บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันได้มีการนำความรู้ทางสถิติไปประยุกต์ใช้กับงานต่าง ๆ เป็นอันมากโดยเฉพาะงานวิจัยในสาขาบริหารธุรกิจ สังคมศาสตร์และเศรษฐศาสตร์ ทั้งนี้เนื่องมาจากวิธีการทางสถิติเป็นวิธีดำเนินการที่เป็นระบบ ซึ่งสามารถช่วยในการวิเคราะห์เพื่อหาคำตอบสำหรับการวิจัยนั้น ๆ ได้โดยเฉพาะอย่างยิ่งการหาคำตอบเพื่อคาดคะเนเหตุการณ์ล่วงหน้าหรือการพยากรณ์ เป็นวิธีการหนึ่งที่ผู้วิจัยเลือกใช้ในปัจจุบัน

เหตุผลที่การวิเคราะห์ข้อมูลอนุกรมเวลาเป็นที่นิยมใช้ คือ ลักษณะข้อมูลที่มีความถูกต้องแม่นยำ ผลการวิเคราะห์มีประสิทธิภาพและน่าเชื่อถือ ในทางปฏิบัติการเก็บรวบรวมข้อมูลเพื่อใช้ในการวิเคราะห์บางครั้งข้อมูลที่ได้มาไม่เป็นไปตามสภาวะการณ์ที่ศึกษาหรือควบคุมอยู่ ทำให้ข้อมูลบางค่าแตกต่างไปจากข้อมูลอื่น บางค่ามีค่าสูงมากบางค่ามีค่าต่ำมาก ผลของเหตุการณ์เหล่านี้ จะก่อให้เกิดค่าสังเกตที่ไม่เสถียรภาพมั่นคงของข้อมูลทั้งหมดที่มีอยู่ ค่าสังเกตเหล่านี้เรียกว่าค่าสังเกตที่ผิดปกติ* (Outliers) ความแตกต่างที่เกิดขึ้นมีสาเหตุสำคัญสองประการ (Ascomebe F.J. : 1960) ประการแรก คือ ความคลาดเคลื่อนที่เกิดจากการวัด (measurment error) หรือ ความคลาดเคลื่อนที่เกิดจากการปฏิบัติการ (execution error) เกิดขึ้นจากการใช้เครื่องมือเครื่องใช้ในการวัดที่มีคุณภาพต่ำหรือการจดบันทึกข้อมูลที่ผิดพลาด เช่น การลงรหัส

---

* ค่าสังเกตที่ผิดปกติ หมายถึง Residual ที่มีขนาด (Magnitude) สูงกว่า Residual อื่น ๆ 3σ หรือ 4σ หมายความว่า เมื่อนำค่า Residual a_t มาหาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน พร้อมทั้งสร้างช่วงเชื่อมั่น 99% และ Residual ตัวใดตำอยู่นอกช่วงเรียกว่า ค่าสังเกตที่ผิดปกติ

การเจาะบัตร เป็นต้น ประการที่สอง เป็นความคลาดเคลื่อนที่เกิดขึ้นโดยธรรมชาติ (inherent errror) เป็นความคลาดเคลื่อนที่เกิดจากผลกระทบของสิ่งแวดล้อมภาพนอกที่ไม่สามารถหลีกเลี่ยงได้แม้ว่าจะมีการควบคุมการวัดการปฏิบัติการอย่างดี เช่น ดัชนีราคาหุ้นตกอย่างผิดปกติเมื่อมีการเปลี่ยนแปลงทางการเมือง ราคาสินค้าอุปโภคบริโภคสูงขึ้นอย่างผิดปกติเมื่อเกิดภาวะสงคราม เป็นต้น ค่าสังเกตที่ผิดปกติเป็นสิ่งที่ทำให้เกิดความเสียหายในการวิเคราะห์ข้อมูลด้านความถูกต้องแม่นยำ ทำให้สรุปผลอ้างอิงที่ขาดความน่าเชื่อถือหรือบางครั้งไม่สามารถนำผลที่วิเคราะห์ไปใช้ได้

