บทที่ 1

บทนำ

ที่มาและความสำคัญของปัญหา

การวิเคราะห์อนุกรมเวลา(Time Series Analysis) เป็นวิธีการทางสถิติวิธีหนึ่งที่สามารถพยากรณ์หรือประมาณค่าในอนาคต และแสดงให้เห็นถึงการเปลี่ยนแปลงของสิ่งที่สนใจในแต่ละช่วงเวลาที่เปลี่ยนไป(โดยอาศัยข้อมูลในอดีต) ซึ่งมีบทบาทสำคัญมากในการวางแผนการดำเนินงานต่างๆ โดยเฉพาะทางด้านธุรกิจและเศรษฐศาสตร์ เช่น การวางแผนการผลิต การวางแผนการตลาด การวางแผนการจัดการบุคลากร และ การวางแผนการจัดการสินค้าคงคลัง เป็นต้น การนำเทคนิคการวิเคราะห์อนุกรมเวลามาใช้เพื่อให้ได้ค่าพยากรณ์ที่เชื่อถือได้สูง จำเป็นต้องเลือกวิธีการพยากรณ์ที่เหมาะสมกับข้อมูลนั้นๆ ซึ่งเทคนิคการวิเคราะห์อนุกรมเวลามีอยู่หลายวิธี เช่นเทคนิคการทำให้เรียบ(Smoothing Techniques) การกรองแบบปรับได้(Adaptive Filtering) วิธีอนุกรมเวลาแบบคลาสสิค(Classical Decomposition Methods) และวิธีอนุกรมเวลาแบบบอกซ์-เจนกินส์(Box-Jenkins Method) เป็นต้น

นอกเหนือจากการเลือกใช้เทคนิคที่เหมาะสมแล้ว ข้อมูลในอดีตก็เป็นอีกปัจจัยหนึ่งที่ส่งผลถึงความถูกต้องและเชื่อถือได้ของค่าพยากรณ์ เนื่องจากการวิเคราะห์อนุกรมเวลาจะต้องอาศัยข้อมูลหรือค่าสังเกตในอดีตที่ผ่านมา แต่ผู้วิเคราะห์อาจประสบปัญหาเกี่ยวกับข้อมูล เช่น ข้อมูลบางค่ามีค่าสูงหรือต่ำกว่าปกติซึ่งลักษณะข้อมูลที่มีค่าผิดปกตินี้อาจเป็นผลเนื่องมาจากลักษณะตามธรรมชาติของค่าสังเกตซึ่งไม่สามารถหลีกเลี่ยงได้ ถึงแม้ว่าเราจะควบคุมการวัดค่าเป็นอย่างดี เช่น การเปลี่ยนนโยบายทางเศรษฐกิจที่มีต่อราคาสินค้า ผลกระทบจากการลดค่าของเงิน ผลกระทบจากการลดราคาสินค้าที่มีต่อยอดขาย ผลกระทบจากการนัดหยุดงานของพนักงานในบริษัทที่มีต่อปริมาณการผลิต ฯลฯ

หนทางหนึ่งในการแก้ปัญหาข้อมูลผิดปกติ คือ การหาตำแหน่ง และประเภทของข้อมูลที่ผิดปกตินั้น แล้วใช้แนวทางของตัวแบบของอินเตอร์เวนชัน(Intervention Models) ในการปรับข้อมูลผิดปกตินั้นให้เหมาะสม ซึ่งวิธีการนี้จำเป็นต้องใช้ขบวนการทำซ้ำ(Iterations) ระหว่างขั้นตอนของการตรวจสอบข้อมูลผิดปกติ และการประมาณค่าพารามิเตอร์ของตัวแบบอินเตอร์เวนชัน(Intervention Model)

ในปี ค.ศ.1993 Chung Chen และ Lon-Mu Liu ได้นำเสนอวิธีการประมาณค่าพารามิเตอร์ร่วมกับการประมาณค่าผิดปกติในอนุกรมเวลา (Joint Estimation of Model Parameters and Outlier Effects in Time Series [1]) มาใช้กับข้อมูลในลักษณะข้างต้น โดยวิธีประมาณค่าที่เสนอคือ วิธีตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimators Method : MLE) หลักการของวิธีตัวประมาณภาวะน่าจะเป็นสูงสุดคือการหาค่าตัวประมาณที่ทำให้สมการภาวะน่าจะเป็นมีค่าสูงที่สุด จากหลักการนี้สิ่งที่จำเป็นต้องทราบคือสมการภาวะน่าจะเป็นของตัวแบบ

