บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ที่ได้จากวิธีการคัดเลือกตัวแบบการถดถอย ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ ซึ่งวิธีการคัดเลือกสมการถดถอยที่นำมาใช้ในการสร้างตัวแบบการถดถอยเชิงเส้นพหุคูณมี 5 วิธีดังนี้

1) วิธีการคัดเลือกตัวแบบที่ดีที่สุด โดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC))
2) วิธีการคัดเลือกตัวแปรของเบส์(Bayesian Variable Selection Method (BVS))
3) วิธีการเฉลี่ยตัวแบบของเบส์(Bayesian Model Averaging Method) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam ' s Window) (BMA_{OCC})
4) วิธีการเฉลี่ยตัวแบบของเบส์(Bayesian Model Averaging Method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC3})
5) วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

โดย 4 วิธีการแรกนั้นเป็นวิธีการภายใต้แนวทางของเบส์ ซึ่งมีการประยุกต์หลักการของเบส์นำไปใช้ในลักษณะต่าง ๆ เพื่อหาตัวแบบที่ให้ค่าพยากรณ์ที่เหมาะสมที่สุด ผู้วิจัยจึงได้มีการนำวิธีการภายใต้แนวทางของเบส์เหล่านี้มาเปรียบเทียบกับวิธีการพื้นฐานที่นิยมใช้กันอย่างแพร่หลายสำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ นั่นคือวิธีการถดถอยแบบขั้นบันได เพื่อศึกษาว่าวิธีการภายใต้แนวทางของเบส์จะมีประสิทธิภาพมากกว่าวิธีการพื้นฐานหรือไม่ นอกจากนี้การนำวิธีการเฉลี่ยตัวแบบของเบส์มาศึกษายังจะทำให้ทราบว่าวิธีการเฉลี่ยตัวแบบของเบส์ซึ่งมีการคำนึงถึงหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบจะมีประสิทธิภาพสูงกว่าวิธีการอื่น ๆ ที่ไม่ได้คำนึงถึงหลักการนี้หรือไม่ ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงได้ทำการเปรียบวิธีการคัดเลือกสมการถดถอยทั้ง 5 วิธีข้างต้น เพื่อศึกษาและเปรียบเทียบว่าวิธีการใดจะให้ค่าพยากรณ์ที่เหมาะสมที่สุด ซึ่งมีการกำหนดสถานการณ์ต่าง ๆ ในการวิจัยครั้งนี้ไว้ดังนี้

1) การแจกแจงของค่าคลาดเคลื่อนเป็นแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 10 20 และ 25
2) ขนาดตัวอย่างที่ศึกษาคือ 25 50 75 และ 100

3) ระดับนัยสำคัญ (α) คือ 0.01 และ 0.05
4) จำนวนตัวแปรอิสระที่ศึกษา คือ 3 5 10 12 และ 15 ตัวแปร
5) ค่าคงที่สำหรับวิธี BVS และวิธี BMA_{MC3} มี 4 ระดับ คือ (1,5) (1,10) (10,100) และ (10,500)
6) ค่าคงที่ที่ใช้เป็นเกณฑ์ในการค้นหาปริภูมิตัวแบบสำหรับวิธี BMA_{OCC} เป็น 20 ในกรณีที่ระดับนัยสำคัญเท่ากับ 0.05 และเป็น 100 ในกรณีที่ระดับนัยสำคัญเท่ากับ 0.01

โดยเกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการคัดเลือกตัวแบบการถดถอยวิธีใดจะมีความถูกต้องและแม่นยำมากที่สุดพิจารณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(Average of Mean Square Error (AMSE)) และเกณฑ์ที่ใช้ประกอบการตัดสินใจจะใช้เกณฑ์ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(Ratio of Different Average Mean Square Error (RDAMSE)) ซึ่งวิธีใดให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด จะเป็นวิธีที่ดีที่สุด ผลการวิจัยได้ข้อสรุปดังนี้