การวิเคราะห์ข้อมูลอนุกรมเวลาสามารถเขียนให้อยู่ในรูปตัวแบบ

$$\Phi(B)Z_t = \Theta(B)a_t \qquad (1.1)$$

เมื่อ $\Phi(B)$ เป็นฟังก์ชันพารามิเตอร์ของขบวนการตัวแบบ AR(p)

$\Theta(B)$ เป็นฟังก์ชันพารามิเตอร์ของขบวนการตัวแบบ MA(q)

B เป็นตัวดำเนินการย้อนเวลา (back-shiff operator)

ซึ่ง $\Phi(B) = (1 - \Phi_1 B - \ldots - \Phi_p B^p)$

$\Theta(B) = (1 - \Theta_1 B - \ldots - \Theta_q B^q)$

$\{a_t\}$ เป็นความคลาดเคลื่อน

ในกรณีที่ข้อมูลที่กำลังศึกษาเกิดมีค่าสังเกตที่ผิดปกติ เราสามารถทำการพยากรณ์โดยใช้เทคนิคการพยากรณ์ตามปกติ แต่จะส่งผลกระทบถึงผลการวิเคราะห์ในการประมาณค่าพารามิเตอร์ของตัวแบบสูงหรือต่ำกว่าความเป็นจริง ทำให้เกิดความเอนเอียง(bias) อาจให้ค่าพยากรณ์ที่มีความคลาดเคลื่อน (error) สูง มีความถูกต้องแม่นยำ (precision) ต่ำ และขาดความน่าเชื่อถือ (reliability) ซึ่งอาจนำไปสู่การตัดสินใจ หรือ การวางแผนงานผิดพลาดได้

ดังนั้น จึงควรจะมีวิธีการที่ใช้ตรวจหาและปรับแก้ข้อมูลที่เกิดค่าสังเกตที่ผิดปกติ การตรวจหาค่าสังเกตที่ผิดปกติในอนุกรมเวลาได้รับการศึกษาเป็นครั้งแรกโดย Fox (1972) ซึ่งได้

ศึกษาตัวแบบทางสถิติเกี่ยวกับการปรับแต่งค่าสังที่ผิดปกติ * (Additive Outlier : AO) และตัวแบบทางสถิติเกี่ยวกับชี้แนะค่าสังเกตที่ผิดปกติ** (Innovation Outlier :IO) ต่อมาได้ทำการศึกษาเพิ่มเติมโดย Abraham และ Box(1979), Martin (1980), Chang และ Tiao (1983), Hillmer bell และ Tiao(1983),Cheng Tiao และ Chen (1988) วิธีการดังกล่าวมีแนวคิดว่า ค่าสังเกตที่ผิดปกติมีอิทธิพลต่อค่าความคลาดเคลื่อนการพยากรณ์ จึงพิจารณาค่าความคลาดเคลื่อนเป็นตัวทดสอบค่าสังเกตที่ผิดปกติ

ดังนั้น ในการศึกษาวิจัยครั้งนี้จะทำการเปรียบเทียบสัดส่วนของความผิดพลาดทั้งหมด อำนาจการทดสอบ ค่าร้อยละเฉลี่ยความแตกต่างระหว่างค่าสังเกตที่ผิดปกติ ค่าสังเกตปกติและค่าสังเกตที่ผิดปกติเมื่อมีการปรับแก้ไขแล้ว ณ ตำแหน่งคาบเวลาที่ตรวจสอบพบตรงตำแหน่ง และค่าร้อยละเฉลี่ยสัมบูรณ์ของความถูกต้องการพยากรณ์ล่วงหน้าของอนุกรมเวลาคงที่ ที่มีค่าสังเกตที่ผิดปกติเฉพาะในตัวแบบการปรับแก้ค่าสังเกตที่ผิดปกติเท่านั้น โดยศึกษาภายใต้เทคนิคการจำลองแบบมอนติคาร์โล (MonteCarlo Simulation Technique) ภายใต้ขนาดตัวอย่าง ลักษณะการแจกแจงความคลาดเคลื่อน ค่าเฉลี่ย และความแปรปรวนตามที่กำหนด เพื่อจะทำให้เข้าใจและรู้โครงสร้างของข้อมูลได้ดียิ่งขึ้นเพื่อเป็นแนวทางการตัดสินใจได้อย่างมีประสิทธิภาพ

