บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เพื่อการพยากรณ์ในข้อมูลอนุกรมเวลาเมื่อมีค่าผิดปกติ ซึ่งวิธีการทั้ง 3 คือ วิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว และ วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว โดยศึกษาภายใต้รูปแบบของอนุกรมเวลา 2 รูปแบบ ขนาดตัวอย่าง 4 ระดับ ความคลาดเคลื่อนมีการแจกแจงใน 2 รูปแบบ ระดับของการปลอมปน 4 ระดับ และระดับของพารามิเตอร์ 5 ระดับ สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบความสามารถในการพยากรณ์คือ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) เฉลี่ย 12 คาบเวลาพยากรณ์

วิธีการดำเนินการวิจัยครั้งนี้ ใช้วิธีการจำลองแบบการทดลองด้วยเทคนิคมอนติคาร์โล ด้วยเครื่องคอมพิวเตอร์ AMDAHL 5860 โดยใช้โปรแกรมภาษาฟอร์แทรน เพื่อสร้างข้อมูลให้มีลักษณะตามแผนการทดลองที่กำหนด และกำหนดให้เครื่องคอมพิวเตอร์ทำงานซ้ำๆกัน 500 รอบในแต่ละสถานการณ์

สรุปผลการวิเคราะห์

จากการทดลองเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี ได้ข้อสรุปดังนี้

1. รูปแบบอัตตสัมพันธ์อันดับที่หนึ่ง (AR(1))

1.1 เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน

วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้วมีความคลาดเคลื่อนในการพยากรณ์ต่ำสุด เมื่อขนาดตัวอย่างเท่ากับ 40 ในทุกระดับของสเกลแฟกเตอร์(5 และ 10) ทุกระดับของสัดส่วนการปลอมปน(0.03 , 0.05 , 0.08 และ 0.10) และทุกระดับของ สัมประสิทธิ์การถดถอย(0.3 , 0.5 , 0.6 , 0.7 และ 0.8)

วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว จะให้ความคลาดเคลื่อนในการพยากรณ์ต่ำสุด เมื่อขนาดตัวอย่างเท่ากับ 60 , 80 และ 120 ในทุกระดับของสเกลแฟกเตอร์(5 และ 10) ทุกระดับของสัดส่วนการปลอมปน(0.03 , 0.05 , 0.08 และ 0.10) และทุกระดับของสัมประสิทธิ์การถดถอย(0.3 , 0.5 , 0.6 , 0.7 และ 0.8)

1.2 เมื่อความคลาดเคลื่อนถูกปลอมปนด้วยการแจกแจงลาปลาซ

วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้วมีความคลาดเคลื่อนในการพยากรณ์ต่ำสุด ในทุกระดับของขนาดตัวอย่าง(40 , 60 , 80 และ 120) ทุกระดับของ β (1 และ 10) ทุกระดับของสัดส่วนการปลอมปน(0.03 , 0.05 , 0.08 และ 0.10) เมื่อสัมประสิทธิ์การถดถอยมีค่าค่อนข้างสูง(0.5 , 0.6 , 0.7 และ 0.8)

สำหรับที่ระดับของสัมประสิทธิ์การถดถอยเท่ากับ 0.3 ไม่มีวิธีการใดให้ค่าพยากรณ์ต่ำอย่างเด่นชัด ในทุกระดับของขนาดตัวอย่าง(40 , 60 , 80 และ 120) ทุกระดับของ β (1 และ 10) และทุกระดับของสัดส่วนการปลอมปน(0.03 , 0.05 , 0.08 และ 0.10)

2. รูปแบบค่าเฉลี่ยเคลื่อนที่อันดับที่หนึ่ง (MA(1))

2.1 เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน

วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้วมีความคลาดเคลื่อนในการพยากรณ์ต่ำสุดในทุกสถานการณ์

2.2 เมื่อความคลาดเคลื่อนถูกปลอมปนด้วยการแจกแจงลาปลาซ

วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้วมีความคลาดเคลื่อนในการพยากรณ์ต่ำสุด ในทุกสถานการณ์ ยกเว้นเมื่อตัวอย่างมีขนาดใหญ่ (n = 120) ในทุกระดับของ β (1 และ 10) ทุกระดับของสัดส่วนการปลอมปน(0.03 , 0.05 , 0.08 และ 0.10) และค่าสัมประสิทธิ์การเคลื่อนที่มีค่าค่อนข้างสูง(0.6 , 0.7 และ 0.8) วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว จะให้ความคลาดเคลื่อนในการพยากรณ์ต่ำสุด