5.1 สรุปผลการวิจัย

5.1.1 ผลการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

จากการเปรียบเทียบค่า AMSE ของทั้ง 5 วิธี พบว่า ค่า AMSE ของแต่ละวิธีเรียงลำดับจากน้อยไปมาก ได้แก่ วิธี BMA_{MC3} BMA_{OCC} BVS BIC และ SR ตามลำดับ สำหรับทุกระดับขนาดตัวอย่าง ค่าส่วนเบี่ยงเบนมาตรฐาน จำนวนตัวแปรอิสระ ระดับนัยสำคัญ และค่าคงที่วิธี BMA_{MC3} และวิธี BVS แสดงว่าวิธี BMA_{MC3} ให้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากที่สุด ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ วิธี BMA_{OCC} นั้นมีความสามารถในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณใกล้เคียงกับวิธี BMA_{MC3} ซึ่งจะสังเกตได้จากค่า AMSE ของวิธี BMA_{OCC} จะสูงกว่าวิธี BMA_{MC3} เพียงเล็กน้อยเท่านั้น โดยเฉพาะอย่างยิ่งเมื่อระดับนัยสำคัญลดลง สำหรับวิธี BVS นั้นจะให้ค่า AMSE ใกล้เคียงกับวิธี BMA_{MC3} และวิธี BMA_{OCC} ก็ต่อเมื่อค่าคงที่ของวิธี BVS มีค่าต่ำ ๆ ส่วนวิธี BIC และ SR มีค่า AMSE แตกต่างจากวิธี BMA_{MC3} และวิธี BMA_{OCC} อย่างชัดเจนในทุกกรณี ซึ่งจากผลการวิจัยที่ได้นี้นอกจากจะสรุปได้ว่าวิธี BMA_{MC3} มีความสามารถในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณได้ดีที่สุดแล้ว ยังสามารถสรุปได้ว่าวิธีการภายใต้แนวทางของเบส์มีความสามารถในการ

คัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณดีกว่าวิธีการพื้นฐานซึ่งในที่นี้คือวิธีการถดถอยแบบขั้นบันได โดยจะเห็นได้จากวิธีการภายใต้แนวทางของเบส์จะให้ค่า AMSE ต่ำสุด 4 อันดับทุกกรณี ส่วนข้อสรุปเกี่ยวกับหลักการความไม่แน่นอนของตัวแบบ จะพบว่าวิธี BMA_{MC3} และวิธี BMA_{OCC} ซึ่งคำนึงถึงหลักการนี้จะให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ ในทุกกรณี แสดงว่าการคำนึงถึงหลักการความไม่แน่นอนของตัวแบบจะช่วยให้ได้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากยิ่งขึ้นในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ

5.1.2 ปัจจัยที่มีผลต่อค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละวิธี

1) ขนาดตัวอย่าง (n)

เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่า AMSE ของทุกวิธีมีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความเบี่ยงเบนที่ไม่ทราบสาเหตุลงได้ และเมื่อขนาดตัวอย่างเพิ่มขึ้นวิธี BMA_{MC3} และวิธี BMA_{OCC} จะมีค่า AMSE ต่ำกว่าวิธีอื่น ๆ อย่างชัดเจน

2) ระดับนัยสำคัญ (α)

ระดับนัยสำคัญจะมีผลต่อค่า AMSE ของวิธี BMA_{OCC} และวิธี SR เท่านั้น โดยค่า AMSE จะเพิ่มขึ้นเมื่อระดับนัยสำคัญเพิ่มขึ้น เนื่องจากเมื่อระดับนัยสำคัญเพิ่มขึ้นจะทำให้ปริภูมิตัวแบบของวิธี BMA_{OCC} มีตัวแบบที่ได้รับการยอมรับน้อยลง ส่งผลให้อาจมีการละเลยบางตัวแบบไป ทำให้การเฉลี่ยตัวแบบที่ได้อาจนำไปสู่การพยากรณ์ที่ไม่เหมาะสม สำหรับวิธี SR นั้น การเพิ่มระดับนัยสำคัญจะทำให้ตัวแปรอิสระมีโอกาสเข้าออกจากตัวแบบมากขึ้น ซึ่งการวิจัยครั้งนี้กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูปที่มีตัวแปรอิสระครบทุกตัว ดังนั้นเมื่อตัวแปรอิสระมีโอกาสเข้าออกจากตัวแบบมากขึ้นก็จะทำให้การพยากรณ์มีความคลาดเคลื่อนมากยิ่งขึ้น ส่วนวิธีอื่น ๆ นั้นไม่ได้นำระดับนัยสำคัญมาพิจารณา ค่า AMSE จึงไม่เปลี่ยนแปลง

3) ค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่ม (σ)

เมื่อค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสุ่มมีค่าเพิ่มขึ้น ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นตัวประมาณค่าความแปรปรวนของค่าคลาดเคลื่อน (σ^2)