---

* การปรับแต่งค่าสังเกตที่ผิดปกติ หมายถึง การเพิ่มเติม ปรับแก้ข้อมูลอนุกรมเวลา ณ ตำแหน่งคาบเวลาที่มีค่าสังเกตที่ผิดปกติเพื่อให้ข้อมูลอนุกรมเวลา ณ คาบเวลานั้นมีความเหมาะสมมากยิ่งขึ้น

** ชี้แนะค่าสังเกตที่ผิดปกติ หมายถึง ค่าสังเกตของอนุกรมเวลา ณ คาบเวลาที่ชี้ให้ก่อนจะเกิดภาวะค่าสังเกตที่ผิดปกติในอนุกรมเวลานั้น

1.2 วัตถุประสงค์การวิจัย

1.2.1 เปรียบเทียบร้อยละสัดส่วนของความผิดพลาดการตรวจสอบค่าสังเกตที่ผิดปกติทั้ง 2 วิธี โดยวัดสัดส่วนความผิดพลาดประเภทที่ 1

1.2.2 เปรียบเทียบอำนาจการทดสอบของวิธีการทั้ง 2 เมื่อมีค่าสังเกตที่ผิดปกติเป็น 1 และ 2 ค่า

1.2.3 เปรียบเทียบค่าร้อยละเฉลี่ยความแตกต่างระหว่างค่าสังเกตที่ปกติกับค่าสังเกตที่ผิดปกติเมื่อมีการปรับแก้ไขแล้ว ณ ตำแหน่งคาบเวลาที่ตรวจพบ

1.2.4 เปรียบเทียบความถูกต้องการพยากรณ์ล่วงหน้าอนุกรมเวลาคงที่ ในรูปค่าร้อยละเฉลี่ยสัมบูรณ์

1.3 สมมติฐานสำหรับการวิจัย

การพยากรณ์อนุกรมเวลาที่มีการตรวจหาปรับแก้ค่าสังเกตที่ผิดปกติก่อนทำการพยากรณ์จะทำให้เกิดความถูกต้องแม่นยำสูงกว่าการไม่มีการปรับแก้ค่าสังเกตนั้น

1.4 ข้อตกลงเบื้องต้น

เพื่อหาข้อสรุปที่เหมาะสมในการเลือกใช้ตัวสถิติสำหรับตรวจสอบค่าสังเกตที่มีอิทธิพลและปรับแก้ค่าสังเกตที่ผิดปกติในข้อมูลอนุกรมเวลาด้วยวิธีการทำซ้ำ (Iteration) 2 วิธี คือ

1. วิธีการแบบวี (Method V)
2. วิธีการแบบเอ็ม (Method M)

1.4.1 ความคลาดเคลื่อน (a_t) เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติเหมือนกันและเป็นอิสระจากกัน โดยมีค่าเฉลี่ยเป็น 0 ความแปรปรวนเป็น 1 ($a_t \sim N(0,1)$)

1.4.2 การวิจัยครั้งนี้ถือว่า การสร้างอนุกรมเวลาคงที่ภายใต้ลักษณะการแจกแจงของค่าความคลาดเคลื่อนแบบปกติปลอมปน

1.4.3 ในการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลาคงที่จะใช้วิธีการฟังก์ชันสหสัมพันธ์ในตัวเอง

1.5 ขอบเขตของการวิจัย

1.5.1 สร้างอนุกรมเวลา โดยมีรูปแบบดังนี้

1.5.1.1 อนุกรมเวลาคงที่และมีตัวแบบอัตดสัมพันธ์อันดับที่หนึ่ง AR(1) (first order autoregressive model : AR(1))

$$Z_t = \mu + \Phi_1 Z_{t-1} + a_t \quad ; t = 1,2,\ldots,n \qquad (1.2)$$

โดยที่ n เป็นขนาดตัวอย่าง

Z_t เป็นอนุกรมเวลาที่ต้องการ

a_t เป็นความคลาดเคลื่อนสุ่ม

μ เป็นค่าคงที่ของอนุกรมเวลา

โดยที่ $|\Phi_1| < 1$

$$E(Z_t) = \mu/(1 - \Phi_1)$$

$$V(Z_t) = \sigma_a^2/(1 - \Phi_1^2) \quad ; t = 1,2, \ldots , n$$

$$E(Z_i Z_j) \neq 0 \ , \ i \neq j$$

$$E(Z_t Z_{t-1}) = \Phi_1 V(Z_t)$$

1.5.1.2 อนุกรมเวลาคงที่มีตัวแบบเฉลี่ยเคลื่อนที่อันดับที่หนึ่ง MA(1) (first order moving-average model : MA(1))

$$Z_t = \beta - \Theta_1 a_{t-1} + a_t \quad ; t = 1,2, \ldots , n \qquad (1.3)$$

