บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การพยากรณ์หมายถึงการคาดคะเนหรือทำนายการเกิดของเหตุการณ์หรือสภาพการณ์ต่าง ๆ ในอนาคต โดยการพยากรณ์จะทำจากการศึกษาแนวโน้มและรูปแบบการเกิดของเหตุการณ์หรือสภาพการณ์จากข้อมูลในอดีต การพยากรณ์มีความสำคัญต่อการวางแผนและการตัดสินใจเกี่ยวกับการดำเนินงาของค์กรในสาขาต่าง ๆ เช่น ธุรกิจ อุตสาหกรรม เกษตร เป็นต้น ไม่ว่าองค์กรนั้นจะเป็นองค์กรขนาดเล็ก หรือขนาดใหญ่ องค์กรของรัฐหรือของเอกชน หากนักวางแผนหรือผู้ตัดสินใจในองค์กรทราบว่าเหตุการณ์หรือสภาพการณ์ใดจะเกิดขึ้นหรืออาจจะเกิดขึ้นในอนาคตด้วยความเชื่อมั่นในระดับหนึ่งจะทำให้การวางแผนหรือการตัดสินใจในการดำเนินงานเป็นไปอย่างถูกต้องหรือมีความผิดพลาดน้อย อย่างไรก็ตามการเกิดเหตุการณ์หรือสภาพการณ์หนึ่งเป็นการเกิดภายใต้ความไม่แน่นอน ดังนั้นการพยากรณ์ที่ให้ความถูกต้องสูงจึงมีความจำเป็นอย่างยิ่ง

วิธีของ Box และ Jenkins เป็นวิธีการพยากรณ์วิธีการหนึ่งที่ได้รับความนิยมมาก สำหรับข้อมูลอนุกรมเวลา โดยวิธีการนี้จะมี 4 ขั้นตอนหลัก ๆ คือ 1) กำหนดรูปแบบ (Identification) 2) การประมาณค่าพารามิเตอร์ในตัวแบบ (Estimation) 3) การตรวจสอบรูปแบบ (Diagnostic Checking) และ 4) การพยากรณ์ (Forecasting)

สำหรับปัญหาหนึ่งในการวิเคราะห์อนุกรมเวลา คือ ปัญหาข้อมูลนั้นมีค่าที่สูงหรือต่ำจนผิดปกติ (Outliers) ซึ่งปัญหานี้จะส่งผลกระทบต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

สำหรับผลงานวิจัยที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ในตัวแบบอนุกรมเวลาเมื่อมีค่าผิดปกติมีดังนี้

Denby และ Martin (1979) ได้ศึกษาเปรียบเทียบวิธีการประมาณแบบเอ็ม (M-Estimate)และวิธีประมาณแบบกำลังสองน้อยสุด ในอนุกรมเวลา ตัวแบบอัตสัมพันธ์ อันดับ 1 AR(1) พบว่าในกรณีที่เกิดค่าผิดปกตินั้น วิธีกำลังสองน้อยที่สุดจะมีประสิทธิภาพน้อยกว่าวิธีประมาณแบบเอ็ม

Chang , Tao และ Chen (1988) ได้นำตัวสถิติทดสอบของ Tsay (1986) มาประยุกต์ใช้กับวิธีการแก้ปัญหาค่าผิดปกติ โดยอาศัยหลักการของ Intervention Analysis และ การวนซ้ำ

ตรวจหา และปรับปรุงข้อมูล ซึ่งวิธีการนี้นั้นเป็นที่นิยมและมักจะถูกกล่าวถึงบ่อย ๆ ในการแก้ปัญหาค่าผิดปกติ

Chen และ Lui (1993) ได้คิดวิธีการประมาณร่วมของพารามิเตอร์ตัวแบบและผลกระทบของค่าผิดปกติ (Joint Estimation of Model Parameters and Outliers Effect) สำหรับแก้ปัญหาค่าผิดปกติ ซึ่งวิธีนี้พัฒนามาจากวิธีการของ Chang ซึ่งมีจุดเสียในกรณีที่มีค่าผิดปกติเกิดขึ้นหลายตัว โดยปรับปรุงขั้นตอนในการวนซ้ำและ ตรวจหา ซึ่งในการวนซ้ำแต่ละรอบนั้น จะมีการประมาณค่าพารามิเตอร์ตัวแบบ และผลกระทบของค่าผิดปกติไปพร้อม ๆ กัน