4) จำนวนตัวแปรอิสระ

เมื่อจำนวนตัวแปรอิสระในตัวแบบการถดถอยเพิ่มขึ้น จะส่งผลให้ค่า AMSE ของทุกวิธีมีแนวโน้มสูงขึ้น เนื่องจากในการวิจัยครั้งนี้ได้กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูป เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้นโอกาสที่จะได้ตัวแบบที่ไม่เหมาะสมก็จะมีมากขึ้น ทำให้ค่า AMSE ของทุกวิธีมีแนวโน้มสูงขึ้น โดยเฉพาะอย่างยิ่งวิธี SR ซึ่งใช้หลักการในการคัดเลือกตัวแปรเข้าออกจากตัวแบบผ่านการทดสอบสมมติฐานเท่านั้น ในขณะที่วิธีการภายใต้แนวทางของเบส์ทั้ง 4 วิธี คือ BIC BVS BMA_{OCC} BMA_{MC3} ค่า AMSE จะมีอัตราการเพิ่มขึ้นน้อยกว่า วิธี SR เพราะวิธีการภายใต้แนวทางของเบส์จะมีการพิจารณาถึงความเหมาะสมของตัวแบบด้วย เช่น มีการคำนึงถึงความน่าจะเป็นภายหลังของตัวแบบ โดยเฉพาะอย่างยิ่งวิธี BMA_{OCC} และ วิธี BMA_{MC3} จะได้รับผลกระทบน้อยมากจากการเพิ่มจำนวนตัวแปรอิสระ สังเกตได้จากค่า AMSE ที่เพิ่มขึ้นเพียงเล็กน้อยเท่านั้น