โดยที่ Z_t เป็นอนุกรมเวลาที่ต้องการ

a_t เป็นความคลาดเคลื่อนสุ่ม

β เป็นค่าเฉลี่ยคงที่ของอนุกรมเวลา

n เป็นขนาดตัวอย่าง

$|\theta_1| < 1$

$E(Z_t) = \beta$

$V(Z_t) = (\theta_1^2 + 1)\sigma_a^2 \quad ; t = 1,2, \ldots , n$

$E(Z_t Z_{t-1}) = -\theta_1 \sigma_a^2$

1.5.1.3 อนุกรมเวลาคงที่มีตัวแบบผสมระหว่าง อัตตสัมพันธ์และเฉลี่ยเคลื่อนที่อันดับที่หนึ่ง ARMA(1,1) (mixed autoregressive-moving average first order model : ARMA(1))

$$Z_t = \mu + \phi_1 Z_{t-1} - \theta_1 a_{t-1} + a_t \quad ; t = 1,2, \ldots , n \qquad (1.4)$$

โดยที่ Z_t เป็นอนุกรมเวลาที่ต้องการ

a_t เป็นความคลาดเคลื่อนสุ่ม

μ เป็นค่าคงที่ของอนุกรมเวลา

n เป็นขนาดตัวอย่าง

$|\phi_1| < 1 \quad ; \quad |\theta_1| < 1$

$E(Z_t) = \mu / (1 - \phi_1)$

$E(Z_t Z_{t-1}) = (1 - \phi_1 \theta_1)(\phi - \theta_1)\sigma_a^2 / (1 - \phi_1^2)$

$V(Z_t) = (1 + \theta_1^2 - 2\phi_1 \theta_1)\sigma_a^2$

1.5.2 เมื่อค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน (Scale - contaminate normal distibution) ในการสร้างค่าสังเกตที่ผิดปกติ จะสร้างเป็น 2 ลักษณะ คือ

1.5.2.1 เมื่อเปรียบเทียบค่าร้อยละสัดส่วนของความผิดพลาดประเภทที่ 1 อำนาจการทดสอบและค่าร้อยละความแตกต่างระหว่างค่าสังเกตที่ผิดปกติกับค่าสังเกตที่ปกติ เมื่อมีการปรับแก้ค่าสังเกตที่ผิดปกติแล้ว จะสร้างค่าสังเกตที่ผิดปกติจากลักษณะการแจกแจงของความคลาดเคลื่อนแบบปกติปลอมปน คือ สเกลคอนทามิเนต มีรูปแบบการแจงแจงเป็น