3. ผลกระทบจากสัดส่วนของการปลอมปน และความรุนแรงของข้อมูลผิดปกติที่มีผลต่อการพยากรณ์

3.1 เมื่อกำหนดให้ปัจจัยอื่นๆ(ขนาดตัวอย่าง ลักษณะการแจกแจงของความคลาดเคลื่อน และความรุนแรงของข้อมูลผิดปกติ)คงที่ เมื่อข้อมูลผิดปกติมีจำนวนเพิ่มขึ้น(สัดส่วนของการปลอมปนเพิ่มขึ้น) จะมีผลให้ค่าความคลาดเคลื่อนจากการพยากรณ์มีค่าสูงขึ้น

3.2 เมื่อกำหนดให้ปัจจัยอื่นๆ(ขนาดตัวอย่าง ลักษณะการแจกแจงของความคลาดเคลื่อน และสัดส่วนของการปลอมปน)คงที่ เมื่อข้อมูลผิดปกติมีความรุนแรงมากขึ้น จะมีผลให้ค่าความคลาดเคลื่อนจากการพยากรณ์มีค่าสูงขึ้น

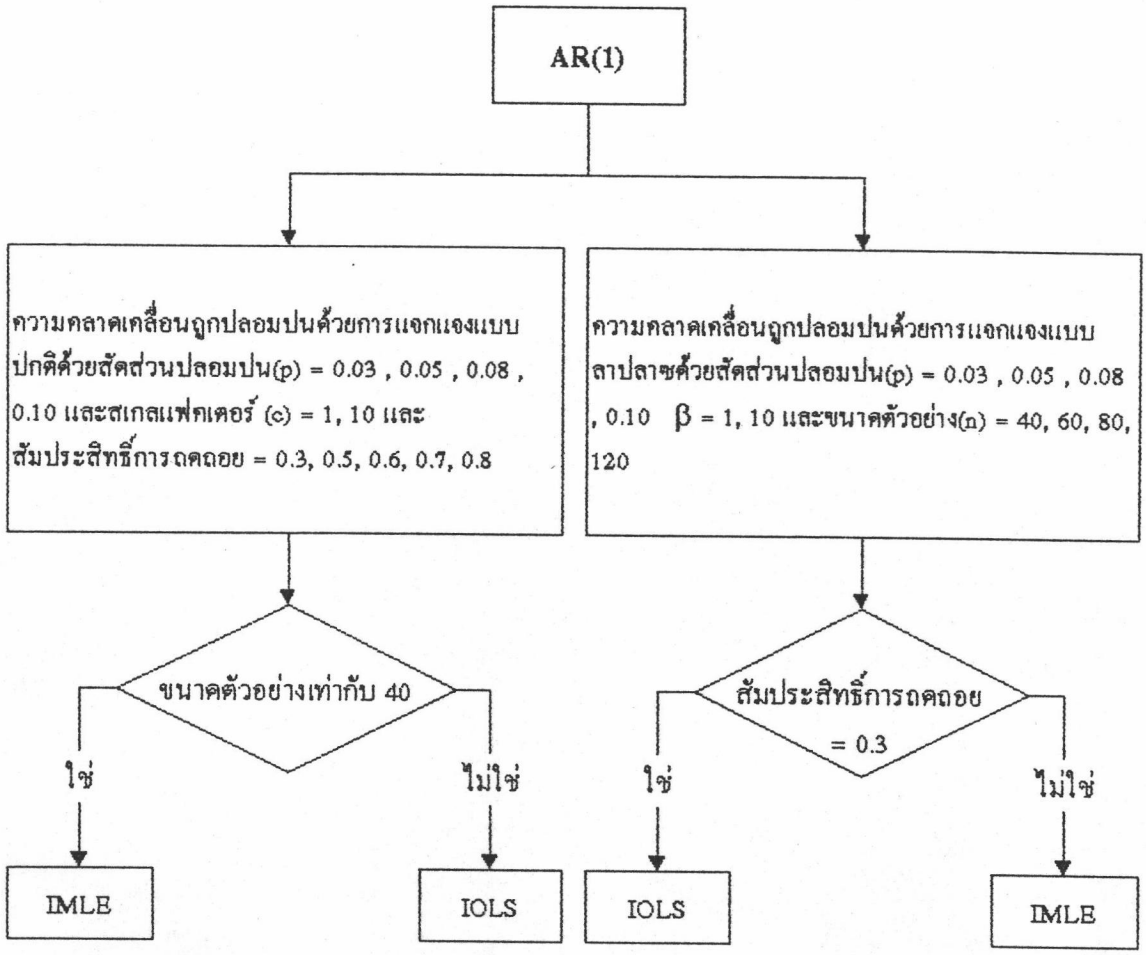
4. ผลจากตำแหน่งของข้อมูลผิดปกติที่มีต่อการพยากรณ์