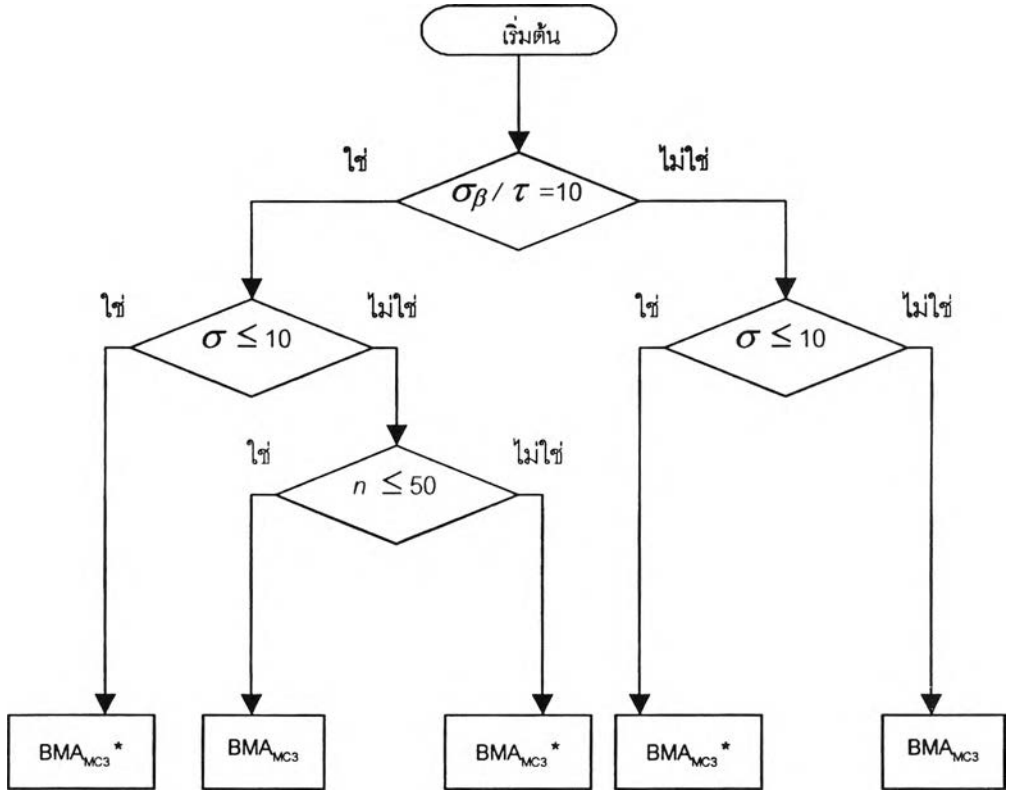
5) ค่าคงที่ σ_β/τ และ c

ค่าคงที่ σ_β/τ และ c จะมีผลต่อค่า AMSE ของวิธี BVS และวิธี BMA_{MC3} เท่านั้น โดยค่า AMSE จะเพิ่มขึ้นเมื่อค่าคงที่ σ_β/τ และ c สูงขึ้น เนื่องจากค่าคงที่ σ_β/τ และ c เป็นค่าที่กำหนดลักษณะการกระจายของพารามิเตอร์สัมประสิทธิ์การถดถอยที่สุ่มได้ ค่าคงที่ σ_β/τ และ c ที่สูงขึ้นจะทำให้การกระจายของพารามิเตอร์สัมประสิทธิ์การถดถอยมีการกระจายมากขึ้น ทำให้ค่าที่สุ่มได้มีความแม่นยำลดลง จึงส่งผลให้ค่า AMSE มีค่าสูงขึ้น การเปลี่ยนแปลงของค่าคงที่นี้จะส่งผลกระทบต่อวิธี BMA_{MC3} เพียงเล็กน้อย แต่จะส่งผลต่อวิธี BVS ค่อนข้างมาก ซึ่งสังเกตได้จากอัตราการเพิ่มขึ้นของค่า AMSE เมื่อค่าคงที่ σ_β/τ และ c เพิ่มขึ้น

5.1.3 ผลสรุปการเลือกวิธีการสร้างตัวแบบการถดถอยในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ

ผลสรุปการเลือกวิธีการสร้างตัวแบบการถดถอยในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณในการวิจัยครั้งนี้ พบว่าการสร้างตัวแบบการถดถอยด้วยวิธี BMA_{MC3} จะส่งผลให้ได้ค่าพยากรณ์ที่มีความถูกต้องแม่นยำมากที่สุด ส่วนวิธีการสร้างตัวแบบการถดถอยที่มีประสิทธิภาพในการพยากรณ์รองลงมา ได้แก่วิธี BMA_{OCC} BVS BIC และ SR ตามลำดับ สำหรับทุกระดับขนาดตัวอย่าง ค่าส่วนเบี่ยงเบนมาตรฐาน จำนวนตัวแปรอิสระ ระดับ

นัยสำคัญ และค่าคงที่วิธี BMA_{MC3} และวิธี BVS ซึ่งวิธี BMA_{OCC} จะมีประสิทธิภาพใกล้เคียงกับวิธี BMA_{MC3} ค่อนข้างมากในทุกกรณี และจากผลการวิจัยเปรียบเทียบค่า AMSE ของทั้ง 5 วิธี แม้ว่าค่า AMSE ของแต่ละวิธีเรียงลำดับจากน้อยไปมาก ได้แก่ วิธี BMA_{MC3} BMA_{OCC} BVS BIC และ SR ตามลำดับ ในทุก ๆ กรณีก็ตาม แต่พบว่ามีบางกรณีที่วิธี BMA_{MC3} มีค่า AMSE ต่ำกว่าวิธีอื่น ๆ ค่อนข้างมาก ซึ่งแสดงให้เห็นว่าในกรณีดังกล่าวเราควรเลือกวิธี BMA_{MC3} ในการสร้างตัวแบบการถดถอยเพื่อให้ได้ค่าพยากรณ์ที่ถูกต้องแม่นยำมากที่สุด ส่วนในกรณีอื่น ๆ นั้น แม้ว่าวิธี BMA_{MC3} จะให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ แต่ในการเลือกใช้วิธีการสร้างตัวแบบการถดถอย เราอาจคำนึงถึงปัจจัยด้านอื่น ๆ ด้วย เช่น ระยะเวลาในการคำนวณ ซึ่งวิธี BMA_{MC3} นั้นใช้ระยะเวลาในการคำนวณค่อนข้างมากเมื่อเทียบกับวิธี BMA_{OCC} โดยแผนผังแสดงข้อสรุปของกรณีต่าง ๆ ที่วิธี BMA_{MC3} จะให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ อย่างชัดเจนแสดงดังรูปที่ 5.1



รูปที่ 5.1 แสดงแผนผังผลสรุปการเลือกวิธีการสร้างตัวแบบในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ

หมายเหตุ : * หมายถึงกรณีที่วิธี BMA_{MC3} ให้ค่า AMSE ต่ำกว่าวิธีอื่น ๆ ค่อนข้างมาก

5.2 ข้อเสนอแนะ

1) การวิจัยครั้งนี้ศึกษาเฉพาะการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ ไม่รวมถึงตัวแบบความถดถอยพหุนาม ซึ่งเป็นกรณีเฉพาะที่มักพบได้บ่อยครั้งในการวิเคราะห์ความถดถอย ดังนั้นในงานวิจัยแบบเต็มรูปควรมีการศึกษาเปรียบเทียบวิธีการคัดเลือกสมการถดถอยสำหรับตัวแบบความถดถอยพหุนามด้วย ทั้งนี้เพื่อศึกษาว่าการนำวิธีการคัดเลือกสมการถดถอยทั้ง 5 วิธีไปใช้กับการวิเคราะห์ความถดถอยพหุนาม จะให้ผลสรุปเหมือนหรือแตกต่างกับการนำไปใช้กับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณดังเช่นในงานวิจัยครั้งนี้อย่างไร