$$F = (N-N_1)\, N(0,1) + N_1\, N(0,C^2(1)) \qquad (1.5)$$

เมื่อ N เป็นจำนวนขนาดตัวอย่าง

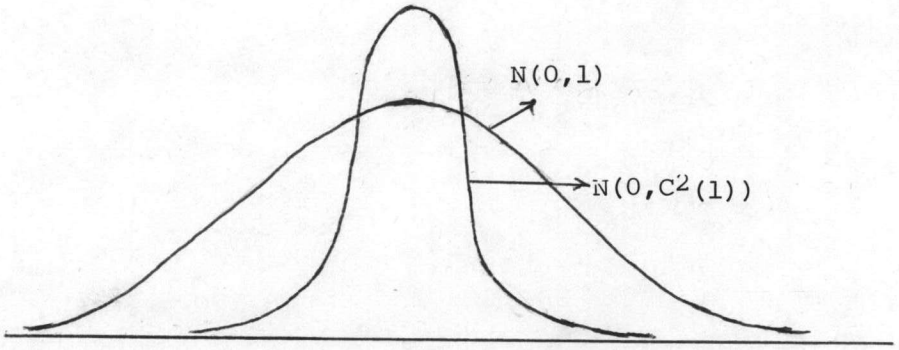
N_1 เป็นจำนวนค่าสังเกตที่ผิดปกติ

C เป็นค่าสเกลแฟคเตอร์ เราจะใช้ C=3 C=4 C=5 และ C=6

1.5.2.2 เมื่อเปรียบเทียบค่าร้อยละเฉลี่ยสัมบูรณ์ของความคลาดเคลื่อนการพยากรณ์ล่วงหน้าในอนุกรมเวลาคงที่ จะสร้างค่าสังเกตที่ผิดปกติจากลักษณะการแจกแจงของความคลาดเคลื่อนแบปกติปลอมปน คือ สเกลคอนทามิเนต มีรูปแบบการแจกแจงเป็น

$$F = (1-p)(N)\, N(0,1) + p(N)\, N(0,C^2(1)) \qquad (1.6)$$

เมื่อ p เป็นร้อยละของการปลอมปน (percentage of contamination) เราจะใช้ค่า $p = 5$, $p = 15$, $p = 25$



รูปที่ 1.1 แสดงลักษณะการแจกแจงแบบปกติปลอมปน

1.5.3 ศึกษาเมื่อค่าอัตตสัมพันธ์ในตัวเอง (ρ_1) ของตัวแบบอัตตสัมพันธ์อันดับที่หนึ่งเป็น 0.1 0.5 และ 0.9

1.5.4 ศึกษาเมื่อค่าอัตตสัมพันธ์ในตัวเอง (ρ_1) ของตัวแบบเฉลี่ยเคลื่อนที่อันดับที่หนึ่งเป็น 0.1 และ 0.4

1.5.5 ศึกษาเมื่อค่าอัตตสัมพันธ์ในตัวเอง (ρ_1) ของตัวแบบผสมระหว่างอัตตสัมพันธ์อันดับที่หนึ่งและเฉลี่ยเคลื่อนที่อันดับที่หนึ่งเป็น 0.1 0.5 และ 0.9

1.5.6 ศึกษาเมื่อขนาดตัวอย่าง (n) เท่ากับ 50 80 100 และ 120 ในแต่ละขนาดตัวอย่างพารามิเตอร์ทำซ้ำ 100 ครั้ง

1.5.7 ตรวจสอบค่าสังเกตที่ผิดปกติจากอนุกรมเวลาคงที่ที่สร้างขึ้น โดยวิธีการแบบ วี และวิธีการแบบ เอ็ม เสนอโดย Hsu (1977) Chang(1982) และ Hillmer(1983) Cheng Tiao และ Chen (1988) ทำการตรวจหาและปรับแก้ค่าสังเกตที่ผิดปกติที่เกิดขึ้นโดยใช้ระดับนัยสำคัญเป็น 0.05 และ 0.01

1.5.8 การวิจัยครั้งนี้ได้จำลองข้อมูลให้มีสภาพการณ์ตามที่ต้องการศึกษาโดยใช้เทคนิคการจำลองแบบมอนติ คาร์โล (Monte Carlo Simulation Technique) จากการเครื่องไมโครคอมพิวเตอร์ เขียนด้วยภาษาฟอร์แทรน 77 (FORTRAN 77) และโปรแกรมสำเร็จรูป SAS/ETS (Statistical Analysis Econmetrics and Time Series Library) และ AUTOBOX 3.0

1.6 ประโยชน์ของการวิจัย

1.6.1 ทำให้ผู้วิจัยเข้าใจโครงสร้างของปัญหา และรู้ข้อบกพร่องของข้อมูล เพื่อเป็นแนวทางในการแก้ปัญหาและหาทางป้องกันมิให้เกิดความผิดพลาดในการทดลองและการเก็บรวบรวมข้อมูล

1.6.2 เพื่อเป็นแนวทางในการศึกษาความคลาดเคลื่อน เพื่อความถูกต้องแม่นยำสำหรับการพยากรณ์

1.6.3 เพื่อให้ผู้ใช้สามารถใช้วิธีการตรวจสอบอนุกรมเวลา ในกรณีที่มีค่าสังเกตที่ผิดปกติ