นอกเหนือจากวิธีการประมาณค่าพารามิเตอร์โดยวิธีการข้างต้นแล้วยังมีวิธีการประมาณวิธีอื่นอีก เช่น วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS) วิธีตัวประมาณเบส์ (Baye's Estimators Method) และวิธีตัวประมาณเอ็ม (M-Estimators Method) เป็นต้น โดยแต่ละวิธีจะมีหลักการในการหาตัวประมาณที่แตกต่างกันไป วิธีกำลังสองน้อยที่สุดเป็นวิธีการหนึ่งในการประมาณค่าพารามิเตอร์ ที่ไม่จำเป็นต้องทราบสมการภาวะน่าจะเป็นของตัวแบบเหมือนกับวิธีตัวประมาณภาวะน่าจะเป็นสูงสุด นอกจากนี้วิธีกำลังสองน้อยที่สุดยังเป็นวิธีที่มีขั้นตอนในการคำนวนที่ไม่ซับซ้อนโดยมีหลักการคือการทำให้ผลรวมกำลังสองของความคลาดเคลื่อนภายใต้พารามิเตอร์ของตัวแบบมีค่าต่ำสุด ดังนั้นจึงเป็นเรื่องที่น่าสนใจว่า การประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุดในการประมาณค่าพารามิเตอร์ร่วมกับการประมาณค่าผิดปกติในอนุกรมเวลา กับวิธีการประมาณค่าพารามิเตอร์โดยวิธีตัวประมาณภาวะน่าจะเป็นสูงสุดร่วมกับการประมาณค่าผิดปกติในอนุกรมเวลา จะได้ค่าพารามิเตอร์ที่ให้ค่าพยากรณ์ต่ำกว่ากัน

สำหรับการวิจัยครั้งนี้ ผู้วิจัยมีความสนใจที่จะศึกษาเปรียบเทียบประสิทธิภาพของวิธีประมาณ 3 วิธีคือ

1. วิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล
2. วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว
3. วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว

ข้อมูลอนุกรมเวลาที่ศึกษาในครั้งนี้ เป็นอนุกรมเวลารูปแบบ AR(1) และ MA(1) ที่มีลักษณะคงที่(Stationary) ภายใต้สถานการณ์ทดลองที่มีค่าผิดปกติ แล้วทำการเปรียบเทียบประสิทธิภาพของทุกวิธีด้วยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE) เฉลี่ย 12 คาบเวลาพยากรณ์

---

[1] Chen,C.,and Lui,L. Journal of the American Statistical Association , March 1993,Vol 88 , No. 421 ,pp.284-297

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลา ARMA เมื่อค่าสังเกตมีค่าผิดปกติด้วยวิธีการประมาณ 3 วิธีคือ

1. วิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล
2. วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว
3. วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว

สมมติฐานของการวิจัย

ภายใต้สถานการณ์ที่ค่าสังเกตมีค่าผิดปกติ วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว จะให้ตัวประมาณที่ดีกว่าตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว และวิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล

ข้อตกลงเบื้องต้น

อนุกรมเวลา z_t ที่ศึกษาในครั้งนี้เป็นอนุกรมเวลา ARMA และเป็นอนุกรมเวลาคงที่ ซึ่งเขียนในรูปแบบค่าเฉลี่ยเคลื่อนที่และการถดถอยทั่วไป ARMA(p,q) ได้ดังนี้

$$\Phi(B)z_t = K + \Theta(B)a_t$$

โดยที่

$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$

$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$

$\phi_1, \ldots, \phi_p$ คือสัมประสิทธิ์ความถดถอย(Autoregressive Coefficients)

$\theta_1, \ldots, \theta_q$ คือสัมประสิทธิ์ค่าเฉลี่ยเคลื่อนที่(Moving-Average Coefficients)

B คือตัวถอยหลังเวลา (Backward Shift Operation)

นั่นคือ $B^m z_t = z_{t-m}$

K คือเทอมคงที่

a_t คือตัวแปรสุ่มอิสระและมีการแจกแจงแบบปกติ ค่าเฉลี่ย 0 ความแปรปรวนคงที่ σ_a^2 เรียก a_t ว่า ค่าผิดพลาดสุ่ม หรือกระตุกสุ่ม (Random Shocks)

