการพยากรณ์ แสดงไว้ในตารางที่ 1 ดังต่อไปนี้

ตารางที่ 1 การวัดความถูกต้องในการเลือกสมการที่ใช้ในการพยากรณ์ด้วย Root Mean Square Error

เทคนิคของการพยากรณ์	Root Mean Square Error $\sqrt{\sum (X_{t+i}^f - X_{t+i})^2}$
Regression Model	0.170
Vector Autoregressive Model : VAR	0.008

ที่มา : จากการคำนวณ

จากตารางที่ 1 พบว่า Vector Autoregressive Model : VAR เป็นโมเดลที่ทำการประมาณค่าแล้วมีค่าความคลาดเคลื่อนต่ำกว่า Regression Model เมื่อเทียบกับข้อมูลจริง สำหรับข้อมูลที่เป็นอนุกรมเวลา

สรุป

สำหรับการเปรียบเทียบประสิทธิภาพของ Regression Model กับ Vector Autoregressive Model นั้น พบว่า แบบ

จำลอง Regression Model ให้ความสำคัญกับการประมาณค่าปลอม (Spurious Estimation) ของข้อมูลตามเงื่อนไขความเป็น BLUE ประกอบกับพยายามขจัดปัญหา Autocorrelation Multicollinearity และ Heteroscedasticity การประมาณค่า (Estimation) สำหรับข้อมูลอนุกรมเวลา (Time Series) เพื่อให้ผลการประมาณค่า (Result) ใกล้เคียงกับข้อมูลจริงมากที่สุด (Actual) ถือได้ว่าเป็น Model ที่พยายามเนื่องจากไม่ได้คำนึงถึงความเป็น Stationary

ของข้อมูล ซึ่งข้อมูลที่น่ามาวิเคราะห์จำเป็นจะต้องเป็นข้อมูลที่เป็น Stationary เพราะ ค่าเฉลี่ยของตัวแปร (Long Run Mean Level) เพียงชั่วคราว เมื่อเวลาเปลี่ยนไปค่าที่เบี่ยงเบนจะกลับเข้ามาสู่ค่าเฉลี่ยของอนุกรมเวลานั้นจะเป็นค่าคงที่ แต่ถ้าตัวแปรมีลักษณะ Non Stationary จะเกิดผลในทางตรงข้าม และ ผลกระทบที่เกิดขึ้นจะส่งผลกระทบต่อแนวโน้มตัวแปรที่เกิดขึ้นในช่วงเวลาอื่นๆ โดยจะทำให้ค่าเฉลี่ยแปรผันไปเรื่อยๆ ตามเวลาที่เปลี่ยนไป แสดงให้เห็นได้ว่าข้อมูลที่เป็น Non Stationary จะไม่มีค่าเฉลี่ยระยะยาว (Long Run Mean Level) นั่นเอง อย่างไรก็ตาม การคำนวณหาสมการถดถอยที่ใช้ข้อมูลอนุกรมเวลาที่มีลักษณะเป็น Non Stationary แม้ผลการคำนวณให้ค่าทางสถิติ คือ R^2 มีค่าสูง ค่า t -statistic มีนัยสำคัญทางสถิติก็ตาม แต่ไม่สามารถนำไปอธิบายใดๆ ได้เหมือนกรณีวิเคราะห์สมการถดถอยทั่วไป ซึ่งจะแตกต่างจาก Vector Autoregressive Model (VAR) ที่จะทำให้ความสำคัญกับข้อมูลที่น่ามาวิเคราะห์นั้นจะต้องมีคุณสมบัติความเป็น Stationary เท่านั้น ส่งผลให้ผลการประมาณค่า (Estimation) มีความถูกต้องน่าเชื่อถือสูงกว่า Regression Model แม้ว่าค่าสถิติ R^2 มีค่าไม่สูง และ/หรือ ค่า t -statistic ไม่มีนัยสำคัญทางสถิติก็ตาม แบบจำลองที่ถือได้ว่ามีความถูกต้องและเหมาะสมสามารถนำมาอธิบายได้นั้น เรียกว่า The Best Model และสิ่งหนึ่งที่สามารถนำมาใช้ในการวัด The Best Model คือ ค่าพยากรณ์

