บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

การบริหารงานด้านต่าง ๆในปัจจุบันนี้ได้มีการนำเอาสถิติเข้ามาช่วยในการตัดสินใจมากขึ้น เนื่องจากสถิติสามารถชี้ให้เห็นแนวโน้มของกิจการตลอดจนสามารถช่วยกำหนดการวางแผนการบริหาร การศึกษาข้อมูลในอดีตโดยอาศัยเทคนิคทางสถิติเพื่อการพยากรณ์ก่อให้เกิดผลดีต่อกิจการนั้น ๆ เพราะกิจการนั้นสามารถที่จะคาดการณ์ล่วงหน้าได้อย่างถูกต้องหรือใกล้เคียงความจริงมากที่สุด ซึ่งการวิเคราะห์อนุกรมเวลา (Time Series Analysis) เป็นวิธีการทางสถิติวิธีหนึ่งที่สามารถพยากรณ์หรือคาดคะเนค่าในอนาคตและแสดงให้เห็นถึงการเปลี่ยนแปลงของสิ่งที่สนใจในแต่ละช่วงเวลาที่เปลี่ยนไป โดยมีวัตถุประสงค์เพื่อประโยชน์ในการกำหนดนโยบายและหาลู่ทางแก้ปัญหาต่าง ๆ และยังใช้เพื่อเปรียบเทียบแนวโน้มของเรื่องต่าง ๆ กับคู่แข่งได้อีกด้วย ซึ่งมีความจำเป็นอย่างยิ่งในการวางแผนการดำเนินงานต่าง ๆ โดยเฉพาะทางด้านธุรกิจและเศรษฐศาสตร์ เช่น การวางนโยบายผลิตภัณฑ์ การวางแผนการตลาด การวางแผนการจัดการบุคลากรและการวางแผนการจัดการสินค้าคงคลัง เป็นต้น

เนื่องจากการวิเคราะห์อนุกรมเวลานั้นต้องศึกษาถึงความเคลื่อนไหวของข้อมูลชุดหนึ่งตามระยะเวลาต่าง ๆ กัน ซึ่งเรียงลำดับโดยคำนึงถึงเวลาของการเกิดเป็นระยะเวลาที่เท่า ๆ กันอาจเป็นรายเดือน รายวัน เป็นต้น ถ้าค่าสังเกตกระทำในเวลาที่ต่อเนื่องกัน อนุกรมเวลานี้เรียกว่า อนุกรมเวลาต่อเนื่อง (Continuous time series) แต่ถ้าค่าสังเกตกระทำ ณ จุดเวลาที่ไม่ต่อเนื่องกัน เรียกอนุกรมนี้ว่า อนุกรมเวลาไม่ต่อเนื่อง (Discrete time series) ซึ่งค่าสังเกต ณ จุดเวลาต่าง ๆ มีระยะห่างเท่า ๆกันและเป็นอนุกรมเวลาที่พบมากในทางปฏิบัติ สำหรับเทคนิคการวิเคราะห์อนุกรมเวลามีอยู่หลายวิธีเช่น อนุกรมเวลาแบบคลาสสิค และอนุกรมเวลาแบบบ็อกซ์-เจนกินส์ เป็นต้น ซึ่งการวิเคราะห์อนุกรมเวลาแบบคลาสสิคจะต้องกำหนดรูปแบบของแนวโน้มหรือรูปแบบของฤดูกาลขึ้น ซึ่งมีความลำบากในการกำหนดรูปแบบโดยเฉพาะเมื่ออนุกรม

เวลาไม่มีรูปแบบของแนวโน้มหรือฤดูกาลที่เด่นชัด ส่วนการวิเคราะห์อนุกรมเวลาแบบบ็อกซ์-เจนกินส์นั้นไม่มีการกำหนดรูปแบบแน่ชัดขึ้นก่อนทำการวิเคราะห์ รูปแบบจะถูกกำหนดขึ้นมาเองในระหว่างทำการวิเคราะห์ จะเห็นได้ว่าอนุกรมเวลาจะต้องอาศัยข้อมูลหรือค่าสังเกตในอดีตที่ผ่านมา แต่ผู้วิเคราะห์มักประสบปัญหาเกี่ยวกับการเก็บรวบรวมข้อมูล เช่น ข้อมูลบางค่าสูญหายไป จึงมีผลกระทบทำให้ผลการวิเคราะห์และการพยากรณ์คลาดเคลื่อนหรืออาจไม่ถูกต้องซึ่งอาจจะนำไปสู่การวางแผนที่ผิดพลาดได้ ดังนั้นก่อนที่จะทำการวิเคราะห์ ผู้วิเคราะห์จึงควรตรวจสอบข้อมูลว่ามีข้อมูลสูญหายบ้างหรือไม่