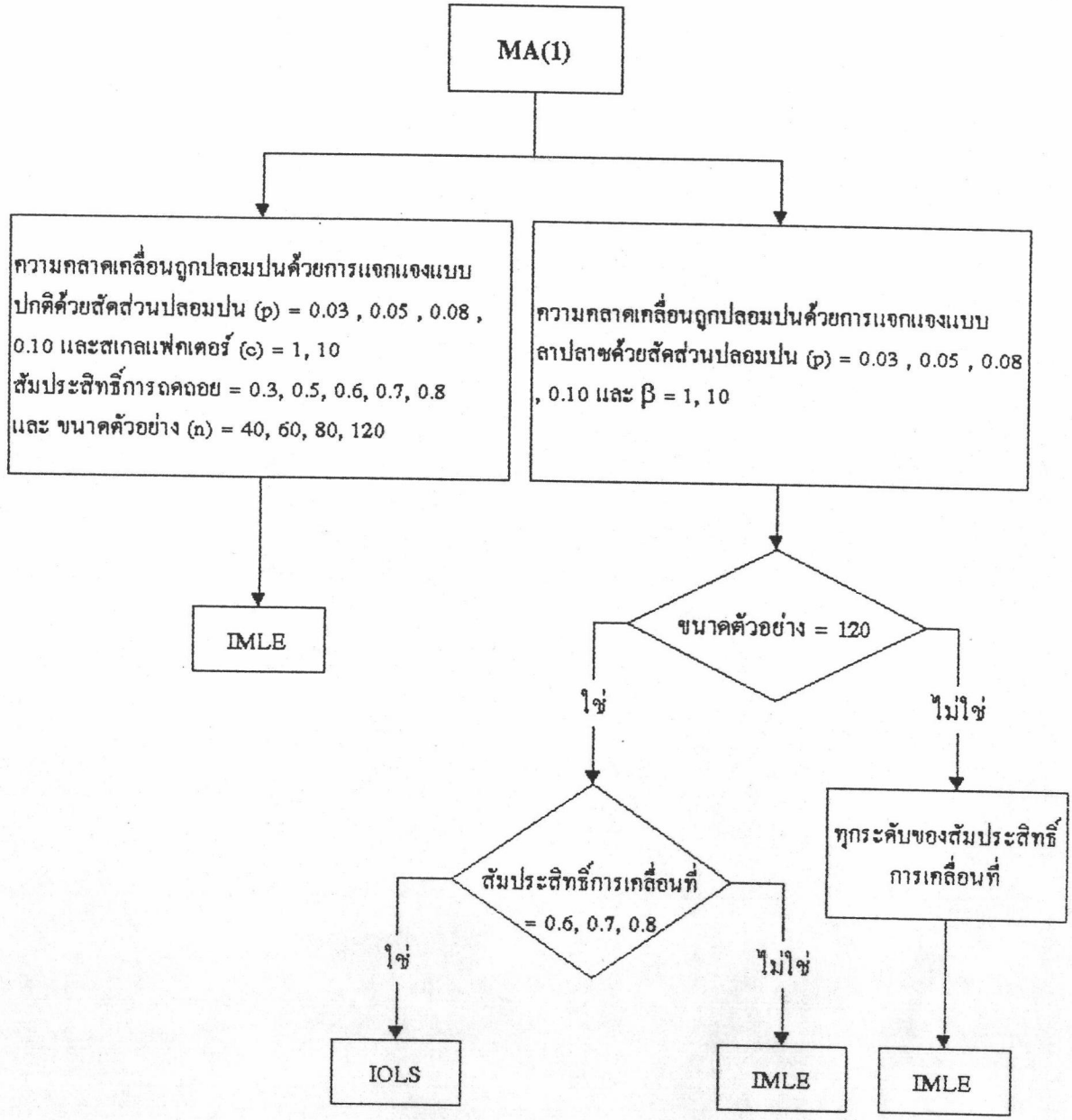
จากการศึกษาค่า RMSE เฉลี่ย 12 คาบเวลาของวิธีการพยากรณ์ทั้ง 3 วิธีในการวิเคราะห์ข้อมูลอนุกรมเวลารูปแบบอัตตสัมพันธ์อันดับที่หนึ่ง (AR(1)) โดยความคลาดเคลื่อนมีฟังก์ชันการแจกแจงในรูปของ $f(x) = (1-p)N(0,100) + pN(0,C^2\sigma^2)$ เมื่อสเกลแฟกเตอร์(C) มีขนาดเท่ากับ 10 จำแนกตามระดับพารามิเตอร์ (ϕ) ขนาดตัวอย่าง (n) และสัดส่วนของการปลอมปน (p) โดยกำหนดให้ข้อมูลผิดปกติเกิด ณ.ตำแหน่งต้นและท้ายอนุกรม (ดังตารางที่ ง.1 และ ง.2) เปรียบเทียบกับกรณีที่ข้อมูลผิดปกติเกิดโดยสุ่ม (ตารางที่ 4.1) ได้ข้อสรุปคือ