Agostinelli (2001) ได้คิดวิธีกำลังสองน้อยสุดแบบถ่วงน้ำหนักด้วยบูทสแทร็พ (Bootstrap Weighted Least Squares Method) ซึ่งค่าน้ำหนัก (weight) นั้นจะถูกคำนวณมาจาก nonparametric kernel density และ smooth model ซึ่งวิธีนี้นั้นนอกจากสามารถใช้สำหรับการประมาณค่าพารามิเตอร์ตัวแบบอนุกรมเวลาได้แล้ว ยังสามารถที่จะนำไปปรับใช้กับตัวแบบถดถอย หรือ ประมาณค่าเฉลี่ย และ ความแปรปรวนได้อีกด้วย

เฉลิม สิงห์สนอง (2536) ได้ทำวิทยานิพนธ์เรื่อง การเปรียบเทียบประสิทธิภาพของวิธีการตรวจหาแล้วปรับปรุงข้อมูล ด้วยวิธีการของ Tsay (Tsay 's Method) และวิธีการแบบวี (V-Method) โดยทำในกรณีที่มีค่าผิดปกติ ภายใต้ตัวแบบ AR(1) MA(1) และ ARMA(1,1) ได้ผลสรุปคือโดยรวมแล้ว วิธีการของ Tsay ให้ประสิทธิภาพดีกว่าวิธีการแบบวี

สุพร ฉัตรแก้วรัตนกุล (2538) ได้ทำวิทยานิพนธ์เรื่อง การประมาณค่าพารามิเตอร์เมื่อมีค่าผิดปกติในตัวแบบ AR(1)และ MA(1) โดยทำการเปรียบเทียบวิธีการประมาณค่า 3 วิธีคือ วิธีกำลังสองน้อยที่สุด (Least Squares Method) วิธีความควรจะเป็นสูงสุดเมื่อข้อมูลได้ปรับแก้แล้ว (Maximum Likelihood Method with Iterative Adjust Series) และ วิธีกำลังสองน้อยสุดเมื่อข้อมูลได้ปรับแก้แล้ว แล้ว (Least Squares Method with Iterative Adjust Series) ได้ผลโดยสรุปคือโดยรวมแล้ววิธีความควรจะเป็นสูงสุดเมื่อข้อมูลได้ปรับแก้แล้วให้ประสิทธิภาพดีที่สุด

ดังที่กล่าวมาแล้วว่า ข้อมูลที่ผิดปกตินั้น จะส่งผลกระทบต่อประสิทธิภาพของพารามิเตอร์ที่ประมาณได้ ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงมีความสนใจที่จะทำการเปรียบเทียบ วิธีการประมาณ 4 วิธีที่ยังไม่มีใครเปรียบเทียบกันมาก่อน ว่าวิธีใดจะมีประสิทธิภาพดีที่สุด โดยวิธีประมาณทั้ง 4 วิธีคือ

1.วิธีกำลังสองน้อยสุดแบบมีเงื่อนไข (Conditional Least Squares Method: CLS)

2.วิธีประมาณร่วมของพารามิเตอร์ตัวแบบและผลกระทบของค่าผิดปกติ (Joint Estimation of Model Parameters and Outliers Effect : JEMPOE)

3.วิธีประมาณแบบเอ็ม (M-Method : M)

4. วิธีกำลังสองน้อยสุดแบบถ่วงน้ำหนักด้วยบูทสแทร็พ (Bootstrap Weighted Least Squares Method: BWLS)

ข้อมูลอนุกรมเวลาที่ศึกษาในครั้งนี้เป็นอนุกรมเวลาที่มีค่าผิดปกติเกิดขึ้น ในตัวแบบ AR(1) MA(1) และ ARMA(1,1) จากนั้นทำการเปรียบเทียบประสิทธิภาพของวิธีประมาณทุกวิธีด้วย ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณพารามิเตอร์ (Average-Mean Square Error : AMSE)

1.2 วัตถุประสงค์ของการวิจัย

ในการวิจัยครั้งนี้มีวัตถุประสงค์ของการวิจัยดังนี้

1. เพื่อศึกษาและเปรียบเทียบ ประสิทธิภาพของวิธีประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลาเมื่อมีค่าผิดปกติเกิดขึ้น ด้วยวิธีประมาณ 4 วิธี คือ วิธีกำลังสองน้อยสุดแบบมีเงื่อนไข (วิธี CLS) วิธีประมาณร่วมของพารามิเตอร์ตัวแบบอิทธิพลของค่าผิดปกติ (วิธี JEMPOE) วิธีประมาณแบบเอ็ม (วิธี M) และวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักด้วยบูทสแทร็พ (วิธี BWLS)