ขอบเขตของการวิจัย

1. อนุกรมเวลา $\{z_t\}$ เป็นอนุกรมเวลาแบบ 1 ตัวแปร (Univariate Time Series) และมีคุณสมบัติเป็นอนุกรมเวลาคงที่ (Stationary) และศึกษาในรูปแบบ AR(1) [หรือ ARMA (1,0)] และ MA(1) [หรือ ARMA(0,1)]
2. กำหนดค่าพารามิเตอร์ในแต่ละอนุกรมที่จะศึกษาดังนี้

   รูปแบบ AR(1) พารามิเตอร์ : ϕ_1 = 0.3 , 0.5 , 0.6 , 0.7 , 0.8

   รูปแบบ MA(1) พารามิเตอร์ : θ_1 = 0.3 , 0.5 , 0.6 , 0.7 , 0.8
3. ลักษณะของการแจกแจงความคลาดเคลื่อนสุ่ม a_t

   3.1 การแจกแจงแบบปกติ(Normal Distribution)

ฟังก์ชั่นความหนาแน่นอยู่ในรูปของ

$$f(x) = 1/(\sigma\sqrt{2\Pi})\exp[(-1/2)((x-\mu)^2/\sigma^2)] \quad ; -\infty < x < \infty$$

ในที่นี้ให้ $\mu = 0$ และ $\sigma = 10$ [2]

3.2 การแจกแจงแบบปกติปลอมปน(Contaminated Normal Distribution)

ฟังก์ชั่นความหนาแน่นอยู่ในรูปของ

$$f(x) = (1-p).\ N(0,100) + p.N(0, c^2\sigma^2)$$

เมื่อ p คือ สัดส่วนของการปลอมปน(Percent of Contaminate) และกำหนด p = 0.03 , 0.05 , 0.08 และ 0.10

C คือ สเกลแฟคเตอร์(Scale Factor) และกำหนด C = 5 และ 10

3.3 ฟังก์ชันการแจกแจงอยู่ในรูปของ

$$f(x) = (1-p)N(0,100) + p\,L(0,\beta)$$

---

[2] ขนาดของความแปรปรวนไม่มีผลต่อการประมาณค่าพารามิเตอร์ของอนุกรม

เมื่อ p คือ สัดส่วนของการปลอมปน และกำหนด p = 0.03 , 0.05 , 0.08 และ 0.10

$L(0,\beta)$ คือการแจกแจงลาปลาซ (Laplace Distribution) ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \frac{1}{2\beta}\exp\left[\frac{-|x - \alpha|}{\beta}\right]$$

เมื่อ $-\infty < x < \infty$, $-\infty < \alpha < \infty$, $\beta > 0$
โดยที่ $E(x) = \alpha$, $Var(x) = 2\beta^2$
และกำหนด $\alpha = 0$ และ $\beta = 1$ และ 10

4. ขนาดตัวอย่างที่ใช้ในการศึกษามี 4 ระดับ คือ 40 [3] , 60 , 80 และ 120

เกณฑ์การตัดสินใจ

เกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพของตัวประมาณ ภายใต้สถานการณ์ต่างๆที่กำหนดในขอบเขตการวิจัยจะพิจารณาจากค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) เฉลี่ย 12 คาบเวลาพยากรณ์ [4]

$$RMSE = \frac{1}{12}\sum_{t=1}^{12}\sqrt{\sum_{i=1}^{500}\frac{(z_{ti} - \hat{z}_{ti})^2}{500}}$$

z_{ti} คือ ค่าสังเกต ณ คาบเวลา t ในการทำซ้ำรอบที่ i

$\hat{z}_{ti}$ คือ ค่าพยากรณ์ณ คาบเวลา t ในการทำซ้ำรอบที่ i

i คือ จำนวนรอบของการทำซ้ำ ; i = 1,...,500

t คือ คาบเวลาของการพยากรณ์ ; t = 1,2,...,12

---

[3] เลือกขนาดตัวอย่างต่ำสุดเท่ากับ 40 เพื่อให้ข้อมูลเป็นไปตามเงื่อนไขของการแจกแจงปกติ

[4] คาบเวลาหมายถึงช่วงเวลาที่มีระยะห่างเท่าๆกัน เช่น วัน เดือน ปี

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางให้ผู้วิจัยสามารถเลือกใช้วิธีการประมาณค่าพารามิเตอร์ของข้อมูลอนุกรมเวลาได้อย่างเหมาะสมและมีประสิทธิภาพ เมื่อข้อมูลมีค่าผิดปกติ

2. เพื่อเป็นแนวทางในการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของข้อมูลอนุกรมเวลา เมื่อข้อมูลอนุกรมเวลามีผลกระทบจากปัจจัยอื่น เช่น อิทธิพลของฤดูกาล