โดยทั่วไปเรามักจะพบว่าวิธีการแก้ปัญหาการสูญหายของข้อมูลนั้น ในทางปฏิบัติส่วนมากจะแก้ปัญหาโดยการตัดข้อมูลที่สูญหายทิ้ง และนำข้อมูลช่วงที่สมบูรณ์เท่านั้นมาทำการวิเคราะห์จึงมีผลทำให้ข้อมูลมีปริมาณลดน้อยลงและไม่มีประสิทธิภาพเพียงพอ แต่ในบางครั้งข้อมูลในอดีตมีปริมาณไม่มากนัก ผู้วิเคราะห์จึงไม่สามารถที่จะตัดข้อมูลออกได้และบางครั้งข้อมูลที่สูญหายไปนั้นมีความสำคัญ เช่น การวางแผนทางการผลิต ซึ่งต้องการทราบถึงการเปลี่ยนแปลงของยอดขายของสินค้าในแต่ละเดือน ด้วยเหตุนี้จึงได้มีนักสถิติหลายท่านพยายามคิดค้นหาวิธีการประมาณค่าสูญหายของอนุกรมเวลาขึ้น ดังที่จะได้กล่าวต่อไป

ค.ศ.1976 Brubacher and Wilson ได้เสนอวิธีการประมาณค่าสูญหายโดยอนุกรมเวลาช่วงก่อนและหลังข้อมูลสูญหายมีจำนวนมากพอ ซึ่งวิธีการประมาณค่าสูญหายนี้ได้อาศัยเทคนิคของบ็อกซ์-เจนกินส์ และสามารถประมาณค่าพารามิเตอร์โดยใช้ข้อมูลช่วงก่อนข้อมูลสูญหายและทำการประมาณค่าพารามิเตอร์ซ้ำ (reestimation) จากข้อมูลอนุกรมเวลาช่วงหลังข้อมูลสูญหาย

ค.ศ.1980 Eivind Damsleth ได้เสนอวิธีการประมาณค่าสูญหายโดยการรวมเชิงเส้น (linear combination) ที่เหมาะสมระหว่างการพยากรณ์ล่วงหน้าที่ได้จากข้อมูลก่อนข้อมูลสูญหาย และการพยากรณ์ย้อนหลังจากข้อมูลหลังข้อมูลสูญหาย เรียกว่า between-forecast ด้วยหลักการความคลาดเคลื่อนโดยประมาณน้อยที่สุด (minimum approximation error) โดยค่าพารามิเตอร์ที่ใช้ในการพยากรณ์ล่วงหน้าและย้อนหลังเป็นค่าเดียวกัน วิธีการนี้ได้รับการปรับปรุงและหาค่าถ่วงน้ำหนักในรูปแบบบางรูปแบบโดย Abraham (1981)

นอกจากนี้ในปี ค.ศ.1980 Richard H.John ได้เสนอวิธีการประมาณค่าพารามิเตอร์ด้วยขั้นตอนของ Kalman Bucy filter และอาศัยค่าคาดหวังอย่างมีเงื่อนไข (conditional expectation) มาทำการประมาณค่าสูญหาย

ในปี ค.ศ.1984 Miller and Ferreiro นำเสนอวิธี PEM (Pseudo Estimation Maximization) มาทำการประมาณค่าสูญหาย แต่ได้กล่าวว่าวิธีการ PEM เป็นวิธีการประมาณค่าสูญหายที่ไม่ดีกว่าวิธีประมาณค่าสูญหายด้วยค่าคาดหวังอย่างมีเงื่อนไข[1]