อย่างไรก็ตาม ถ้าเปรียบเทียบความถูกต้องของค่าที่ได้จากการพยากรณ์ (Forecast) กับข้อมูลจริง (Actual Data) ใน Regression Model และ Vector Autoregressive Model พบว่าการเลือกใช้ Vector Autoregressive Model จะทำให้ผลการพยากรณ์ใกล้เคียงกับความเป็นจริงมากกว่า การเลือกใช้ Regression Model

ข้อเสนอแนะ

การเลือกแบบจำลองการประมาณค่าไม่ว่าวิธีใด ๆ ก็ตาม จำเป็นจะต้องเลือกให้เหมาะสมกับวัตถุประสงค์และจะต้องคำนึงถึงประโยชน์ที่ได้รับเป็นสำคัญ เพราะแบบจำลองที่แตกต่างกันย่อมให้ผลลัพธ์ (Out put) แตกต่างกันด้วย

สำหรับการนำ Vector Autoregressive Model: VAR มาใช้วิเคราะห์ในทางธุรกิจต่างๆ หรือในเศรษฐกิจนั้น บางกรณีจำเป็นจะต้องคำนวณค่าพารามิเตอร์โดยใช้ Structural VAR เพราะจะทำให้ผลการคำนวณที่ได้รับมีคุณสมบัติเป็น Efficiency นอกจากนี้ยังมีแนวทางแก้ไขคือ คำนวณค่าพารามิเตอร์ของ Structural VAR โดยผ่านค่าพารามิเตอร์ของ Standard VAR ที่คำนวณได้โดยวิธี OLS เรียกว่า Indirect Least Squares แต่ถ้าผลการคำนวณดังกล่าวส่งผลให้เกิดปัญหา Under Identification ผู้วิจัยอาจจะหันมาเลือกแนวทางแก้ไขอย่างอื่น เช่น Choleski Decomposition เป็นวิธีกำหนดเงื่อนไขเกี่ยวกับค่าพารามิเตอร์ของตัวแบบ Structural VAR เพิ่มเติม นั่นเอง เป็นต้น

เอกสารอ้างอิง

- Akaike, H, "**Autoregressive Model Fitting for Control**" Annals of the Institute of Statistical Mathematics, Vol.22, 1970, pp. 163-180,
- Anderson, O.D., "**Time Series Analysis and Forecasting – The Box – Jenkins Approach,**" Butterworths, London, 1975.
- Armstrong, J.S. and ed. **Principles of Forecasting: A Handbook for Researchers and Practitioners.** Boston : Kulwar,1993.
- Bartlett, M.S., "**on the Theoretical Specification of Sampling Properties of Auto correlated Time Series,**" **Journal of Royal Statistical Society,** B8,Vol, 1946, 27
- Christ, C.F., "**Econometric Models and Methods,**" John Wiley & Son, New York, 1986.
- Croxtan, F.E., Cowden, D.J. and Klein, S., "**Applied General Statistic,**" 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey, 1967
- DeJong, D.V., J. C. Nankervis, N. E. Savin and H. Whiteman , "**The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors,**" **Journal of Econometrics,** 53, 1992, pp. 323-343.
- 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,"Distribution of the Estimators for Auto –Regressive Time Series with a Unit Root", **journal of American Statistical Association,** 74 , 1979, 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
- Gujarati, D.N., **Basic Econometrics.** Third Edition, Mc Graw-Hill, Now York. 1995.
- Hendall, M.G., "**Time Series,**" Griffin, London.1973.

- Johansen, S. and K. Juselius , "Maximum Likelihood Estimation and Inference on Cointegration: 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
- Pindyck, R.S. and Rubinfeld, D.L., "**Econometric Models and Economic Forecast,**" 3rd Edition, McGraw – Hill, New York, 1991.
- Sims, Christopher. "**Bayesian Skepticism on Unit Root Econometrics.**" Journal of Economic Dynamics and Control 12, 1998, 463-74
- Willam W.S. Wei. **Time Series Analysis : Univariate and Multivariate method.** New York USA: Addison – Wesley Publishing company, 1990.