2. เพื่อหาข้อสรุปเกี่ยวกับวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสมภายใต้สถานการณ์ต่าง ๆ ดังที่จะกล่าวต่อไปในขอบเขตของการวิจัย

1.3 สมมุติฐานของการวิจัย

สมมุติฐานของการวิจัยครั้งนี้คือ

วิธีกำลังสองน้อยสุดแบบถ่วงน้ำหนักด้วยบูทสแทร็พเป็นวิธีที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของพารามิเตอร์ ต่ำกว่า วิธีกำลังสองน้อยสุดแบบมีเงื่อนไข วิธีประมาณร่วมของพารามิเตอร์ตัวแบบและผลกระทบของค่าผิดปกติ และวิธีประมาณแบบเอ็ม

1.4 ข้อตกลงเบื้องต้น

ในการศึกษาครั้งนี้จะศึกษาการประมาณค่าพารามิเตอร์ตัวแบบอนุกรมเวลาเมื่อมีค่าผิดปกติเกิดขึ้น โดยจะทำการศึกษาตัวแบบอนุกรมเวลาทั้งสิ้น 3 ตัวแบบคือ

1) ตัวแบบอัตถดถอยอันดับที่ 1 AR(1)
2) ตัวแบบค่าเฉลี่ยเคลื่อนที่อันดับที่ 1 MA(1)
3) ตัวแบบอัตถดถอย-ค่าเฉลี่ยเคลื่อนที่อันดับที่ (1,1) ARMA(1,1)

1.5 ขอบเขตการวิจัย

ขอบเขตในวิจัยครั้งนี้คือ

1. อนุกรมเวลา z_t เป็นอนุกรมเวลาหนึ่งตัวแปร มีลักษณะคงที่ในค่าเฉลี่ยและคงที่ในความแปรแปรวน และมีค่าผิดปกติเกิดขึ้น

2. การกำหนดค่าพารามิเตอร์ในแต่ละตัวแบบที่จะศึกษา มีหลักการกำหนดให้เป็นไปตามคุณสมบัติของการเป็นกระบวนการสเตชันนารี (Stationary) และ อินเวอร์ทิเบิล (Invertible) โดยมีรายละเอียดในแต่ละตัวแบบดังนี้

2.1 ตัวแบบอัตสหสัมพันธ์อันดับที่หนึ่ง AR(1) มีตัวแบบคือ

$$z_t = \mu(1-\phi) + \phi z_{t-1} + a_t$$

กระบวนการนี้จะมีเงื่อนไข คือ $|\phi|<1$

กำหนดค่าพารามิเตอร์ ϕ ทั้งหมด 6 ค่า ดังนี้ -0.9,-0.5,-0.3 , 0.3 , 0.5 และ 0.9

2.2 ตัวแบบเฉลี่ยเคลื่อนที่อันดับที่หนึ่ง MA(1) มีตัวแบบคือ

$$z_t = \mu + a_t - \theta a_{t-1}$$

กระบวนการนี้จะมีเงื่อนไข คือ $|\theta|<1$

กำหนดค่าพารามิเตอร์ θ ทั้งหมด 6 ค่า ดังนี้ -0.9,-0.5,-0.3 , 0.3 ,0.5 และ 0.9

2.3 ตัวแบบ อัตตสัมพันธ์และค่าเฉลี่ยเคลื่อนที่ ARMA(1,1) มีตัวแบบคือ

$$z_t = \mu(1-\phi) + \phi z_{t-1} - \theta a_{t-1} + a_t$$

กระบวนการนี้จะมีเงื่อนไข คือ $|\phi|<1$ และ $|\theta|<1$

กำหนดค่าพารามิเตอร์ (ϕ,θ) ทั้งหมด 5 ชุด ดังนี้ (0.2,0.7),(0.8 ,0.3) ,(-0.5,0.4) (-0.4,-0.9) และ (0.7,-0.9)