ค.ศ.1986 R.Harvey and C.F.Pierse ได้นำขั้นตอนของ Kalman filter และ fixed-point smoothing (Anderson and Moore ค.ศ.1979) ซึ่งได้ปรับปรุงโดย Harvey and Pierse (1984) และ Ansley and Kohn (1985) มาช่วยในการประมาณค่าสูญหายให้มีประสิทธิภาพยิ่งขึ้น ซึ่งวิธีการนี้มีความเหมาะสมไม่ว่าข้อมูลสูญหายเป็นจำนวนมากหรือน้อย

ดังนั้นผู้วิจัยจึงสนใจศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในข้อมูลอนุกรมที่เสนอโดย Damsleth และ Abraham ซึ่งใช้เทคนิค Between-Forecast estimation กับวิธี Fixed-Point Smoothing ที่เสนอโดย Harvey and Pierse โดยข้อมูลอนุกรมเวลาที่ศึกษาในครั้งนี้เป็นข้อมูลอนุกรมเวลาไม่ต่อเนื่องที่มีคุณสมบัติไม่เปลี่ยนแปลงเมื่อจุดเริ่มต้นของเวลาเปลี่ยนไป ซึ่งอนุกรมเวลาที่มีคุณสมบัติดังกล่าวเป็นอนุกรมเวลาคงที่ (stationary) โดยใช้วิธีการวิเคราะห์อนุกรมเวลา บ็อกซ์-เจนกินส์ (Box-Jenkins)

1.2 วัตถุประสงค์ของการวิจัย

ในการวิจัยต้องการศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในข้อมูลอนุกรมเวลา 2 วิธีดังนี้

1. วิธี Fixed-Point Smoothing
2. วิธี Between-Forecast Estimation

---

[1] O.Ferreiro, Methhodologies for The Estimation of Missing Observation in Time Series (Statistics & Probability Letters. vol.5, No.1, January, 1987)

เพื่อหาข้อสรุปเกี่ยวกับความเหมาะสมของการประมาณค่าสูญหายดังกล่าว โดยพิจารณาที่ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error ; MSE)

1.3 สมมุติฐานของการวิจัย

วิธีการประมาณค่าสูญหายในข้อมูลอนุกรมเวลาจะให้ค่าประมาณที่แตกต่างกัน โดยวิธีการ Fixed-Point Smoothing จะให้ค่าประมาณที่ใกล้เคียงค่าจริงมากกว่า วิธี Between-Forecast Estimation

1.4 ข้อตกลงเบื้องต้น

อนุกรมเวลา (Z_t) ที่ศึกษาในครั้งนี้เป็นอนุกรมเวลาไม่ต่อเนื่องที่มีตัวแปร 1 ตัวแปรและเป็นอนุกรมเวลาคงที่ (Stationary) จาก Box-Jenkins (1976) สามารถเขียนในรูปแบบ Mixed Autoregressive-Moving Average ; ARMA (p,q) ได้ดังนี้

$$\Phi(B)Z_t = \Theta(B)a_t \qquad \text{----------} \ (1.1)$$

หรือ

$$Z_t = \psi(B)a_t \qquad \text{----------} \ (1.2)$$

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

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

$\phi_1, \_\_\_, \phi_p$: คือสัมประสิทธิ์ความถดถอย (Autoregressive parameter)

$\theta_1, \_\_\_, \theta_q$: คือสัมประสิทธิ์ค่าเฉลี่ยเคลื่อนที่ (Moving-average parameter)

B : คือ backward shift operation นั่นคือ $B^m Z_t = Z_{t-m}$

a_t : เรียกว่า random shock คือตัวแปรสุ่มอิสระและมีการแจกแจงแบบปกติ ค่าเฉลี่ย 0 ความแปรปรวนคงที่ σ_a^2

$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \_\_\_ \qquad ; \sum_{j=0}^{\infty} \psi_j =$ ค่าคงที่

นั่นคือ $\sum_{j=0}^{\infty} \psi_j$ จะลู่เข้าสู่ค่าคงที่ จึงจะทำให้อนุกรม (Z_t) มีคุณสมบัติคงที่ และ $\sum_{j=0}^{\infty} \psi_j$ จะลู่เข้าสู่ค่าคงที่เมื่อรากของสมการ $\Phi(B) = 0$ และสมการ $\Theta(B) = 0$ จะต้องอยู่นอกวงกลมรัศมี

1 หน่วย (unit circle) ซึ่งสามารถสรุปได้ในภาคผนวก ก.