4.1 กรณีที่ข้อมูลผิดปกติเกิดณ.ตำแหน่งต้นและท้ายอนุกรม ค่า RMSE ของวิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูลจะมีค่าสูงกว่ากรณีที่ข้อมูลผิดปกติเกิดโดยสุ่ม ส่วนค่า RMSE ของวิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้วและวิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้วจะมีค่าไม่แตกต่างจากกรณีที่ข้อมูลผิดปกติเกิดโดยสุ่ม

4.2 กรณีที่ข้อมูลผิดปกติเกิดณ.ตำแหน่งต้นและท้ายอนุกรมวิธีการพยากรณ์ที่ให้ความคลาดเคลื่อนในการพยากรณ์ต่ำสุดเป็นวิธีเดียวกับกรณีที่ข้อมูลผิดปกติเกิดโดยสุ่ม

จากการสรุปผลข้างต้นสามารถเขียนเป็นแผนภาพได้ดังนี้





หมายเหตุ IOLS หมายถึง วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว
IMLE หมายถึง วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว

ข้อเสนอแนะ

สำหรับข้อเสนอแนะในการวิจัยครั้งนี้สามารถแยกได้เป็น 2 ด้านคือ

1. ด้านการนำไปใช้ประโยชน์

1.1 รูปแบบอัตสัมพันธ์อันดับที่หนึ่งในการหาสมการการพยากรณ์เมื่อข้อมูลมีค่าผิดปกติ ควรพิจารณาจากขนาดตัวอย่าง ในกรณีที่ตัวอย่างมีขนาดเล็ก (n = 40) ควรใช้วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้วในการหาสมการการพยากรณ์ เมื่อตัวอย่างมีขนาดมากขึ้นควรใช้วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้วในการหาสมการการพยากรณ์ เนื่องจากในทางปฏิบัติเราไม่ทราบรูปแบบของการแจกแจงที่มาปลอมปน และวิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้วสะดวกต่อการคำนวน

1.2 รูปแบบค่าเฉลี่ยเคลื่อนที่อันดับที่หนึ่ง ควรใช้วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้วในการหาสมการการพยากรณ์ แต่ถ้าตัวอย่างมีขนาดใหญ่ (n = 120) ควรใช้วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้วในการหาสมการการพยากรณ์

2. ด้านการศึกษาวิจัย

2.1 ควรศึกษาในกรณีที่ข้อมูลอนุกรมเวลามีผลกระทบจากปัจจัยอื่น เช่น อิทธิพลของฤดูกาล หรือศึกษาในกรณีที่อนุกรมเวลามีรูปแบบอื่น เช่น AR(2) , MA(2) หรือ ARMA(1,1)

2.2 สำหรับการประมาณค่าพารามิเตอร์ครั้งนี้ ศึกษาเฉพาะวิธีการประมาณค่าพารามิเตอร์ 3 วิธีดังกล่าวข้างต้น ยังมีวิธีการอื่นๆที่น่าสนใจ เช่น วิธีตัวประมาณเอ็ม (M-Estimators) วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดโดยใช้สมการภาวะน่าจะเป็นในรูปของการแจกแจงแบบปกติปลอมปน