3. กำหนดค่าเฉลี่ยอนุกรมเวลา μ เท่ากับ 100

4. กำหนดขนาดตัวอย่างเท่ากับ 40 , 50 , 60 , 80 , 100 และ 120 โดยในแต่ละสถานการณ์ที่ศึกษาจะทำซ้ำ 1,000 ครั้ง

5.ในการศึกษาครั้งนี้จะกระทำภายใต้เมื่อค่าความคลาดเคลื่อนสุ่ม (Random Error) a_t มีการแจกแจงแบบปกติปลอมปน (Contaminated Normal Distribution)

$$a_t \sim (1-P)N(0,\sigma_a^2) + PN(0,C^2\sigma_a^2)$$

โดย $N(\mu_a,\sigma_a^2)$ คือ การแจกแจงแบบปกติมีฟังก์ชันการแจกแจงเป็น

หอสมุดกลาง จุฬาลงกรณ์มหาวิทยาลัย

$$f(a)=\frac{1}{\sigma_a\sqrt{2\pi}}\exp(-\frac{1}{2\sigma_a^2}(a-\mu_a)^2)$$

โดยกำหนด $\mu_a=0$ และ σ_a^2 = 100

P คือ สัดส่วนการปลอมปน ที่มีค่าระหว่าง 0 ถึง 1 โดยกำหนดให้มีค่าเท่ากับ 0.05 0.10 , 0.15 และ 0.20

C คือ องค์ประกอบสเกล (Scale Factor) กำหนดให้มีค่าเท่ากับ 3 แทน ขนาดของค่าผิดปกติแบบไม่รุนแรง (Mild Outliers) และ 10 แทน ขนาดของค่าผิดปกติแบบรุนแรง (Extreme Outliers)

จากตัวแบบ a_t จะหมายถึง a_t มีการแจกแจงแบบ $N(0,\sigma_a^2)$ ในสัดส่วน $1-P$ และ มีการแจกแจงแบบ $N(0,C^2\sigma_a^2)$ ในสัดส่วน P

6. ทำการทดลองหาค่าประมาณ AMSE โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) ซึ่งเขียนด้วยโปรแกรมภาษาเดลไฟล์ (Delphi) จำลองซ้ำ 1,000 รอบในแต่ละสถานการณ์ของการทดลอง

1.6 เกณฑ์การประเมินประสิทธิภาพของวิธีประมาณ

ในการประเมินประสิทธิภาพของวิธีประมาณในการวิจัยครั้งนี้จะพิจารณา ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average-Mean Square Error : AMSE) ของตัวประมาณพารามิเตอร์ โดยที่ MSE นั้นจะเป็นการรวมระหว่าง ความเอนเอียง และ ความแปรปรวนของตัวประมาณ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณพารามิเตอร์ตัวที่ j คือ

$$MSE(\hat{\beta}_j)=\sum_{k=1}^{1000}\frac{(\hat{\beta}_{jk}-\beta_j)^2}{1,000}$$

โดยที่ β_j คือ ค่าจริงของพารามิเตอร์ เช่น μ , θ หรือ ϕ

$\hat{\beta}_{jk}$ คือ ค่าประมาณพารามิเตอร์ตัวที่ j ในการทำซ้ำรอบที่ k

k. คือ รอบที่ของการทำซ้ำ

จากนั้นพิจารณาจากค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ของวิธีประมาณ (AMSE)

$$AMSE = \frac{\sum_{j=1}^{p} MSE(\hat{\beta}_j)}{p}$$

โดยที่ $MSE(\hat{\beta}_j)$ คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ $\hat{\beta}_j$

p คือ จำนวนพารามิเตอร์ทั้งหมด

ในการพิจารณาประสิทธิภาพของวิธีประมาณนั้นจะพิจารณาว่าวิธีประมาณที่ให้ค่า AMSE ต่ำกว่าจะมีประสิทธิภาพมากกว่า

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัยครั้งนี้คือ

1. ทราบประสิทธิภาพของตัวประมาณ ในสถานการณ์ต่าง ๆ

2. เป็นแนวทางให้ผู้วิจัยสามารถเลือกใช้วิธีประมาณค่าพารามิเตอร์ของข้อมูลอนุกรมเวลาเมื่อมีค่าผิดปกติ ได้อย่างเหมาะสมและมีประสิทธิภาพ

3. เป็นแนวทางในการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของข้อมูลอนุกรมเวลา เมื่ออนุกรมเวลามีผลกระทบจากปัจจัยอื่น ๆ