ในทำนองเดียวกันกับสมการที่ (1.1) เราสามารถเขียนในรูปของสมการล่วงหน้า (foreward) ได้ดังนี้

$$\Phi(F)Z_t = \Theta(F)e_t \qquad ---------- (1.3)$$

โดย

F : คือ fordward shift operator นั่นคือ $F^m Z_t = Z_{t+m}$

e_t : คือ ตัวแปรสุ่มอิสระที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ย 0 ความแปรปรวนคงที่ $\sigma^2_e = \sigma^2_a$

Z_t : คือ ค่าสังเกตของอนุกรมเวลาหลังจากลบกับค่าเฉลี่ยแล้ว

1.5 ขอบเขตของการวิจัย

1. อนุกรมเวลา (Z_t) เป็นอนุกรมเวลาแบบ 1 ตัวแปร (Univariate Time Series) และมีคุณสมบัติเป็นอนุกรมเวลาคงที่ (Stationary) และเขียนในรูปแบบ AR(1),AR(2),MA(1),MA(2) และ ARMA(1,1)
2. a_t เป็นอิสระกันและมีการแจกแจงปกติ ค่าเฉลี่ย 0 และความแปรปรวนคงที่ $\sigma^2_a = 100$
3. อนุกรมเวลา (Z_t) มีข้อมูลจำนวน 50, 75 และ 100
4. กำหนดอัตตสหสัมพันธ์ (Autocorrelation) มีค่า สูง กลาง และต่ำ
5. ในการศึกษาครั้งนี้ข้อมูลมีการสูญหายอย่างสุ่ม และเป็นช่วง 1 ช่วง (single gap) โดยจำนวนข้อมูลสูญหายคือ 1 ค่า และ 2 ค่า
6. ตำแหน่งของข้อมูลสูญหายจะแบ่งเป็น 3 ช่วง คือ ต้น กลาง และปลาย
7. การวิจัยในครั้งนี้ได้จำลองข้อมูลขึ้นตามสถานการณ์ที่ต้องการศึกษา โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) และทำการจำลองแบบซ้ำๆกัน 50 รอบในแต่ละสถานการณ์ของการวิจัย

8. กำหนดให้พารามิเตอร์มีค่าดังนี้

รูปแบบ	พารามิเตอร์
AR(1)	ϕ_1 = 0.3,0.6,0.8
AR(2)	ϕ_1 = 0.3 ϕ_2 = 0.6, ϕ_1 = 0.4 ϕ_2 = -0.6
	ϕ_1 = 0.6 ϕ_2 = -0.8, ϕ_1 = 1.0 ϕ_2 = -0.6
	ϕ_1 = 1.0 ϕ_2 = -0.8, ϕ_1 = 1.5 ϕ_2 = -0.8
MA(1)	θ_1 = 0.3,0.8
MA(2)	θ_1 = 0.4 θ_2 = 0.3, θ_1 = 0.6 θ_2 = -0.3
	θ_1 = 0.2 θ_2 = -0.6, θ_1 = 0.6 θ_2 = -0.8
	θ_1 = 1.2 θ_2 = -0.8
ARMA(1,1)	ϕ_1 = 0.3 θ_1 = 0.8, ϕ_1 = 0.3 θ_1 = -0.6
	ϕ_1 = 0.8 θ_1 = -0.6

การกำหนดค่าพารามิเตอร์ดังตารางข้างต้นมีวัตถุประสงค์ให้อนุกรมเวลาเป็นอนุกรมเวลาคงที่ตามเงื่อนไขของแต่ละรูปแบบของอนุกรมเวลาและระดับความรุนแรงของอัตตสหสัมพันธ์

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการประมาณค่าสูญหายในข้อมูลอนุกรมเวลาที่เหมาะสมกับรูปแบบของข้อมูลและสถานการณ์ต่าง ๆ
2. เพื่อเป็นแนวทางในการศึกษา และเปรียบเทียบวิธีการประมาณค่าสูญหายในสถานการณ์อื่น ๆ อีกต่อไป