2) วิธีการคัดเลือกสมการถดถอยภายใต้แนวทางของเบส์จะได้รับผลกระทบจากการแจกแจงก่อนค่อนข้างมาก ดังนั้นถ้ามีข้อมูลเพิ่มเติมซึ่งเป็นความควรจะเป็นเกี่ยวกับการแจกแจงของพารามิเตอร์ ก็ควรทำการปรับการแจกแจงก่อนตามความเหมาะสมกับปัญหาและข้อมูลที่มีอยู่ และแนวทางหนึ่งในการปรับการแจกแจงก่อน คือ สร้างการแจกแจงความน่าจะเป็นภายหลังใหม่ เมื่อกำหนดการแจกแจงความน่าจะเป็นภายหลังเดิมให้ทำหน้าที่เป็นหลักเกณฑ์ใหม่ตามรูปแบบดังนี้

การแจกแจงความน่าจะเป็นภายหลัง$_1$ α การแจกแจงความน่าจะเป็นก่อน$_1$ x ความควรจะเป็น$_1$

การแจกแจงความน่าจะเป็นภายหลัง$_2$ α การแจกแจงความน่าจะเป็นภายหลัง$_1$ x

ความควรจะเป็น$_2$

กล่าวคือ ถ้าการแจกแจงความน่าจะเป็นก่อน$_1$ เป็นหลักเกณฑ์แบบไม่มีข้อสนเทศ หลักเกณฑ์ใหม่สำหรับสร้างการแจกแจงความน่าจะเป็นภายหลังใหม่จึงมาจากความควรจะเป็น$_1$ ของข้อมูล ซึ่งคือการนำลักษณะของข้อมูลที่มีอยู่จริงมาปรับใช้เป็นหลักเกณฑ์ใหม่นั่นเอง

3) การกำหนดค่าคงที่ σ_β/τ และ c จะมีผลต่อค่า AMSE ของวิธี BVS และวิธี BMA_{MC3} โดยเฉพาะอย่างยิ่งวิธี BVS จะได้รับผลกระทบค่อนข้างมาก ดังนั้นอาจมีการใช้วิธีการลู่เข้ากึ่งอัตโนมัติ (Semiautomatic Approach) ในการกำหนดค่าคงที่ σ_β/τ และ c ที่เหมาะสมเพื่อเปรียบเทียบกับการกำหนดค่า σ_β/τ และ c เป็นค่าคงที่ดังที่ใช้ในการวิจัยครั้งนี้

4) การประมาณค่าพารามิเตอร์เวกเตอร์สัมประสิทธิ์การถดถอย (β) และค่าความแปรปรวน (σ^2) ด้วยวิธีกำลังสองน้อยสุด เพื่อใช้เป็นค่าเริ่มต้นสำหรับวิธี BVS และวิธี BMA_{MC3} นั้น เนื่องจากการวิจัยครั้งนี้ศึกษาเฉพาะการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณในกรณีที่ข้อตกลงเบื้องต้นของการวิเคราะห์ความถดถอยเป็นจริง ดังนั้นการใช้วิธีกำลังสองน้อยสุดในการ

ประมาณค่าเริ่มต้นจึงไม่ได้รับผลกระทบมากนัก แต่หากเป็นการวิเคราะห์ความถดถอยในกรณีที่ข้อตกลงเบื้องต้นไม่เป็นจริง เช่น การวิเคราะห์ความถดถอยพหุนาม การใช้วิธีกำลังสองน้อยสุดในการประมาณค่าเริ่มต้นอาจไม่เหมาะสมนัก จึงควรใช้วิธีการหาฐานนิยมของการแจกแจงภายหลังขอบ [1] ได้แก่ วิธีของนิวตัน (Newton 's Method) และวิธีการหาค่าสูงสุดแบบมีเงื่อนไข (Conditional Maximization Method) เป็นต้น ในการประมาณค่าเริ่มต้น

---

[1]Andrew Gelman, John B. Carlin, and Donald B. Rubin, **Bayesian Data Analysis** (London : Chapman & Hall, 1995), pp. 271-273.